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

Basic cognitive processes of intelligence

presented by

Gidon T. Frischkorn

year of submission

2018

Dean: Prof. Dr. Dirk Hagemann

Advisor: Prof. Dr. Dirk Hagemann

Adress: Gidon T. Frischkorn
Militärstrasse 91
CH-8004 Zürich
Schweiz
Mail: gidon.frischkorn@yahoo.de

First Reviewer: Prof. Dr. Dirk Hagemann
dirk.hagemann@psychologie.uni-heidelberg.de

Second Reviewer: Dr. Jan Rummel
jan.rummel@psychologie.uni-heidelberg.de

Acknowledgements

First and foremost, I want to thank my supervisor Prof. Dr. Dirk Hagemann. Not only have you guided and supported me throughout my student times and Ph.D, but you have also been critical and demanding while never losing appreciation for my personal ideas. Accordingly, you have contributed significantly to my development as a scientist.

Likewise, I would like to thank Dr. Jan Rummel who kindly agreed to review my thesis. And in addition, Jan, you too have been a role model for my work as a scientist.

During the past three years, I spent many hours discussing my work with colleagues and friends. A very special mention is warranted to Dr. Anna-Lena Schubert. It was, and still is, a gift to work with such a clear and critical thinker while also having a friend to talk about more than what is going on in science and making the office a place to dwell. In a similar vein, I want to thank our research assistants, Ben, Christoph, Isabel, and Larissa. So much of the work in our lab would not have been possible without you. You were a special and wonderful team like a second family to me. And I want to thank Dr. Andreas Neubauer: your curiosity and methodological interest has always inspired me and I hope to watch more than a few future super bowls with you.

My gratitude also goes to those not primarily involved in this thesis. A big thank you to my parents, Ursula and Hans-Georg, for your always open ears and ever lasting support. Thank you to my brothers, Moritz and Philip: you constantly keep my mind open to all the cultural things going on in your lives. And thank you to Andi: for your friendship and backing when I hassled. Ultimately, I want to thank my wife: Lea, you are the strongest company I could hope for in the rough seas of everyday life and my personal struggles. I will always have a hard time crafting to-do-lists. Still, every single motivation boost and every bit of love you give me is present in these pages.

Contents

List of Publications
of the publiction based dissertation

Manuscript 1

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

*both authors contributed equally to this work.

Manuscript 2

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

Manuscript 3

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

Manuscript 4

Frischkorn, G. T., Schubert, A.-L. & Hagemann (submitted). Processing Speed, Working Memory, and Exectuive Functions: Independent or inter-related predictors of general intelligence? *Intelligence*.

1 Introduction

The ability to think and reason and to mentally solve problems is among the most intriguing traits of mankind. Unsurprisingly, the scientific study of *intelligence*, as we call it, is one of the most and longest studied topics in psychology. In the beginning of the 20th century, Spearman (1904) initiated the branch of *correlational psychology* and established the concept of general intelligence. In the same year, Binet and Simon (1904) published the first intelligence test with the aim to classify the cognitive progress of school children. The Binet-Simon scale was well received, and soon further refined and translated (Binet & Simon, 1911; Burt, 1922; Terman, 1916). These first steps in the development of intelligence testing represented the starting point of the modern measurement of intelligence. In fact, until today a revised form of the Binet-Simon scale, the Stanford-Binet test, is still used to measure intelligence (Roid, 2003).

Following the success of intelligence tests, one of the first definitions of intelligence was simply operational as: "intelligence is simply what the tests of intelligence test" (Boring, 1923). Arguably, such an operational definition of intelligence by its measurement is no longer up-to date (Hunt & Jaeggi, 2013; Johnson, 2013). Today, intelligence is understood as "the ability to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77) or as "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (Wechsler, 1944, p.3). Despite these theoretical conceptualizations of intelligence[1], intelligence research remained strongly focused on psychometric questions such as the factor structure of intelligence.

[1]Some researchers argue and criticize that intelligence is still defined operationally (van der Maas, Kan, & Borsboom, 2014).

In the tradition of Spearman (1904), a plethora of factor analytic studies addressed the question which psychometric model most adequately describes the structure of intelligence. Starting with the concept of general intelligence (*g*) - the shared variance of cognitive tests - researchers were in debate whether *g* is formed by different uncorrelated factors (Kovacs & Conway, 2016; Thurstone, 1938) or represents a broad and domain general factor and domain-specific residuals (Carroll, 1993; Horn & Cattell, 1966; Jensen, 1998; Spearman, 1927). Modern psychometric analyses indicate that intelligence consists of both, a general factor determining performance in a broad variety of tasks, and specific uncorrelated factors (i.e. residual factors) that are related to domain- or task-specific cognitive performance (McGrew, 2005, 2009).

Nevertheless, the positive manifold of cognitive tests still is an enigma demanding for a theoretical explanation. Empirically, results show that heterogeneous sets of cognitive tasks share between 40 to 50 percent of variance (Jensen, 1998; Johnson, Bouchard Jr., Krueger, McGue, & Gottesman, 2004). More importantly, *g* factors from different cognitive test batteries show large overlaps (r_s = .77 – .99; Johnson, te Nijenhuis, & Bouchard, 2008). Both the emergence of *g* in any set of cognitive tests and the functional invariance across different test sets were often interpreted as an indicator that few domain-general cognitive processes underlie individual differences in intelligence (Jensen, 1998; Spearman, 1927).

Already in the beginning of intelligence research, cognitive process domains such as processing speed or attention have been linked to general intelligence (Peak & Boring, 1926; Spearman, 1904). More recent correlational studies further emphasize the role of processing speed (Jensen, 2006; Schubert, Hagemann, & Frischkorn, 2017) as an important covariate of *g*. Furthermore, later established psychological constructs such as working memory (Baddeley & Hitch, 1974) and attention regulation mechanisms such as executive functions (Miyake et al., 2000) have been linked to *g* as well (Ackerman, Beier, & Boyle, 2005; Engle & Kane, 2003; Kyllonen & Christal, 1990; Unsworth & Spillers, 2010). Nonetheless, a comprehensive theoretical account why and how these different cognitive processes are related to individual differences in intelligence is still due.

The present dissertation aims to contribute conceptually and empirically to this very question: *what are the basic cognitive processes of intelligence?*. To address this question, I identified three issues that my dissertation is focused on:

1. How can we adequately measure individual differences in cognitive processes?
2. How can we test which cognitive processes underlie individual differences in intelligence?
3. How can different cognitive processes related to intelligence be integrated?

In detail, the first two manuscripts of my thesis were directed on answering the first question and discussed how to incorporate theoretically founded measures of a cognitive process (e.g. parameters of a cognitive model) into intelligence research (Manuscript 1), and collected empirical evidence assessing which parameters of the drift-diffusion model possess trait-like properties qualifying them as predictors of individual differences in intelligence (Manuscript 2). The third manuscript contributed to the second question and developed a parametrization for the worst performance rule to provide future research with means to asses which cognitive processes may explain this phenomenon. And the fourth manuscript of my thesis addressed the last question and aimed to integrate the currently most prominent candidate processes underlying individual differences in intelligence.

2 Measuring individual differences in cognitive processes

In order to meaningfully link cognitive processes to intelligence, adequate measures of individual differences in cognitive processes are needed. Two points play an important role for this: (1) the measurement of a cognitive process should meet the three basic psychometric characteristics of objectivity, reliability, and validity (Adams, 1936), and (2) the cognitive process should exhibit trait-like properties similar to intelligence to ensure a symmetry in interpretation between the cognitive process and intelligence (Wittmann, 1988; Wittmann & Süß, 1999). In this, we have to evaluate the psychometric properties of individual difference measures of a cognitive process, and we have to ensure that the cognitive process shows consistency across different situations and tasks, just as intelligence does (Danner, Hagemann, Schankin, Hager, & Funke, 2011; Deary, Whalley, Lemmon, Crawford, & Starr, 2000) .

Evaluating the psychometric properties of a measurement

A psychometric measurement is understood as the mapping of an empirical observation onto a formal – mostly numerical – representation (Narens & Luce, 1986; v. Helmholtz, 1921). This mapping procedure is required to follow three main quality criteria: objectivity, reliability, and validity. Objectivity means that the mapping has to be independent of the observer, reliability that the measurement needs to be sufficiently precise with respect to the construct to be measured, and validity that the measurement actually represents the construct, trait or process of interest. In the end, the adequacy of a measurement for a cognitive process is evaluated by how well these psychometric criteria are met.

Cognitive process parameters related to intelligence (e.g. working memory capacity, processing speed, or executive functions) are usually measured by the response time and accuracy in a task. These measures can be determined more or less objectively.[1] With some additional assumptions, for example from models of classical test theory (CTT) or structural equation models (SEMs), the reliability of such measures can be estimated as well (Lord & Novick, 1968). Although estimating and interpreting reliability estimates is certainly not trivial (Cronbach, 1951; Cronbach & Shavelson, 2004; Guttman, 1945; Jackson & Agunwamba, 1977; Mellenbergh, 1996), assessing the validity of these measures is by far the most complex problem (Borsboom, 2005).

Following one of the current conceptions of validity, the critical point for a valid measure of a cognitive process is that variations in the underlying process are responsible for variation in the measurement (Borsboom, Mellenbergh, & van Heerden, 2004). This problem has often been resolved by implicitly equating the process to be measured with its measurement. For example, the number of correctly recalled items from a memory set after a short-delay is said to represent the capacity of short-term memory. By conceptually adopting a reflective measurement model (see Figure 2.1, p. 6), it is then stated that any variation in the capacity of short-term memory causes different numbers of correctly recalled items. However, such a simple reflective measurement model does not provide an account for how the latent cognitive process translates to the observed behavior and why changes in the cognitive process lead to differences in behavior (Rhemtulla, van Bork, & Borsboom, 2015). Moreover, it postulates the existence of a cognitive process without any evidence regarding its ontological status (Borsboom, 2005, 2008).

Finding valid measures for cognitive processes is thus concerned with identifying how a latent cognitive process relates to or determines the associated observed behavior and discussing the ontological status of the cognitive processes to be measured.

[1] For instance, response time is usually defined as the time taken from the beginning to the end of the processing of a task and accuracy is operationalized as a match of the response with a logically derived correct solution.

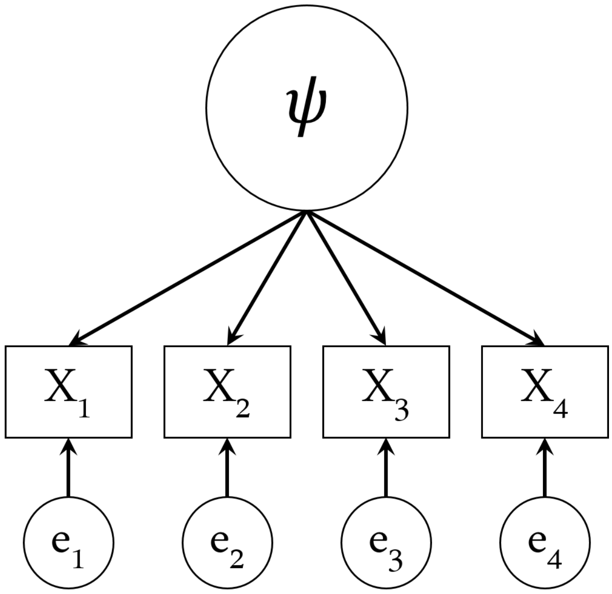

Figure 2.1: Path diagram of a reflective measurement model assuming that variations in the observed variables X_{1-4} stem from variation in the latent cognitive process ψ and measurement error e_{1-4}.

This would address the two core problems of reflective measurement models (Borsboom, 2005). In order to achieve this, we first have to develop a theoretical model that specifies how the observed behavior (i.e. reaction time or accuracy) is determined by the cognitive process to be measured. Although such a model may well be wrong (Box, 1976), I would argue that it will always provide a more useful approach than equating cognitive processes and observed behavior without accounting for their relation. Secondly, we have to establish the ontological status of the cognitive process by linking it to undeniably existing measures.[2] For example, assuming a physical basis of a cognitive process, biological measures can be used as a preliminary reference for the existence of a cognitive process.

[2]How to establish the ontology or existence of a process or measure, is an epistemological problem. The methods used in this realm depend significantly on the philosophical position researchers take (Borsboom, 2005). A *realist* position assumes that psychological constructs such as cognitive processes exist independent of the observer. In contrast, *anti-realist* positions, such as positivism or constructivism, assume that psychological constructs are part of our imagination or at least perception and thus not independent of the observer. Unfortunately, the epistemological problems related to these positions cannot be discussed in more detail in this dissertation.

Structural properties of personality traits

Apart from the valid measurement of cognitive processes and their psychometric properties, any cognitive process supposedly related to intelligence or any other personality trait has to exhibit similar trait-like properties. Despite that the ontological and theoretical status of personality traits such as intelligence have been under debate in personality theory ever since (Allport, 1937; Cattell, 1946; Eysenck, 1967, 1981), all personality theories expressed similar structural properties defining a trait of personality.

The two most important structural properties of a personality trait and its measurement are situational consistency and independence of assessment methods. Conceptually, this relates to the discussion on how far observed behavior is comprised of situational and testing contexts, a person's traits and the interaction between these variables (Cronbach & Snow, 1977; Endler & Magnusson, 1976). Two approaches to dissolve the contribution of situational characteristics and personality traits to overt behavior have been put forth. One proposal was to randomly vary situations and tasks for a sample of people in order to separate the variance components of the different factors with variance analytic methods (Brennan, 2001). The other approach suggested to use correlation matrices (Campbell & Fiske, 1959) and later SEMs or confirmatory factor analysis (CFA) to decompose variance proportions unique to persons from situation and method related variance (Eid, 2000; Steyer, Schmitt, & Eid, 1999).

In general, a trait-like cognitive process should be characterized by a strong consistency across situations and methods. Empirically, this can be estimated by measuring the same cognitive process in multiple situations with different tasks. Intelligence, for example, has shown a remarkable differential stability with longitudinal correlations of $r \approx .70$ over a time span of more than 60 years (Deary et al., 2000). Moreover, between 40 to 70 percent of variance is captured by a task general factor of general intelligence (g) in heterogeneous sets of cognitive tasks (Danner et al., 2011; Johnson et al., 2004). Such a large proportion of variance consistent across situations and tasks, together with little situation- and task-specific variance proportions, represents strong evidence for the trait characteristics of intelligence. Any cognitive process underlying individual differences in *g* should thus exhibit similar trait-like properties.

Towards more adequate representations of cognitive processes.

Taken together, there are two important aspects when measuring cognitive processes related to intelligence that I want to address within my dissertation and the associated manuscripts. First, I aim at incorporating measures that are theoretically and biologically more closely connected to the cognitive processes of interest. And second, I want to provide evidence that these measures exhibit trait-like properties and thus can be interpreted symmetrically to intelligence. The following two sections will introduce two different kinds of measures that better meet these requirements: (1) parameters from cognitive models, and (2) neuro-physiological measures such as the latency of event-related potential (ERP) components. Specifically, I will first explain how these two approaches lead to a more adequate representation of individual differences in cognitive processes. And second, I present evidence that both these measures exhibit trait-like properties qualifying them as symmetric to intelligence.

2.1 Cognitive models as measurement tools

(Manuscript 1)

As pointed out in the outline of this section (see p. 5), one problem of reflective measurement models is that they do not specify how the latent cognitive processes translates into observed behavior. To address this problem, the first manuscript of my dissertation (Frischkorn & Schubert, 2018, see p. 50) advocates cognitive models as a solution. Cognitive models are mathematical formalizations of a cognitive process.[3] In detail, cognitive models translate verbal theories (i.e. descriptive models of a cognitive process) into mathematical formalizations of these theories (Farrell & Lewandowsky, 2018). Such models provide a detailed formalization of the interplay of different processes that generate the observed behavior.

[3]Cognitive models have been referred to with various terms such as computational models, mathematical models or formal models. For consistency, I will use the term cognitive models to refer to mathematically formalized models of a cognitive process within the present thesis.

Specifically, I identify three major benefits of cognitive models for intelligence research in the first manuscript of my dissertation (Frischkorn & Schubert, 2018). Namely, (1) cognitive models provide a testable conceptualization of cognitive processes, (2) parameters from cognitive models can be interpreted more objectively, and (3) individual differences in cognitive model parameters can be linked to individual difference in intelligence. In addition, the first manuscript presents models for processing speed, working memory and attention that are particularly relevant for intelligence research and provides guidelines for their application.

In general, I argue within this manuscript that due to the clear formalization of parameters in a cognitive model, individual differences in cognitive model parameters provide a more adequate representation of individual differences in a cognitive process. Most importantly, they provide a clear formalization how difference parameters of a cognitive process are translated into the observed behavior. Beyond that, individual differences in cognitive model parameters that show strong relationships with intelligence provide theory guided evidence for the basic cognitive processes of intelligence.

Finally, with respect to the interpretation and ontology of cognitive model parameters, the first manuscript of my dissertation highlights two important points: First, the semantic interpretation of model parameters is always dependent on the task and context they are estimated in. And second, cognitive models are always simplifications of a latent cognitive process and can hardly capture the entirety of its existence and architecture. However, I also point out that linking cognitive process parameters to biological measures, such as psycho-physiological or neuro-imaging parameters, can serve as a reference to inform us in how far model parameters are reflected in neural correlates. This idea will be further elaborated in section 2.2 (p. 14).

Models for performance in intelligence tests

If cognitive models provide an account for the cognitive processes underlying observed behavior in a measurement, why do we not formulate a cognitive model for intelligence tests? There actually have been attempts to formalize individual differences in intelligence test performance. Carpenter, Just, and Shell (1990) formulated a model for performance in the raven matrices test (Raven & Raven, 2003). In detail, this model described different processes and strategies that could be used while solving a raven test item. Interestingly, some of these processes (e.g. incremental encoding or rule induction) were used by all participants, while other processes (e.g. induction of abstract rules or dynamic management of different rules) were only used by above-average performing participants.

The cognitive model by Carpenter et al. (1990) represents an important step towards a more formalized and process oriented account of individual differences in intelligence. Still, its application is limited to matrix reasoning tasks. While matrix reasoning tasks may be the best single item indicator of *g* (Martinez, 2013; Spearman, 1946), it has been argued that *g* is always better represented by a set of heterogeneous tasks than one single indicator (Gignac, 2015). As the model by Carpenter et al. (1990) is restricted to matrix reasoning, and cannot be adapted for other measures of intelligence, its parameters cannot exhibit trait characteristics such as task independence. In sum, Carpenter et al. (1990) attempt was a start but does not provide a exhaustive account of individual differences in *g*.

Therefore, I argue to take a step back and exploit advances in cognitive psychology; an area in which various cognitive models for processes that may underly *g* have been developed over the past decades. Linking person-specific parameter estimates from such models with individual differences in *g* provides insights into which specific aspects of a cognitive process are related to intelligence. This evidence then provides the basis for combining different processes related to *g* into a comprehensive process theory of intelligence.

One promising model that has already been used in intelligence research (e.g. Ratcliff, Thapar, & McKoon, 2010; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schmitz & Wilhelm, 2016; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015) is the drift-diffusion model (DDM). The DDM assumes that a decision in binary choice tasks stems from an information accumulation process toward one of two response alternatives (see Fig. 2.2, p. 11). This accumulation process is defined as a random

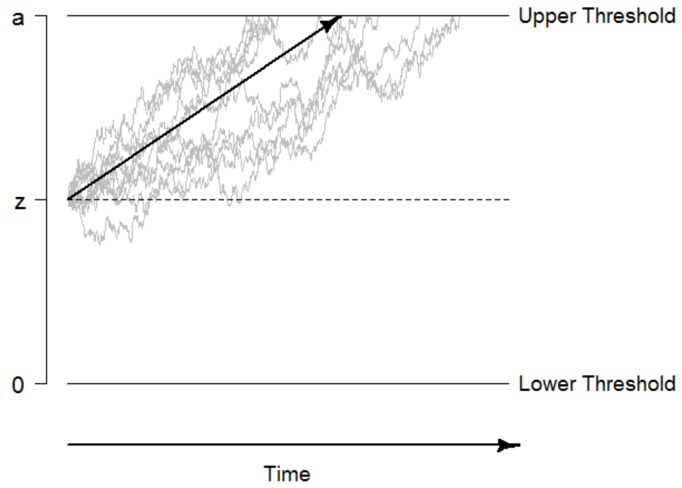

Figure 2.2: Graphical illustration of the drift-diffusion model (DDM). The drift rate (illustrated by the arrow) determines the speed of information accumulation. The boundary separation a can be interpreted as response caution, and the starting point z represents a bias towards one of the two response alternatives. The non-decision time t_0 is not depicted in this illustration and is simply added to the time of the above shown decision process.

walk with a systematic component, the drift rate (v), and random noise (Ratcliff, 1978). Information accumulation stops until one of the two response thresholds is reached. The distance between these response thresholds is specified by the boundary separation (a), and the accumulation process starts at the starting point (z) between the two boundaries. Finally, time that is not related to this decision process is captured in the non-decision time (t_0).

The DDM nicely demonstrates the benefits of cognitive models. While behavioral reaction timess (RTs) mixes the contribution of different processes within a decision, the DDM decomposes the information from the distribution of response times into distinct parameters representing the different aspects of a decision. More importantly, experimental studies have validated the interpretation of DDM parameters (Voss, Rothermund, & Voss, 2004). In this, drift rate v can be seen as a theoretically more valid representation of processing speed than individual differences in mean RTs.

Trait characteristics of cognitive processes parameters (Manuscript 2)

As the DDM has become a very popular cognitive model in intelligence research, I evaluated which parameters of the DDM exhibit trait-like properties in the second manuscript of my dissertation (Schubert, Frischkorn, Hagemann, & Voss, 2016, see p. 72). In detail, we estimated drift rate, boundary separation and non-decision time in three different elementary cognitive tasks (ECTs) across two measurement occasions. With on average 44% of variance captured by a task- and situation general drift factor, drift rate was similarly consistent across different tasks and measurement occasions as intelligence. In contrast, consistency for boundary separation and non-decision time was lower with the task- and situation-general factors capturing on average 32 to 36% of variance. Interestingly, the variance consistent across the three different task showed no situation specific variance, and state residuals could all be fixed to zero.

On the one hand, the second manuscript of my dissertation (Schubert et al., 2016) thus exhibited how trait-properties for parameters of cognitive models can be evaluated, and on the other hand established that drift rate is the best candidate parameter of the DDM to be related to intelligence. The results however stress that it is important to measure drift rate across different tasks, because single task indicators contain a considerable amount of task-specific variance (up to 17% as estimated by Schubert et al., 2016). This amount of task specific variance in single task drift rates might also explain why empirical estimates of the relationship between drift rate and intelligence varied considerably (r = .50 – .90; Ratcliff et al., 2010; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert et al., 2015).

In summary, the first manuscript of my dissertation (Frischkorn & Schubert, 2018) demonstrated how cognitive models may provide a more valid representation of individual differences in cognitive processes, because they provide more explicit measurement models that connect the latent cognitive process with observed behavior. Moreover, the second manuscript of my dissertation (Schubert et al., 2016) showed that drift rate of the DDM exhibits trait-like properties rendering it a suitable predictor of individual differences in intelligence.

2.2 Neuro-biological measures of cognitive processes

Similar to parameters from cognitive models, neuro-physiological measures such as latencies or amplitudes of the event-related potential (ERP) can provide a more valid representation of individual differences in cognitive processes than observed reaction times or accuracies. The high time-resolution of the electro-encephalogram (EEG) and ERPs allows to decompose the neural stream of information processing between stimulus onset and response into distinct components (e.g. the N1 or P3). These components can be differentiated by polarity, topography, and latency after stimulus onset. In addition, the EEG reflects physical properties of the information processing within the brain and can thus be used as biological reference confirming the ontological status of cognitive processes.

Over the past decades, a plethora of experimental research has sharpened the psychological interpretation of different ERP components. Specifically, a number of studies investigated which experimental manipulations elicit specific ERP components or affect their latency or amplitude (Luck, 2005; Luck & Kappenman, 2011). Ultimately this has led to a functional interpretation of the different ERP components that arguably separates different aspects of cognitive processes more validly than behavioral RTs or accuracy.

Unfortunately, the consistency across different tasks and situations of parameters of ERP components has often been neglected. Specifically, amplitudes and latencies of ERP components have often been determined only for a single task and have then been related to measures of intelligence (Bazana & Stelmack, 2002; McGarry-Roberts, Stelmack, & Campbell, 1992; Troche, Houlihan, Stelmack, & Rammsayer, 2009). This has led to inconsistent results with respect to the relationship of ERP components and intelligence. Beyond that, past research often neglected the question whether ERP components exhibit trait-like properties that are required for an adequate predictor of individual differences in intelligence.

A recent study addressed these problems and suggested that as long as parameters of ERP components are determined across multiple tasks, situational effects may be neglected (Schubert et al., 2017). Specifically, latencies of ERP components contained a considerable amount of task specific variance. However, the variance consistent across different tasks contained no significant situation-specific variance. In this, latencies or amplitudes of ERP components that are measured in different task seem to be relatively stable across time. On the basis of these results, I adopted EEG measures across three different tasks in the fourth manuscript of my dissertation (Frischkorn, Schubert, & Hagemann, submitted) summarized in section 4.1 (p. 25) to enrich the interpretation of behavioral measures of cognitive processes.

Cross validating parameters from cognitive models and neuro-physiological process parameters

Although the manuscripts of my dissertation make no contribution to this point, I want to point out one very promising approach to a more comprehensive understanding of basic cognitive processes of intelligence: the simultaneous analysis of both parameters of cognitive models and neuro-physiological process parameters as predictors of intelligence (for an example, see Schubert, Nunez, Hagemann, & Vandekerckhove, 2018). These so called cognitive latent variable models (CLVMs) combine the advantages of mathematically formalized cognitive models (see section 2.1, p.8ff.) with benefits of neuro-psychological process parameters (see section 2.2, p.13f.). Such CLVMs assume that both parameters of cognitive models derived from behavioral measures and process parameters from neuro-physiological data reflect properties of the same latent cognitive process (Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017). Accordingly, these two kinds of parameters should be strongly interrelated and predict individual differences in intelligence alike.

Empirically, especially parameters of the DDM and neuro-cognitive process parameters have been linked. The P3 component, for example, has been shown to be a neural indicator of the evidence accumulation process captured in the drift rate of the DDM (Kelly & O'Connell, 2013; O'Connell, Dockree, & Kelly, 2012; Ratcliff, Sederberg, Smith, & Childers, 2016; van Ravenzwaaij, Provost, & Brown, 2017). Further, individual differences in P3 amplitude have been shown to explain up to 74 percent of variance in drift rates (Ratcliff, Philiastides, & Sajda, 2009). These results are especially interesting for discussing the ontology of cognitive model parameters and the interpretation of ERP components. The strong link of drift rate and P3 amplitude suggests that both these measures reflect the same neuro-biological process. In this, the use of parameters from cognitive models together with neuro-cognitive process parameters provides a promising way to address both of the critical issues of reflective measurement models (see p. 5), and investigate theoretically and ontologically grounded basic cognitive processes of intelligence in future studies.

3 Testing explanations of important phenomena in intelligence research

The adequate measurement of individual differences in cognitive processes related to intelligence is only one aspect in the realm of searching for the basic cognitive processes of intelligence. Another critical question is, how we can evaluate which cognitive processes underlie individual differences in intelligence and whether they explain important phenomena in intelligence research? The present chapter will focus on two different approaches towards this aim, and highlight how the third manuscript of my dissertation (summarized on p. 18f.; Frischkorn, Schubert, Neubauer, & Hagemann, 2016) contributed to help identifying the basic cognitive processes of intelligence.

To test explanations for individual differences in intelligence, it is necessary to, first, parameterize important phenomena and second, to design experiments that allow for a causal interpretation of their results. Individual differences in intelligence are arguably readily accessible via performance scores from intelligence tests. However, other important phenomena, for example the worst performance rule (WPR), have been approached rather descriptively (Frischkorn et al., 2016; Schubert, 2018). Developing a parametrization of phenomena like the WPR is important in order to statistically test which cognitive process might explain this phenomenon. In this, a parameterization of such findings is a necessary first step to provide statistical inference on possible explanations.

In addition to the parametrization of the phenomena to be explained, study designs that allow to establish a causal relationships between cognitive processes and intelligence are necessary. Hitherto, almost any evidence concerning the basic cognitive processes of intelligence is based on correlations. However, correlations can hardly be interpreted causally (Danner, Hagemann, & Fiedler, 2015; Mooij, Peters, Janzing, Zscheischler, & Schölkopf, 2016; Tabachnik & Fidell, 2013). Nevertheless, a plethora of correlational studies mostly adopting structural equation model (SEM) has put forward claims about different cognitive processes underlying individual differences in intelligence or explaining phenomena such as the WPR. In detail, the capacity of working memory (Ackerman et al., 2005; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhelm, & Süß, 2005) or short-term memory (Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Colom, Jung, & Haier, 2007), the speed of information processing (Jensen, 2006; Schmitz & Wilhelm, 2016; Schubert et al., 2017, 2015; Sheppard & Vernon, 2008), and attention regulation mechanisms (Engle & Kane, 2003; Kane, Conway, Hambrick, & Engle, 2007; Kane et al., 2016) have all been discussed as basic cognitive processes of intelligence and associated phenomena – e.g. the WPR – on the basis of correlations.

In the case of working memory capacity (WMC) as a basic cognitive process of intelligence, there has even been a vivid discussion in how far WMC might be inseparable to intelligence (Ackerman et al., 2005; Kane, Hambrick, & Conway, 2005; Oberauer et al., 2005). It is however important to note that perfectly correlated measures can still refer to different constructs. For example, the pressure and volume of a gas in a closed system are perfectly correlated as described by Boyle's law[1] (Boyle, 1662). However, these two measures are obviously not equal. Insofar, correlational evidence does not suffice if we are to determine the causally basic cognitive processes of intelligence or explanation for the WPR.

To establish which of the cognitive processes correlated with intelligence cause individual differences in intelligence or explain the WPR experimental studies are needed. An experiment provides the possibility to systematically manipulate specific cognitive processes such as processing speed (PS) or WMC. Decreases in performance in intelligence measures or changes in the WPR that can be attributed to these manipulations

[1] This is the case as long as the temperature and amount of gas stay constant.

then provide evidence for their causal link to intelligence. Ultimately, it is thus necessary to bridge this often criticized gap between correlational and experimental psychology (Cronbach, 1957) to identify the causally basic cognitive processes of intelligence.

3.1 Parametrization of the Worst Performance Rule (Manuscript 3)

The third manuscript of my dissertation (Frischkorn et al., 2016, see p. 18) addressed the problem that the WPR has mostly been described rather than tested. The WPR describes the phenomenon that worst rather than average or best performance of participants within a cognitive task shows the largest relationship to intelligence (Coyle, 2003b; Larson & Alderton, 1990). Not only does this contradict basic assumptions of measurement theories such as classical test theory (CTT) but it amplifies in tasks that are highly related to general intelligence (*g*) and in samples with below average cognitive abilities that rely more strongly on *g* related cognitive abilities (Blum & Holling, 2017; Coyle, 2003b; Spearman, 1927). In this, the cognitive processes that are responsible for the WPR might also be strongly related to *g*.

Specifically, the third manuscript identified the problem that studies evaluating the WPR typically computed correlations across performance percentiles with intelligence and then described the increase in correlations from best to worst performance. Any group or task comparisons (e.g. between younger and older adults or tasks with low compared to high *g*-loading) were then made by descriptively comparing the increase in correlations instead of quantifying or testing differences in the WPR (Coyle, 2003a; Rammsayer & Troche, 2016; Ratcliff, Schmiedek, & McKoon, 2008; Ratcliff et al., 2010). Although the difference in correlations can be tested (Steiger, 1980), this test often lacks statistical power (Cohen, 1988). Thus, to effectively review any theoretic explanation of the WPR a parameterization of the WPR is required.

To overcome the problem of describing rather than testing the WPR, I developed a parameterization for the WPR in the third manuscript of my dissertation (Frischkorn et al., 2016). This parameterization reformulated the WPR as a moderation. In detail,

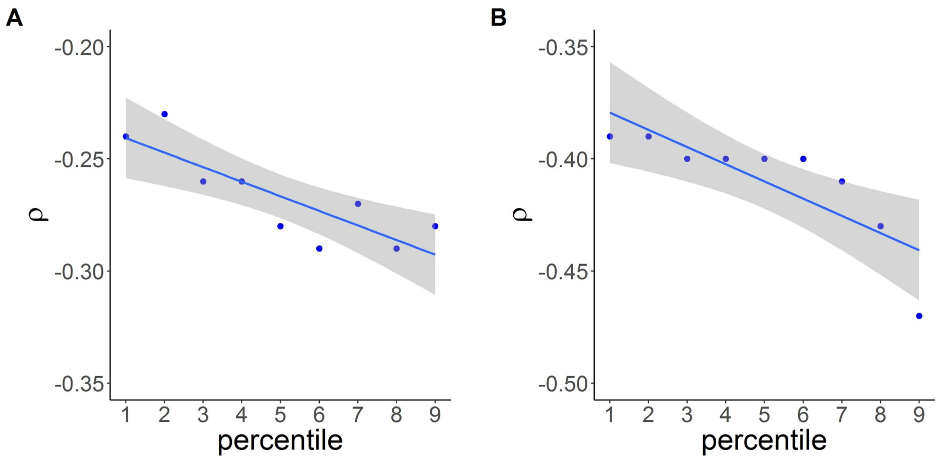

Figure 3.1: Visualisation of the worst performance rule (WPR) as evaluated in Frischkorn et al. (2016). The left side (A) displays the WPR for set size one in the Sternberg memory scanning task, and the right side (B) for set size five. The line displays the regression as described in equation 3.1. The gray area around the regression line displays the standard error of estimation for this linear regression.

the relationship ρ between performance in a given percentile p and intelligence can be described as:

$$\rho_p = \beta_0 + p \cdot \beta_1 \tag{3.1}$$

In this regression, β_0 represents the relationship ρ for the centered percentile, and β_1 represents the increase or decrease of this relationship across percentiles. As long as β_1 is significantly different from zero, the correlation ρ between performance in the cognitive task and intelligence across percentiles will change (see Fig. 3.1, p. 19). Thus testing the significance of the WPR essentially means evaluating whether β_1 differs from zero or not.

While I preliminarily implemented the increase of correlations across percentiles in a sequential manner in the third manuscript (as it was usually done when describing the WPR), this does not account for the uncertainty in correlations as statistical estimates (Skrondal & Laake, 2001). Therefore, I further recommended to implement this moderation in a hierarchical model that simultaneously estimates the relationship between intelligence and performance in a cognitive task while also modeling the increase across percentiles. On the one hand, this conveys the moderation aspect of the new parametrization of the WPR more clearly (see the following paragraph for an extended description). On the other hand, it complies with recent recommendations for hierarchical models from the area of cognitive modeling (Boehm, Marsman, Matzke, & Wagenmakers, 2018), and the possibility of incorrect inferences due to the sequential analysis can be avoided. Specifically, results from the third manuscript showed that a sequential approach underestimates the standard errors of regression parameters and thus overestimates the significance of a possible worst performance effect.

In detail, the hierarchical model of the WPR advocated in the third manuscript is a regression of performance Ψ of the participants i across the percentiles p in a cognitive task on intelligence *g*:

$$\hat{\Psi}_{ip} = \beta_{00} + \beta_{1p} \cdot g_i$$

To additionally implement the WPR in this regression, the regression weight β_{1p} is in turn predicted by the percentile:

$$\beta_{1p} = \beta_{10} + \beta_{11} \cdot p$$

A combination of these two nested, and thus hierarchical, regressions results in an interaction of percentile p and intelligence g across percentiles:

$$\hat{\Psi}_i = \beta_{00} + \beta_{10} \cdot g_i + \beta_{11} \cdot p \times g_i$$

Conceptually, β_{10} then represents the relationship between performance Ψ in the cognitive task and intelligence g for the centered percentile (i.e. $p = 0;$ Wainer, 2000), and β_{11} represents the change of this relationship across percentiles. This formalization elucidates more clearly why the WPR can be seen as a moderation: namely, the interaction quantified in β_{11} of percentile p and intelligence g parameterizes the WPR and can be interpreted as a moderation.

In summary, the parametrization of the WPR developed in the third manuscript of my dissertation (Frischkorn et al., 2016) provided future research with the means to evaluate different theoretical explanations of the WPR. This can be achieved with two approaches: (1) either we additionally control for individual differences in other cognitive processes across percentiles, or (2) we specify a three-way interaction of the additional cognitive process and the WPR interaction (β_{11}). Both these methods would allow to estimate in how far β_{11} changes with respect to additional cognitive processes. This provides future research with the means to evaluate different hypothesis concerning the WPR.

Theoretically, it has been argued that either attentional lapses (Unsworth, Redick, Lakey, & Young, 2010) or the speed of information accumulation (Ratcliff et al., 2008) might underlie the WPR. To empirically test these hypothesis, one very recent study measured and controlled for attentional lapses across percentiles and evaluated in how far the amount of attentional lapses may explain the WPR (Löffler, Schubert, Frischkorn, Rummel, & Hagemann, 2018). The results indicated that controlling for attentional lapses did not affect the WPR. In a similar vein, the drift rate of the drift-diffusion model (DDM) was used to evaluate whether it explains the WPR (Dutilh et al., 2017). But as Dutilh et al. (2017) did not find a WPR in their data, they were not able to evaluate whether drift rate does accutaly explain the WPR.

Beyond that, the parametrization of the WPR also allows to summarize results from multiple studies and obtain meta-analytic evidence on the generalizability of the WPR (Schubert, 2018). This stresses the significance of parameterizing important phenomena of intelligence research both with respect to cumulative science and their theoretical explanation.

3.2 Experimental manipulation of cognitive processes

Still, the parameterization of important phenomena in intelligence research is only the first step in finding and testing explanations for the different phenomena. In addition, studies that go beyond information on bi-directional relationships (e.g. correlations) are necessary.[2] Yet, research on intelligence has traditionally focused on correlations. And while causal evidence can be obtained via correlations under very specific circumstances (c.f. Danner et al., 2015; Mooij et al., 2016), these are hardly given in the case of most correlational studies in the area of intelligence research. It has therefore

[2]Although I did not provide any own publications to the following point within this dissertation, I will still briefly discuss the issue of experimentally studying the basic cognitive processes of intelligence for the sake of completeness.

often been argued that we have to incorporate experimental studies supplementing the correlational results in intelligence research (Cronbach, 1957).

Only few studies have bridged this gap, but these studies in particular often provided important insights on the basic cognitive processes of intelligence. For instance, Rao and Baddeley (2013) used different secondary tasks demanding specific processes in working memory while participants worked on items from a raven matrices test (Raven & Raven, 2003). By comparing processing time and accuracy of matrix reasoning while participants had to count backward to control conditions such as engaging in articulatory suppression or a silent baseline, Rao and Baddeley (2013) were able to show that time taken to solve an intelligence test item increases while counting backwards. While accuracy – the more central measures in raven matrices – showed smaller and inconsistent effects, this study still provides first evidence that especially central executive processes may be causally related to performance in an intelligence test.

Apart from using secondary tasks that capture resources of specific cognitive processes, another possibility to experimentally manipulate cognitive processes is psychopharmacology. Schubert, Hagemann, Frischkorn, and Herpertz (2018), for example, used a nicotine manipulation to increase neural speed of processing and evaluated in how far this leads to increases in intelligence test performance. Interestingly, results of this study showed an effect of nicotine administration on neural speed of processing, but no effect on intelligence test performance. With respect to the well established correlation between mental speed and intelligence (Schmitz & Wilhelm, 2016; Schubert et al., 2017; Sheppard & Vernon, 2008) this result suggests that a third variable not affected by the nicotine administration might underlie the relationship of mental speed and intelligence. This idea will be discussed in more detail in chapter 4 (p. 24ff.)

Taken together, experimentally manipulating specific cognitive processes and evaluating how such manipulations interfere with performance in intelligence tests is a promising approach towards identifying the basic cognitive processes of intelligence. Obviously, the specificity of such manipulations with respect to a single cognitive process needs to be discussed, which is the reason why experimental studies on the basic cognitive processes of intelligence (Hagemann, Schubert, & Frischkorn, 2016; Schubert, Hagemann, et al., 2018) always incorporated manipulation checks. In this realm, again cognitive models could provide further insight which specific part of a cognitive process was tapped by a manipulation. Building on the first promising results, an important future direction would thus be to further close the gap between correlational and experimental studies in intelligence research (Cronbach, 1957)

4 Integrating different cognitive processes correlated with intelligence

Last but not least, after summarizing my contributions to the issues of measuring cognitive processes and testing explanations for individual differences in intelligence, I want to address the third aspect raised in the Introduction (see p. 2): How can different cognitive processes related to intelligence be integrated?

As emphasized throughout this dissertation, different cognitive processes have been linked to individual differences in intelligence over the past decades. The three most prominent of these cognitive process domains are working memory (WM), processing speed (PS), and executive functions (EFs). In detail, working memory capacity (WMC) has been shown to correlate positively with intelligence, especially on a latent level (r = .50 – .90; Ackerman et al., 2005; Kyllonen & Christal, 1990; Oberauer et al., 2005). In contrast, the relationship between PS and intelligence is rather small for single task measures (r = -.50; Sheppard & Vernon, 2008) but increases considerably when PS is measured with a heterogeneous set of tasks ($|r| \approx .50$; Schmitz & Wilhelm, 2016; Schubert et al., 2017). Moreover, neural speed of higher order processing has shown a correlation with intelligence similar in size to WMC (r = -.89; Schubert et al., 2017).

The evidence for the relationship between EFs to intelligence is less consistent. First, it is still unclear in how far EFs can be described as a unitary attention regulation mechanism or as a set of diverse attentional processes (Friedman & Miyake, 2017; Karr et al., 2018; Miyake et al., 2000). Second, due to the separation of different EFs such as inhibition, updating, and shifting (Miyake et al., 2000) there are inconsistent results which EF specifically is related to intelligence. Furthermore, there is considerable heterogeneity in the measurement of EF ranging from average performance, or performance within specific conditions to difference measures between conditions. This is why estimates for the relationship of EFs with intelligence range from $|r|$ = .08 – .12 for shifting and inhibition measured with difference scores (Friedman et al., 2006; Wongupparaj, Kumari, & Morris, 2015) up to r = .44 – .74 for updating measured with average performance (Friedman et al., 2006; Wongupparaj et al., 2015).

Interestingly, these different cognitive processes and their relationship to intelligence have often been investigated separately or at least as independent predictors of intelligence (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008). One aim of my dissertation, in particular the fourth manuscript (summarized on p. 27f.; Frischkorn et al., submitted), was therefore to integrate these three different cognitive process domains (WMC, PS, and EFs) as predictors of intelligence.

4.1 Executive Functions: Bridging the gap between working memory capacity and processing speed as predictors of intelligence?

Previous studies that have linked WMC to intelligence raised the question which specific processes within WM are responsible for the strong relationship between WMC and intelligence. Some researchers have argued that attentional control processes unique to complex span tasks are the main reason for this strong correlation (Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin, & Conway, 1999), while others suggested that the capacity of short-term memory (STM) underlies the relationship of WM and intelligence (Colom et al., 2008, 2005; Shahabi, Abad, & Colom, 2014). In fact, current theories of WM propose that attentional processes may play a major role in maintenance of memory items regardless of concurrent processing (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Souza & Vergauwe, 2018). In sum, the same attentional processes may underlie both the capacity of STM and additional demands posed by complex span tasks (Barrouillet, Portrat, & Camos, 2011; Wilhelm, Hildebrandt, & Oberauer, 2013).

Similarly, early work has already discussed attention as one of the main reasons that PS is related to intelligence (Spearman, 1904). Especially in ageing research, attentional deficits in older adults have been related to changes in processing speed (Lester, Vatterott, & Vecera, 2018). In addition, many experimental paradigms used in attention research (e.g. the Stroop task, the Eriksen Flanker task or the Attention Network Test) measure the effects of their manipulations with response times (Eriksen & Eriksen, 1974; Fan, McCandliss, Sommer, Raz, & Posner, 2002; Stroop, 1935). In conclusion, attention seems to be closely related to PS as well.

Following these theoretical and empirical considerations, one candidate for attentional processes that are responsible for the strong relationship between WMC, PS and intelligence are executive functions (EFs). In reference to the multi-component model of WM (Baddeley & Hitch, 1974), EFs are often conceptualized as attentional control mechanisms within the central executive of working memory (Jurado & Rosselli, 2007).[1] One prominent model of EFs separates three different EFs (Friedman & Miyake, 2017; Miyake et al., 2000): inhibition, updating, and shifting. *Inhibition* refers to the ability to focus attention on relevant information while ignoring irrelevant information, *updating* describes the ability to remove outdated information from memory while encoding new information, and *shifting* represents the ability to effectively switch between different task instructions or mental sets.

Typically, EF tasks are designed so that specific conditions require the to be measured EF while others do not. For example, the Stroop- or Flanker-tasks (Eriksen & Eriksen, 1974; Stroop, 1935) – both proposed as measures of inhibition – contain congruent and incongruent conditions. In the incongruent condition, the irrelevant stimulus information indicates another response than the relevant information, while in the congruent condition both irrelevant and relevant stimulus information indicate the same

[1] The definition of EF is anything but unique. In some areas in psychology, especially developmental psychology, EF subsume all higher order cognitive processes dedicated to control or planning processes (Diamond, 2013), and other areas, mostly cognitive psychology, define EFs more restrictively as a set of specific attention control mechanisms (Miyake et al., 2000). Within the present thesis I use the latter definition of EFs to separate the different cognitive processes such as intelligence, WM, and EFs more clearly.

response. Thus, only the incongruent condition requires that the irrelevant stimulus information is ignored and attention to be focused on the relevant information to respond correctly. The effort of inhibiting the irrelevant information is thus captured by the difference between the two conditions. Accordingly, individual differences in inhibition should be represented by person specific differences in this effect.

Integrating EFs into the relationship of intelligence, WMC, and PS

(Manuscript 4)

The fourth manuscript of my dissertation (Frischkorn et al., submitted, see p. 116) addressed the question in how far EFs may explain the relationship of both WMC and PS with intelligence. In this study, a sample of $N = 101$ participants worked on three EF tasks – one for each of the three EF sensu Miyake et al. (2000) – while their EEG was recorded. Additionally, we assessed participants' intelligence, their WMC, and their PS. I then addressed the following two research questions, by implementing structural equation models (SEMs):

1. How much manipulation specific versus unspecific variance do reaction timess (RTs) in the different conditions of EF tasks contain?
2. How do individual differences in EFs fit into the relationship of WMC, PS, and intelligence?

The first question addresses in how far performance in EF task shows variations that can be attributed to the EF demands (i.e. the manipulation that supposedly requires the EF), and how much variance can be attributed to processes that are not specific for certain conditions and thus represent variations in more general cognitive processes. Building on these results, the second question then aimed at integrating the general and manipulation specific performance parts into the relationships between intelligence, WMC, and PS.

For the first research question, I separated the variance specific to the different conditions in the EF tasks from general variance across conditions for both behavioral RTs and latencies of the N1 and P3 component of the event-related potential (ERP) with bi-factor models. This bi-factor approach essentially computes a factor for general performance across all conditions, and latent difference scores for the specific manipulation in comparison to the general performance. Thus, the variance of the manipulation-specific factors (i.e. latent difference factors) represents individual differences in the EFs tapped by the respective tasks.

Results from the fourth manuscript indicated that the general factors of the bi-factor models captured on average 68% of variance in manifest indicators (both behavioral RTs and ERP latencies), while each of the manipulation specific (i.e. latent difference) factors captured on average only 14% of variance. This explains both the low reliability of difference measures in experimental paradigms (Hedge, Powell, & Sumner, 2018), and the small and inconsistent relationships of inhibition and shifting that are usually measured with difference scores to intelligence (Friedman et al., 2006; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Wongupparaj et al., 2015).

I then joined the bi-factor models from the three EF tasks and estimated the relationship of general and manipulation-specific variance from the EF tasks with intelligence, WMC, and PS. Interestingly, only the general factors from the bi-factor models correlated with the three covariates. Specifically, behavioral RTs were strongly related (r = .77) to PS operationalized with two elementary cognitive tasks, and showed consistent negative relationships with both intelligence (r = -.55), and WMC (r = -.49). This pattern, resembled the typical correlations of PS measured by elementary cognitive tasks (ECTs) with intelligence and WMC (Kyllonen & Christal, 1990). As WMC was strongly correlated with intelligence, I simplified the model by (1) joining all reaction time measures (i.e. RTs in both EFs tasks and ECTs) in one general processing speed factor and (2) joining intelligence and WMC in one factor for *higher cognitive abilities*. These two factors were then negatively correlated and shared approximately 30% of variance. Moreover, none of the manipulation-specific factors for behavioral RT showed any correlations with intelligence, WMC, or PS.

ERP latencies showed less consistent correlations with the three covariates. In detail, neither general nor manipulation-specific factors of N1 latencies in the three EF tasks showed any significant relationship with intelligence, WMC, or PS. The general factor for P3 latencies in the three EF tasks was moderately related to PS in ECTs (r = .40), and showed a small correlation with WMC (r = -.28). However, it was not related to intelligence. Similar to N1 latencies, the manipulation-specific factors for P3 latencies did not show any significant relationships with intelligence, WMC, or PS.

Overall, the results of the fourth manuscript thus conveyed two important points: (1) Performance within EF tasks mainly reflects general individual differences, and (2) only this general variance showed relationships with intelligence, WMC, and PS. The results from the fourth manuscript (Frischkorn et al., submitted) support previous evidence on the relationships of WMC and PS with intelligence (Ackerman et al., 2005; Jensen, 2006; Kyllonen & Christal, 1990; Schmitz & Wilhelm, 2016; Schubert et al., 2017). However, a comprehensive theoretical account for the inter-relations between these three constructs by EFs could not be established. In fact, attention regulation mechanisms such as EF, that were supposed to be related to both WMC and PS (Engle et al., 1999; Kane & Engle, 2003), did not show any relationships with intelligence, or WMC and PS.

Revisiting the difference between correlational and causal relationships between intelligence, WMC, and PS

It is important to note that recent empirical results suggest that PS may not causally underlie individual differences in intelligence (Schubert, Hagemann, et al., 2018). As described in section 3.2 (p.21ff.), a pharmacological manipulation of PS, did only affect RTs and ERP latency but not intelligence test performance. As there is nonetheless a consistent relationship of PS with intelligence and WMC (Schmitz & Wilhelm, 2016; Schubert et al., 2015; Sheppard & Vernon, 2008) there has to be at least one third variable that is responsible for this relationship.

Schubert, Hagemann, et al. (2018) argued that structural brain properties might explain the relationship between processing speed and general intelligence. For instance, organizational properties of the salience network (Menon & Uddin, 2010; Soltani &

Knight, 2000), as well as individual differences in nodal efficiency across brain regions (e.g. the right anterior insula and the dorsal anterior cingulate cortex) have shown correlations with intelligence (Hilger, Ekman, Fiebach, & Basten, 2017). As EFs could not explain the inter-relations between intelligence, WMC, and PS, the structural architecture of the neuro-cognitive system could potentially bridge the gap between WMC and PS as basic cognitive processes of intelligence.

4.2 Bindings: A new account for the relationship between working memory capacity, processing speed and intelligence?

In an attempt to acknowledge the structural properties of the brain, the interference theory of working memory (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Oberauer & Lin, 2017) assumes that one critical process in working memory is binding content information (e.g. words or colors) to arbitrary context information (e.g. serial or visuo-spatial position; Oberauer & Kliegl, 2006; Oberauer & Lin, 2017; Wilhelm et al., 2013). Within this theory, that has often been formalized as a cognitive model, these types of information are typically represented by a distributed activation within a connectionist network model (for an illustration see Figure 4.1, p. 31). In this model a specific activation pattern of context *neurons* is bound to a specific pattern of content *neurons*. On a neural level, this might be reflected in associations between brain regions that code context and content information respectively. These associations may well be related to individual differences in the structural organization of the brain and thus underlie the relationship between processing speed and intelligence, as hypothesized by Schubert, Hagemann, et al. (2018).

This theoretical idea gains further support by studies that have argued that basic reaction time tasks incorporate bindings as well (Meiran & Shahar, 2018; Wilhelm & Oberauer, 2006). In detail, reaction time tasks require participants to represent stimulus-response bindings. These stimulus-response bindings are often very intuitive,

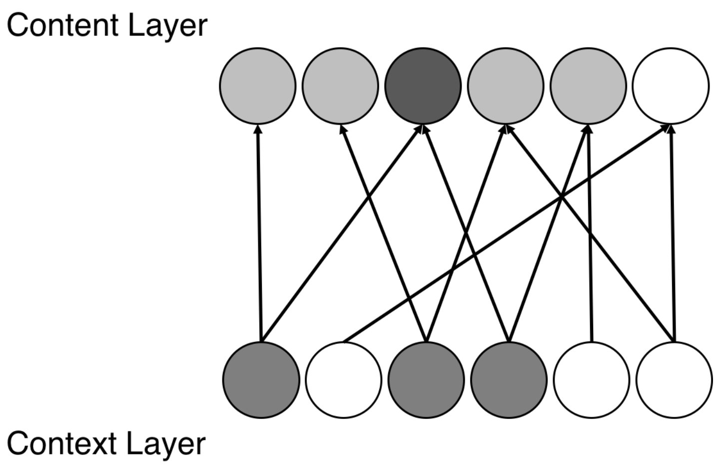

Figure 4.1: Graphical illustration of the connectionist network assumed by the interference theory of working memory. A specific context cue is represented by a distributed activation of specific nodes on the context layer. The bindings between these nodes and the content layer result in an activation pattern on the content layer. This activation pattern ideally should correspond to the distributed code of a specific semantic representation.

like pressing the left response key when an arrow pointing left is shown and the right key for an arrow pointing right. However, if stimulus-response bindings are arbitrary (i.e. the follow no readily accessible rule) rather than intuitive, the correlation between RTs and both working memory and general intelligence increased. Maintaining stimulus-response bindings - especially when arbitrary - thus seems to be a processes that is required in both measures of processing speed and working memory capacity. Thus, cognitive processes related to the formation and maintenance of bindings may explain why WMC and PS are related.

Interestingly, the P3 component of the ERP has been associated with updating of stimulus-response bindings as well (Verleger, 1997, 2010; Verleger, Metzner, Ouyang, Śmigasiewicz, & Zhou, 2014). Conceptually, one interpretation of the P3 component proposes that it reflects aspects of the decision process that are related to response selection processes (Hillyard & Kutas, 1983; Kelly & O'Connell, 2013). This response selection processes might rely on stimulus-response bindings that are necessary in basic reaction time tasks and more complex WM tasks.

Thus, the missing link between processing speed, working memory and general intelligence, could be that all these processes rely on bindings in some form. In sum, individual differences in the ability to form and maintain these bindings might be related to the structural organization of the brain. It is highly likely, that individual differences in this structural organization underlie numerous behavioral and neural measures of cognitive processes (e.g. ERP latencies or RTs) that have been linked to intelligence.

It is important to note, that important phenomena such as the worst performance rule (WPR) have been linked to bindings as well (McVay & Kane, 2012; Schmiedek et al., 2007). For example, Schmiedek et al. (2007) argued that the WPR – often associated with attentional lapses (Unsworth et al., 2010) – might also be explained by variations in the ability to maintain stimulus-response bindings. In contrast to the attentional lapses account for the WPR, Schmiedek et al. (2007) suggest that the variation of stimulus-response bindings across trials does not correspond to attentional lapses, but does reflect individual differences in working memory processes.

How to test whether bindings may explain individual differences in intelligence

So far, bindings have mostly served as an theoretical account for results concerning the relationship between intelligence, WMC, and PS. Future research should now focus on empirically examining this theoretical account. In detail, future studies could exploit the formalization of the interference theory of WM that is based on bindings (Oberauer, 2013; Oberauer & Kliegl, 2006), and relate binding parameters from cognitive measurement models (Oberauer & Lewandowsky, 2018) to individual differences in intelligence. This would (1) profit from the clear translation of the latent binding process into observed behavior within the cognitive model and (2) provide evidence in how far individual differences in the ability to form and maintain bindings are related to intelligence.

In addition to that, a cross-validation of the binding parameters from cognitive models with neuro-physiological measures that capture network properties of the brain could further strengthen the parameter interpretation. Such neuro-physiological measures could be frequency and cross-frequency coupling in the EEG (Jirsa & Müller, 2013; Palva & Palva, 2012) or structural or functional connectivity in the fMRI (Honey et al., 2009; van den Heuvel & Hulshoff Pol, 2010). An important advantage of this cross-validation would be that it can further specify whether the binding parameters from cognitive models are related to connectivity between specific brain regions or across the whole brain. In the light of WM theories, it is reasonable to assume that specifically fronto-parietal connections but not for example fronto-temporal connections should be related to bindings in WM.

Ultimately, future research should then link both these domains, cognitive model parameters and neuro-physiological measures, to individual differences in intelligence. This would incorporate numerous benefits that were pointed out within the present thesis: (1) adequate representations of cognitive processes by both cognitive models and neuro-physiological measures and (2) an integrative account for individual differences in intelligence.

Altogether, the idea of bindings is an interesting perspective on the inter-relations between WMC, PS, and intelligence. However, bindings have been linked to both cognitive domains that show strong relationships with intelligence (e.g. WMC and PS; Meiran & Shahar, 2018; Wilhelm et al., 2013; Wilhelm & Oberauer, 2006) as well as domains that seem to be unrelated to intelligence (e.g. EF; Oberauer, 2005). In this, bindings do not yet provide a exhaustive account for individual differences in intelligence. Thus, future research combining parameters of cognitive models and neuro-physiological measures could aid to specify which kind of bindings are associated with intelligence and which are not.

5 Summary and Conclusion

Starting with the question *What are the basic cognitive processes and intelligence?*, my dissertation aimed to contribute to three essential aspects (see Introduction, p. 2):

1. How can we measures individual differences in cognitive processes adequately?
2. How can we test which cognitive processes explain individual differences in intelligence?
3. How can we integrate different cognitive processes related to intelligence?

These three aspects tapped into core problems within psychology. First, the measurement of psychological traits, second the quest for causal explanations of psychological phenomena and individual differences, and third a comprehensive account for the interrelations between psychological constructs.

With respect to the first point, I demonstrated that cognitive models specify how latent cognitive processes are linked to observed behavior and that neuro-physiological measures provide an ontological reference for parameters from cognitive models. Moreover, I showed that the drift rate from the drift-diffusion model (DDM) exhibits trait-like properties similar to intelligence. Regarding the second point, I conceptualized the worst performance rule (WPR) as moderation and, based on this conceptualization, developed a new parameterization for the worst performance rule. This parameterization has already been adopted to test theoretical explanations for WPR (Löffler, 2018) and gain meta-analytic evidence on the WPR (Schubert, 2018). Finally, concerning the third point, I presented empirical results indicating that attention control mechanisms such as executive functions cannot explain why working memory capacity and processing speed are inter-related predictors of intelligence.

These contributions addressed fundamental problems in intelligence research: On the one hand, a lack of theoretical foundation of intelligence and its measurement, and on the other hand, a lack of results with respect to the connection and integration of basic cognitive processes with individual differences in intelligence. Although the results of my dissertation discarded attention regulation mechanisms such as executive functions as an integration for cognitive processing domains such as working memory capacity and processing speed that are related to intelligence, the different studies of my dissertation convey how future research can profit from integrating theoretically founded measures for cognitive processes with experimental and correlational approaches to individual differences in intelligence. Especially the joint modeling of cognitive model parameters and neuro-cognitive process parameters can provide exciting new insights what the basic cognitive processes of intelligence are.

Gaining a more comprehensive knowledge on the basic cognitive processes of intelligence may provide numerous possibilities. First, this is an important step towards developing a comprehensive process-oriented theory of intelligence that is long due. Additionally, such insights can provide specific suggestions how to enhance the development of cognitive abilities. And finally, such knowledge could bring us one step closer to a solution for the enigmatic positive manifold of cognitive abilities and a better understanding of the intriguing trait of mankind that we call *intelligence*.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin*, *131*, 30-60. doi: 10/drqmhx

Adams, H. F. (1936). Validity, reliability, and objectivity. *Psychological Monographs*, *47*, 329-350. doi: 10/frsdmx

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

Baddeley, A., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, p. 47-89). Academic Press. doi: 10.1016/S0079-7421(08)60452-1

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*, 175-192. doi: 10/bq7rsv

Bazana, P. G., & Stelmack, R. M. (2002). Intelligence and information processing during an auditory discrimination task with backward masking: An event-related potential analysis. *Journal of Personality and Social Psychology*, *83*, 998-1008.

Binet, A., & Simon, T. (1904). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'année psychologique*, *11*, 191-244. doi: 10/br3g44

Binet, A., & Simon, T. (1911). *A method of measuring the development of the intelligence of young children*. Lincoln, Ill. : The Courier Company.

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence*, *65*, 60-66. doi: 10/gckfnk

Boehm, U., Marsman, M., Matzke, D., & Wagenmakers, E.-J. (2018). On the importance of avoiding shortcuts in applying cognitive models to hierarchical data. *Behavior Research Methods*, *50*, 1614-1631. doi: 10/gd6dx2

Boring, E. G. (1923). Intelligence as the Tests Test It. *New Republic*, 35–37.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. New York, NY, US: Cambridge University Press. doi: 10.1017/CBO9780511490026

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, *6*, 25 - 53. doi: 10.1080/15366360802035497

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, *111*, 1061 - 1071. doi: 10.1037/0033-295X.111.4.1061

Box, G. E. P. (1976). Science and Statistics. *Journal of the American Statistical Association*, *71*, 791-799. doi: 10.1080/01621459.1976.10480949

Boyle, R. (1662). *A defence of the doctrine touching the spring and weight of the air, propos'd by Mr. R. Boyle in his New Physico-Mechanical Experiment... by the author of those experiments*. London: Printed by F. G. for Thomas Robinson.

Brennan, R. L. (2001). *Generalizability theory*. New York; Berlin; Heidelberg: Springer.

Burt, C. L. (1922). *Mental and scholastic tests*. London, P.S. King and son, ltd.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105. doi: 10/ds75wg

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*, *37*, 404-431.

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

Cattell, R. B. (1946). Personality structure and measurement I The operational determination of trait unities. *British Journal of Psychology*, *36*, 88-103.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed ed.). Hillsdale, N.J: L. Erlbaum Associates.

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*, 584-606. doi: 10/dn58z8

Colom, R., Flores-Mendoza, C., Quiroga, M. A., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, *39*, 1005-1014. doi: 10/fvckcg

Colom, R., Jung, R. E., & Haier, R. J. (2007). General intelligence and memory span: Evidence for a common neuroanatomic framework. *Cognitive Neuropsychology*, *24*, 867-878. doi: 10/bn8rqz

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163-183. doi: 10/frs9t8

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547 - 552. doi: 10.1016/J.TICS.2003.10.005

Coyle, T. R. (2003a). IQ, the worst performance rule, and Spearman's law: A reanalysis and extension. *Intelligence*, *31*, 473-489.

Coyle, T. R. (2003b). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*, 567-587. doi: 10/d4n5xx

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, *16*, 297-334. doi: 10/cc5

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684. doi: 10/b9h9gg

Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, *64*, 391-418. doi: 10/bwg2pm

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. Oxford: Irvington.

Danner, D., Hagemann, D., & Fiedler, K. (2015). Mediation analysis with structural equation models: Combining theory, design, and statistics. *European Journal of Social Psychology*, *45*, 460-481. doi: 10/gd8b7d

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, *39*, 323-334. doi: 10/c8vrxd

Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R., & Starr, J. M. (2000). The Stability of Individual Differences in Mental Ability from Childhood to Old Age:

Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49-55. doi: 10/ftzt7s

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135-168. doi: 10/b2m2

Dutilh, G., Vandekerckhove, J., Ly, A., Matzke, D., Pedroni, A., Frey, R., ... Wagenmakers, E.-J. (2017). A test of the diffusion model explanation for the worst performance rule using preregistration and blinding. *Attention, Perception, & Psychophysics*, *79*, 713-725. doi: 10/f962n6

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241-261. doi: 10/cwn8pj

Endler, N. S., & Magnusson, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin*, *83*, 956-974. doi: 10/cb6k3x

Engle, R. W., & Kane, M. J. (2003). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In *Psychology of Learning and Motivation* (Vol. 44, p. 145-199). Elsevier. doi: 10.1016/S0079-7421(03)44005-X

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of experimental psychology. General*, *128*, 309–331. doi: 10.1037/0096-3445.128.3.309

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143-149. doi: 10/d23822

Eysenck, H. J. (1967). *The Biological Basis of Personality*. New York, NY, USA: Routledge.

Eysenck, H. J. (Ed.). (1981). *A Model for Personality*. Berlin Heidelberg: Springer-Verlag.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*, 340-347. doi: 10.1162/089892902317361886

Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. Cambridge University Press. doi: 10.1017/CBO9781316272503

Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences*, *15*, 272-279. doi: 10/c7sbd2

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186-204. doi: 10/f9sqwm

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*, 172-179. doi: 10/bmb68s

Frischkorn, G. T., & Schubert, A.-L. (2018). Cognitive Models in Intelligence Research: Advantages and Recommendations for Their Application. *Journal of Intelligence*, *6*, 34. doi: 10/gd3vqn

Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (submitted). Processing Speed, Working Memory, and Executive Functions: Independent or inter-related predictors of general intelligence? *Intelligence*.

Frischkorn, G. T., Schubert, A.-L., Neubauer, A., & Hagemann, D. (2016). The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis.

Journal of Intelligence, *4*, 9. doi: 10/gd3vsz

Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence*, *52*, 71-79. doi: 10/f7rcdj

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*, 255-282. doi: 10/b45tjm

Hagemann, D., Schubert, A.-L., & Frischkorn, G. T. (2016). Intelligenz ist wesentlich mehr als die Kapazität des Arbeitsgedächtnisses: Eine experimentelle Analyse. In *50. Konferenz der Deutschen Gesellschaft für Psychologie (DGPs).* Leipzig, Germany.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*, 1166-1186. doi: 10/gddfm4

Hilger, K., Ekman, M., Fiebach, C. J., & Basten, U. (2017). Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence*, *60*, 10-25. doi: 10/f9jg3q

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of Cognitive Processing. *Annual Review of Psychology*, *34*, 33-61. doi: 10/b9mng9

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., & Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, *106*, 2035-2040. doi: 10/dtm9fb

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*, 253 - 270. doi: 10.1037/H0023816

Hunt, E., & Jaeggi, M. (2013). Challenges for Research on Intelligence. *Journal of Intelligence*, *1*, 36-54. doi: 10.3390/JINTELLIGENCE1010036

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*, 567-578. doi: 10/d98rr5

Jensen, A. R. (1998). *The g Factor - The Science of Mental Ability.* Westport, CT, USA: Praeger Publishers.

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

Jirsa, V., & Müller, V. (2013). Cross-frequency coupling in real and virtual brain networks. *Frontiers in Computational Neuroscience*, *7*. doi: 10/gfkfbp

Johnson, W. (2013). Whither Intelligence Research? *Journal of Intelligence*, *1*, 25-35. doi: 10.3390/JINTELLIGENCE1010025

Johnson, W., Bouchard Jr., T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, *32*, 95 - 107. doi: 10.1016/S0160-2896(03)00062-X

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*, 81-95. doi: 10/c3235x

Jurado, M. B., & Rosselli, M. (2007). The Elusive Nature of Executive Functions: A Review of our Current Understanding. *Neuropsychology Review*, *17*, 213-233. doi: 10/cw6rkt

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in Working Memory Capcacity as Variation in Executive Attention and Control. In A. R. A. Conway & M. Jarrold (Eds.), *Variation in Working Memory* (p. 21-46).

Oxford University Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70. doi: 10/fwhz92

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related onstructs: Comment on Ackerman, Beier and Boyle (2005). *Psychological Bulletin*, *131*, 66-71. doi: 10.1037/0033-2909.131.1.66

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*, 1017-1048. doi: 10/gd3vrj

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*. doi: 10/gd3vsx

Kelly, S. P., & O'Connell, R. G. (2013). Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. *Journal of Neuroscience*, *33*, 19434-19441. doi: 10/f5ktrt

Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, *27*, 151-177. doi: 10/gd3vr6

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389-433. doi: 10/bxmdv4

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

Lester, B. D., Vatterott, D. B., & Vecera, S. P. (2018). Attention and Processing Speed. In *The Wiley Handbook on the Aging Mind and Brain* (p. 337-361). John Wiley & Sons, Ltd. doi: 10.1002/9781118772034.ch17

Löffler, C. (2018). *Exekutive Aufmerksamkeit als Erklärungsansatz der Worst Performance Rule*. Heidelberg: Masterarbeit, Universität Heidelberg, 2018.

Löffler, C., Schubert, A.-L., Frischkorn, G. T., Rummel, J., & Hagemann, D. (2018). Die Rolle exekutiver Aufmerksamkeit im Zusammenhang zwischen Informationsverarbeitungsgeschwindigkeit und Intelligenz.. doi: None

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

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT Press.

Luck, S. J., & Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.001.0001

Martinez, M. E. (2013). *Future Bright: A Transforming Vision of Human Intelligence*. OUP USA.

McGarry-Roberts, P. A., Stelmack, R. M., & Campbell, K. B. (1992). Intelligence, reaction time, and event-related potentials. *Intelligence*, *16*, 289-313. doi: 10/fcrtwf

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In *Contemporary Intellectual Assessment: Theories, Tests,* *and Issues.*

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1-10. doi: 10/cg6qpg

McVay, J. C., & Kane, M. J. (2012). Drifting From Slow to "D'oh!": Working Memory Capacity and Mind Wandering Predict Extreme Reaction Times and Executive Control Errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 525-549. doi: 10/c945qk

Meiran, N., & Shahar, N. (2018). Working memory involvement in reaction time and its contribution to fluid intelligence: An examination of individual differences in reaction-time distributions. *Intelligence*, *69*, 176-185. doi: 10/gdxqtm

Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods*, *1*, 293-299. doi: 10/brn8pm

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*, 655-667. doi: 10/fmz3ch

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*, 49-100. doi: 10/bkksp2

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621-640. doi: 10/cp4w8m

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing Cause from Effect Using Observational Data: Methods and Benchmarks. *Journal of Machine Learning Research*, *17*, 1-102.

Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, *99*, 166-180. doi: 10/dx2qjg

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., ... Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77-101. doi: 10/d987m8

Oberauer, K. (2005). Binding and Inhibition in Working Memory: Individual and Age Differences in Short-Term Recognition. *Journal of Experimental Psychology: General*, *134*, 368-387. doi: 10/cppm38

Oberauer, K. (2013). The focus of attention in working memory - from metaphors to mechanisms. *Frontiers in Human Neuroscience*, *7*, 1 - 16.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*, 758-799. doi: 10.1037/BUL0000046

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of memory and language*, *55*, 601-626. doi: 10.1016/J.JML.2006.08.009

Oberauer, K., & Lewandowsky, S. (2018). Simple Measurement Models for Complex Working-Memory Tasks. *PsyArXiv*.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779–819. doi: 10.3758/S13423-012-0272-4

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory.

Psychological Review, *124*, 21 - 59. doi: 10.1037/REV0000044

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working Memory and Intelligence–Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61-65. doi: 10/cfjxgw

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature Neuroscience*, *15*, 1729-1735. doi: 10/f4dv4x

Palva, S., & Palva, J. M. (2012). Discovering oscillatory interaction networks with M/EEG: Challenges and breakthroughs. *Trends in Cognitive Sciences*, *16*, 219-230. doi: 10/gfkfb9

Peak, H., & Boring, E. G. (1926). The factor of speed in intelligence. *Journal of Experimental Psychology*, *9*, 71-94. doi: 10/dbj88z

Rammsayer, T. H., & Troche, S. J. (2016). Validity of the Worst Performance Rule as a Function of Task Complexity and Psychometric g: On the Crucial Role of g Saturation. *Journal of Intelligence*, *4*, 5. doi: 10.3390/JINTELLIGENCE4010005

Rao, K. V., & Baddeley, A. (2013). Raven's matrices and working memory: A dual-task approach. *The Quarterly Journal of Experimental Psychology*, *66*, 1881-1887. doi: 10/gd59vd

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108. doi: 10/fjwm2f

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, *106*, 6539-6544. doi: 10/ch22rk

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence*, *36*, 10-17. doi: 10/fp5rjz

Ratcliff, R., Sederberg, P. B., Smith, T. A., & Childers, R. (2016). A single trial analysis of EEG in recognition memory: Tracking the neural correlates of memory strength. *Neuropsychologia*, *93*, 128-141. doi: 10/f9hcgn

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127-157. doi: 10/fndcr6

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In R. S. McCallum (Ed.), *Handbook of Nonverbal Assessment* (p. 223-237). Boston, MA: Springer US. doi: 10.1007/978-1-4615-0153-4_11

Rhemtulla, M., van Bork, R., & Borsboom, D. (2015). Calling Models With Causal Indicators "Measurement Models" Implies More Than They Can Deliver. *Measurement: Interdisciplinary Research and Perspectives*, *13*, 59-62. doi: 10/gfkfdg

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales* (5th Edition ed.). Itasca, IL: Riverside Publishing.

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*, 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*. doi: 10/gd3vrp

Schubert, A.-L. (2018). A meta-analysis of the worst performance rule. *PsyArXiv*. doi: 10/gfhz63

Schubert, A.-L., Frischkorn, G. T., Hagemann, D., & Voss, A. (2016). Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence*, *4*, 7. doi: 10.3390/JINTELLIGENCE4030007

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*, 1498-1512. doi: 10/gch83n

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: 10/gffjb9

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/f7mtv3

Schubert, A.-L., Nunez, M. D., Hagemann, D., & Vandekerckhove, J. (2018). Individual differences in cortical processing speed predict cognitive abilities: A model-based cognitive neuroscience account. *bioRxiv*, 374827. doi: 10.1101/374827

Shahabi, S. R., Abad, F. J., & Colom, R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence*, *44*, 134-141. doi: 10/f576d3

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*, 535 - 551. doi: 10.1016/J.PAID.2007.09.015

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563-575. doi: 10/fckhxw

Soltani, M., & Knight, R. T. (2000). Neural Origins of the P300. *Critical Reviews in Neurobiology*, *14*, 26. doi: 10/gdz7mv

Souza, A. S., & Vergauwe, E. (2018). Unravelling the intersections between consolidation, refreshing, and removal. *Annals of the New York Academy of Sciences*, *1424*, 5-7. doi: 10/gdxvb7

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*, 201-292. doi: 10/chp6wk

Spearman, C. (1927). *The Nature Of Intelligence And The Principles Of Cognition*. London: Macmillan.

Spearman, C. (1946). Theory of General Factor1. *British Journal of Psychology. General Section*, *36*, 117-131. doi: 10/bzt5rq

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*, 245-251. doi: 10/dkrqw7

Steyer, R., Schmitt, M., & Eid, M. (1999). Latent state–trait theory and research in personality and individual differences. *European Journal of Personality*, *13*, 389-408. doi: 10/dx7cvw

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643 - 662. doi: 10.1037/H0054651

Tabachnik, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Terman, L. M. (1916). *The Measurement Of Intelligence*. Houghton Mifflin Company.

Thurstone, L. (1938). *Primary mental abilities.* University of Chicago Press: Chicago.

Troche, S. J., Houlihan, M. E., Stelmack, R. M., & Rammsayer, T. H. (2009). Mental ability, P300, and mismatch negativity: Analysis of frequency and duration discrimination. *Intelligence*, *37*, 365-373. doi: 10/dm4nz7

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65-79. doi: 10/f9wfkt

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*, 111-122. doi: 10/b324jf

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392-406. doi: 10/bnbft5

v. Helmholtz, H. (1921). Zählen und Messen, erkenntnistheoretisch betrachtet. In H. v. Helmholtz, P. Hertz, & M. Schlick (Eds.), *Schriften zur Erkenntnistheorie* (p. 70-108). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-36342-3_3

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, *20*, 519-534. doi: 10/dgtvmc

van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). Intelligence Is What the Intelligence Test Measures. Seriously. *Journal of Intelligence*, *2*, 12-15. doi: 10/gchthn

van Ravenzwaaij, D., Provost, A., & Brown, S. D. (2017). A confirmatory approach for integrating neural and behavioral data into a single model. *Journal of Mathematical Psychology*, *76*, 131-141. doi: 10/f9wzhq

Verleger, R. (1997). On the utility of the P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131-156. doi: 10/fvv22s

Verleger, R. (2010). Popper and P300: Can the view ever be falsified that P3 latency is a specific indicator of stimulus evaluation? *Clinical Neurophysiology*, *121*, 1371-1372. doi: 10/fm38mj

Verleger, R., Metzner, M. F., Ouyang, G., Śmigasiewicz, K., & Zhou, C. (2014). Testing the stimulus-to-response bridging function of the oddball-P3 by delayed response signals and residue iteration decomposition (RIDE). *NeuroImage*, *100*, 271-280. doi: 10.1016/J.NEUROIMAGE.2014.06.036

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*, 1206–1220. doi: 10/d6fh2r

Wainer, H. (2000). The Centercept: An Estimable and Meaningful Regression Parameter. *Psychological Science*, *11*, 434-436. doi: 10.1111/1467-9280.00284

Wechsler, D. (1944). *The Measurement of Adult Intelligence*. The Williams & Wilkings Company.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. doi: 10.3389/FPSYG.2013.00433

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*, 18-50. doi: 10/dgsf8g

Wittmann, W. W. (1988). Multivariate Reliability Theory. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of Multivariate Experimental Psychology* (p. 505-560). Boston, MA: Springer US. doi: 10.1007/978-1-4613-0893-5_16

Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants.* (p. 77-108). Washington, DC: American Psychological Association. doi: 10.1037/10315-004

Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence*, *53*, 166-180. doi: 10.1016/J.INTELL.2015.10.007

List of Figures

List of Abbreviations

CFA confirmatory factor analysis. 7

CLVM cognitive latent variable model. 14

CS complex span. 25, 26

CTT classical test theory. 5, 18

DDM drift-diffusion model. 3, 10–12, 14, 21, 34

ECT elementary cognitive task. 12, 28, 29

EEG electro-encephalogram. 13, 14, 32

EF executive function. 24–30, 33, 35

ERP event-related potential. 8, 13–15, 28, 29, 31, 32

fMRI functional magnetic resonance imaging. 33

g general intelligence. 2, 7, 8, 10, 18, 20

PS processing speed. 17, 24–33, 35

RT reaction times. 11, 13, 27–29, 31, 32

SEM structural equation model. 5, 7, 17, 27

STM short-term memory. 25

WM working memory. 24–26, 31–33

WMC working memory capacity. 17, 24–33, 35

WPR worst performance rule. 3, 16–21, 32, 34

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 ______________________________

Datum, Unterschrift
Date, Signature ______________________________

Appendix

Contents of Appendix:

Journal of
Intelligence

MDPI

Article

Cognitive Models in Intelligence Research: Advantages and Recommendations for Their Application

Gidon T. Frischkorn *,† and Anna-Lena Schubert †

Institute of Psychology, Heidelberg University, Hauptstrasse 47-51, D-69117 Heidelberg, Germany; anna-lena.schubert@psychologie.uni-heidelberg.de
* Correspondence: gidon.frischkorn@psychologie.uni-heidelberg.de; Tel.: +49-6221-54-7746
† Both authors contributed equally to this work.

Received: 21 April 2018; Accepted: 9 July 2018; Published: 17 July 2018

check for updates

Abstract: Mathematical models of cognition measure individual differences in cognitive processes, such as processing speed, working memory capacity, and executive functions, that may underlie general intelligence. As such, cognitive models allow identifying associations between specific cognitive processes and tracking the effect of experimental interventions aimed at the enhancement of intelligence on mediating process parameters. Moreover, cognitive models provide an explicit theoretical formalization of theories regarding specific cognitive processes that may help in overcoming ambiguities in the interpretation of fuzzy verbal theories. In this paper, we give an overview of the advantages of cognitive modeling in intelligence research and present models in the domains of processing speed, working memory, and selective attention that may be of particular interest for intelligence research. Moreover, we provide guidelines for the application of cognitive models in intelligence research, including data collection, the evaluation of model fit, and statistical analyses.

Keywords: intelligence; cognitive modeling; methods; measurement; practical guidelines

1. Introduction

One of the greatest challenges in intelligence research is the identification of cognitive processes underlying cognitive abilities and the measurement of process parameters giving rise to individual differences in general intelligence [1]. Traditional as well as current theories of general intelligence either assume that intelligent behavior is the result of individual differences in various independent cognitive abilities [2–4], or that there is a hierarchical structure of cognitive abilities with a domain general and broad factor of general intelligence *g* that determines individual differences in cognitive abilities [5–9]. Theoretically and empirically the most discussed process parameters related to individual differences in general intelligence are the speed of information processing e.g., [9,10], the capacity of short-term memory e.g., [11], working memory e.g., [12–14] or secondary memory e.g., [15,16], and the efficiency of executive functions e.g., [4,17,18].

With respect to these theoretical and empirical considerations, there are three main goals to this process-oriented approach to intelligence research: First, understanding whether *one* or *several* cognitive processes give rise to individual differences in general intelligence will help to decide whether *g* should be conceived of as a single cognitive process, as suggested by Spearman's two-factor theory [5], or as an emerging phenomenon due to several independent or interacting cognitive processes, as suggested by sampling theories [4,19]. Second, a process-oriented approach aims to identify the mechanisms limiting or facilitating performance in certain cognitive processes by developing formal theories of the mechanisms constituting these processes. Third, such a process-oriented approach may ultimately lead to the development of formal theories of general intelligence by combining psychometric approaches and previous insights into the mechanisms of cognitive processes strongly related to general intelligence.

In empirical research, individual differences in these cognitive processes are usually measured by behavioral indicators such as response times and accuracies in tasks supposedly engaging one specific cognitive process. The behavioral performance in these tasks is then used to quantify the relationship of these cognitive processes to overall performance in intelligence tests. This approach presumes that a specific task provides a process-pure measure of a single cognitive process—an assumption that is often violated as most cognitive tasks do not measure one specific cognitive process, but rather a combination of several cognitive processes. For example, tasks measuring the efficiency of inhibitory processes such as the Stroop or Flanker task usually use reaction times as performance measures [20,21]. These reaction times arguably reflect not only the efficiency of inhibitory processes, but also basic information-processing speed. Another example is complex cognitive tasks such as complex span tasks measuring working memory capacity that not only require the storage of information in the face of processing, but may also rely on attentional control processes and speed of information processing [22,23]. In sum, typical measures for a specific cognitive process thus require additional cognitive processes beyond the cognitive process aimed to be measured.

Two approaches are typically pursued to overcome this problem. First, variance decomposition methods may be used to isolate the variance of one latent cognitive process parameter from the influence of other variables e.g., [11,12,17]. This method is feasible as long as there are "pure" measurements of the confounding cognitive processes available that can be controlled for. However, this approach may be resource- and time-consuming, as participants have to complete large test batteries including both measures of interest and of possible confounds.

A second approach to this measurement problem is to design experimental tasks that contain a baseline condition requiring the engagement of all confounding processes and an experimental condition that is equal to the baseline condition except for the insertion of one additional processing requirement of interest. Subtracting performance in the baseline condition from performance from the experimental condition is supposed to isolate the efficiency or speed of the added process [24]. However, it is questionable if the resulting difference scores only contain variance that can be attributed to the inserted process or if the insertion of additional processing demands may affect or interact with other task demands that are also reflected in the difference scores [25,26]. Moreover, the low between-subject variability and low reliability of difference scores in typical cognitive tasks renders the isolation of individual differences in experimental effects by means of difference scores virtually impossible [27,28].

In the present paper, we aim to demonstrate how mathematical models of cognition can be used to partially overcome these measurement problems by directly quantifying specific cognitive processes. Moreover, we will provide practical guidelines and recommendations for the use of cognitive models in intelligence research. While ultimately a formalization of specific theories of intelligence e.g., [3,4] would be desirable, these theories are still too general and abstractly formulated to allow the development of a formalized cognitive model of intelligence. As long as this is the case, the incorporation of mathematical models of the cognitive processes addressed in these theories provides a first necessary step towards a concrete formalized theory of intelligence. Therefore, the present manuscripts focuses on mathematical models of cognitive processes that are related to general intelligence or *g* rather than on cognitive models for general intelligence itself.

2. Advantages of Cognitive Modeling in Intelligence Research

2.1. Statistical Models

Although often not explicitly in mind, each measurement of a cognitive process and more generally any property of a person is based on a model. Most often we use statistical models, such as classical test theory or latent variable models for this measurement procedure [29]. These models typically assume that the measured and observed behavior is the compound of some *true* or *latent* property of a person and of an *error* of measurement [30–32]. Across repeated measurements of the same property, this results in a distribution of observations of which the average or expected value given a person (i.e., the arithmetic mean) is conceptualized as the best estimate of the *true* person property, while deviations from this value (i.e., the standard deviation) correspond to the amount of *error* or uncertainty in the measurement. Taken together, statistical models describe statistical properties of observed variables such as their mean and reliability (according to classical test theory), or the covariances among different variables (according to latent variable models).

Even though statistical models have proven to be very useful in the context of measurement, such models bear serious conceptual problems [29,33] and the selection of an adequate statistical model for measurement is anything but trivial. Apart from these general philosophical and epistemological problems of measurement with statistical models such as the ontological status of true-scores or latent variables and the adoption of a realist or constructionist perspective on science and measurement [29], all of these models have another serious shortcoming: Statistical models do not specify any psychological or cognitive processes underlying the *true* part of the measurement, but rather focus on separating *true* properties of a person from the *error* of measurement.

In response to this problem, it has been recommended to use more elaborate statistical models such as ex-Gaussian- or Wald- distributions for reaction times [34–36], and Binomial-distributions for accuracies or mental test scores [37,38]. Although these distributions correspond more closely to the empirical shape of the distributions of observed variables, the parameters of these distributions do not consistently resemble indicators of distinct cognitive processes, see [39]. More importantly, these models still only describe statistical characteristics of the observed variables and do not provide a theoretical account of the cognitive processes underlying the observed behavior. In sum, statistical models may be useful to quantify the amount of variance in a measurement attributable to the *true* personality trait (i.e., the reliability), however they do not allow any theoretically founded statements about the cognitive processes underlying the observed behavior or the latent personality trait.

2.2. Cognitive Models

Conversely, cognitive models may provide a mathematically-guided quantification of specific cognitive processes [40]. Specifically, cognitive models translate explicit verbal theories of cognitive processes in specific tasks into mathematical formulations of these theories. In this, the behavioral measures within a task are described as the result of different interacting processes or parameters of the model. The detailed interplay and interaction of these processes is specified within the formal architecture of the model and represents the assumptions the model makes with respect to a specific cognitive process. Thus, a cognitive model represents a formalized theory of a cognitive process that objectively states which parameters of the cognitive process affect differences in observed behavior across conditions or individuals. The adequacy and validity of this formalization can be evaluated by parameter recovery studies and by testing the selective effects of theoretically-guided experimental manipulations on model parameters [41].

Taken together, cognitive models provide several advantages over statistical models: (1) They provide falsifiable descriptions of the cognitive process underlying behavioral responses in a specific task; (2) Model parameters can be interpreted in an objective and formally described manner;

and (3) Model parameters can be used to quantify individual differences in specific cognitive processes based on the underlying model architecture.

3. Selecting Cognitive Models Suitable for Intelligence Research

Usually, cognitive models are used with two different aims: (1) A cognitive model aims to *formally describe* the cognitive processes underlying the observed behavior in a specific task and *explain* specific experimental effects observed within this task; (2) The parameters of a cognitive model estimated from the observed behavior in a task are used as *measures* for differences across individuals or experimental conditions. These measures quantify how far people or conditions differ with respect to a specific process of the cognitive model. Within the field of cognitive modeling, cognitive models serving the first aim are often described as explanatory cognitive models or cognitive process models, while cognitive models used with the second aim are often called cognitive measurement models [42]. Accordingly, any cognitive model can be considered both an explanatory cognitive model and a cognitive measurement model depending on the circumstances of its use. Nevertheless, cognitive models that are used to explain the observed behavior within a specific task (i.e., explanatory cognitive models) often differ from cognitive models that are used to quantify differences in their parameters across individuals or conditions (i.e., cognitive measurement models).

In detail, explanatory cognitive models aim to provide formal explanations for variations across experimental conditions in specific paradigms in terms of cognitive processes. These models formally describe the architecture of a cognitive process and focus on the interplay of different mechanisms that lead to specific experimental results. In contrast, cognitive measurement models typically decompose the observed behavior of a person into meaningful parameters of a latent cognitive process. Thus, instead of explaining differences across individuals or experimental conditions, cognitive measurement models are highly flexible tools that reflect these differences in variations of their estimated parameters (for a comparison of these two model types, see [43]). Often cognitive measurement models rely on a more elaborated explanatory cognitive model. However, there are many cognitive measurement models that have been developed independently of any explanatory cognitive model e.g., [44].

With respect to their application, cognitive models used to explain observed behavior, such as the SOB-CS [45], the slot-averaging model [43], or the interference model of visual working memory [42], often resemble highly elaborated model architectures that specify detailed formal models for a cognitive process. These models are often very complex and require high computing power to calculate predictions for a given set of parameters. In contrast, cognitive models used to measure differences of parameters across individuals or conditions, such as signal-detection theory [44], the two-high threshold model for recognition [46], or the drift-diffusion model [47], are mostly simplified descriptions of a cognitive process that can be generalized to a broad set of paradigms and observed variables. Beyond that, such cognitive measurement models are easy to use and parameters of cognitive measurement models can either be readily calculated from observed variables or estimated with adequate fitting procedures.

In intelligence research, the use of cognitive measurement models is far more widespread than the use of explanatory cognitive models. Although explanatory cognitive models provide a powerful tool for comparing different theories with respect to their predictions for experimental paradigms and manipulations, see [48]; their complexity and especially the lack of estimable parameters renders their application in intelligence research difficult. Still, results from explanatory cognitive models may provide the theoretical foundation for deciding for or against a specific cognitive measurement model.

Furthermore, there have been efforts to formulate explanatory models of intelligence test performance such as the Carpenter et al. [49] model for performance in the Raven matrices. In this model, Carpenter et al. [49] described different cognitive processes that are used while solving the Raven matrices. Some of these processes such as incremental encoding processes and rule induction for each matrix were used by all participants, while other processes such as the induction of abstract relations of the dynamic management of different goals in memory were specific to participants with above-average performance. Although this model provides a strong theoretical explanation for individual differences in Raven performance, its application remains limited.

Cognitive measurement models may instead provide person- and condition-specific parameters for distinct cognitive processes. These person-specific parameters can be easily used as measures of individual differences in specific aspects of cognitive processes, which can then be related to performance in intelligence tests. For instance, parameters of the drift-diffusion model, that will be introduced later, have been associated with performance in intelligence test or memory tasks [26,50–52]. In this, parameters from cognitive measurement models may thus provide insights on which cognitive processes are actually linked to intelligence.

While all cognitive models are deliberate simplifications of the cognitive processes within a task and rely on often critically discussed assumptions, there is actually no alternative to the use of a measurement model, may it be statistical or cognitive. While most research does not explicitly decide for a specific measurement model, by calculating the mean performance for a person in a task (as often done) they implicitly adopt a statistical measurement model that makes no explicit statements about the underlying cognitive processes of the measurement. It may even be argued that not explicitly deciding for a specific measurement model is practically similar to implicitly using the most simple cognitive model at hand: A model assuming that the observed variable directly represents the cognitive processes of interest. As already mentioned earlier, this assumption is almost always false. Therefore, we would argue that using explicit measurement models is always superior to equating observed variables with the cognitive process of interest.

To convey an idea of the benefits of the application of cognitive modeling in intelligence research, we will discuss three examples of cognitive models in the following sections. We selected different models describing cognitive processes of particular interest to intelligence research, such as decision making, working memory, and cognitive control, and demonstrate how they may be used to quantify individual differences in the respective cognitive processes. Please note that the three models described below differ in their breadth of application and in their former use as explanatory or measurement model. Following these examples, we then provide guidelines for choosing the appropriate model for a particular research question.

3.1. Different Cognitive Models of Interest for Intelligence Research

3.1.1. The Drift Diffusion Model of Binary Decision Making

The drift diffusion model (DDM) describes performance in two-alternative forced choice decisions tasks. The model assumes that evidence is accumulated in a random walk process until one of two decision thresholds is reached, the decision process is terminated, and a motor response (usually a key press) is initiated (see Figure 1 for an illustration; [47]). This evidence accumulation process can be described by a Wiener diffusion process that consists of a systematic component, the drift rate v, and normally distributed random noise with a mean of 0 and a variance of s^2 (this so-called diffusion constant s is usually fixed to a standardized value such as 0.1 or 1 for reasons of identifiability). The drift rate can be considered as a performance measure that directly quantifies the velocity of information uptake. In addition, the DDM quantifies the distance between decision thresholds as a measure of speed-accuracy trade-offs and decision cautiousness in the boundary separation as the parameter a, the starting point of evidence accumulation as the parameter z, and the time of non-decisional processes such as encoding and response preparation and execution as the parameter t_{er} or t_0. Beyond these basic parameters, intra-individual variability parameters have been added to the DDM (i.e., st_0, sv, and sz) to account for inter-trial variability within a person [47,53].

The validity of DDM parameters has been demonstrated both by parameter recovery studies [54] and by experimental validation studies [55–57]. Moreover, model parameters have been showing satisfying reliabilities estimated with test-retest correlations given sufficiently large trial numbers [58] and at least drift rates have been shown to exhibit trait-like properties [59]. Specifically, Schubert et al. [59] used latent-state trait models with additional method factors [60] to separate different variance sources across three different tasks and two measurement occasions. The results showed that the variance consistent across tasks and measurement occasions was largest for drift rates (on average 44%), while this variance was considerably lower for boundary separations and non-decision times (between 32 to 36%). Although drift rates captured this amount of variance that was consistent across tasks and measurement occasions best, single task estimates of drift rates were only moderately reliable ($Rel = 0.38 - 0.69$) and still contained considerable method specific variance (9 to 17%). Therefore, individual differences in drift rates should always be measured across different tasks if one is interested in individual differences in the underlying latent trait.

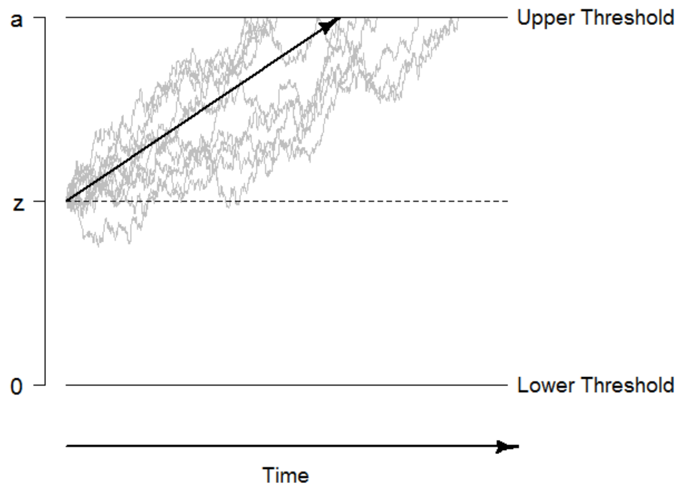

Figure 1. Graphical illustration of the drift-diffusion model. The decision process starts at the starting point z, and information is accumulated until the boundary a is reached. The systematic part of the accumulation process, the drift rate v, is illustrated with the black arrow. The non-decision time t_0 is not included in this figure.

Altogether, it is not surprising that the DDM is the most frequently used cognitive model in intelligence research. By mathematically identifying parameters quantifying the speed of information uptake (v), the decision cautiousness (a), and encoding and movement times (t_{er}), it renders complicated experimental setups that have been used to dissociate these elements of the decision process with little success unnecessary [61]. Several studies have reported positive associations between cognitive abilities and drift rates e.g., [26,50,52,62–64], whereas the other model parameters have been shown to be largely unrelated to fluid intelligence [26,52,64]. The application of the DDM to data sets is made fairly easy by user-friendly software such as *EZ* [65,66] and *fast-dm* [67].

The DDM is part of a larger family of evidence accumulation models that provide a general description of decision processes. Another member of this model family is the linear ballistic accumulator model (LBA; [68]), which presumes that a number of independent accumulators race towards a common response threshold. Hence, where the DDM can only be applied to data from two-choice reaction times tasks, the LBA can be applied to data from both two- and multiple-choice reaction time tasks. Another member of this model family is the leaky, competing accumulator model (LCA; [69]), which entails a number of stochastic accumulators that compete against each other via mutual inhibition to reach a decision threshold. Both models have not been applied in intelligence research yet, probably because they do not provide a single performance measure such as the drift rate of the DDM, as one drift parameter for each of the accumulators is estimated in LBA and LCA models, resulting in several drift rates.

3.1.2. The Time-Based Resource-Sharing Model of Working Memory

The time-based resource sharing (TBRS) model of working memory started out as a verbal theory explaining the performance in complex span tasks measuring working memory capacity [70,71], but has been extended to verbal and visual WM in general [72–74]. The TBRS model claims that processing and the maintenance of stored information rely on the same attentional resource in working memory. Because of this attentional bottleneck, only one of these two processes can be performed at a given time. In detail, the model assumes that information stored in working memory decays over time, unless this decay is counteracted by an attentional refreshing process or verbal rehearsal. Moreover, additional processing demands as imposed in complex span tasks shift attention towards these secondary processing tasks, resulting in the decay of items stored in working memory (see Figure 2 for an illustration). Altogether, working memory as conceptualized in the TBRS model continuously shares attentional resources between maintenance and processing in order to counteract decay of memory items and efficiently process information.

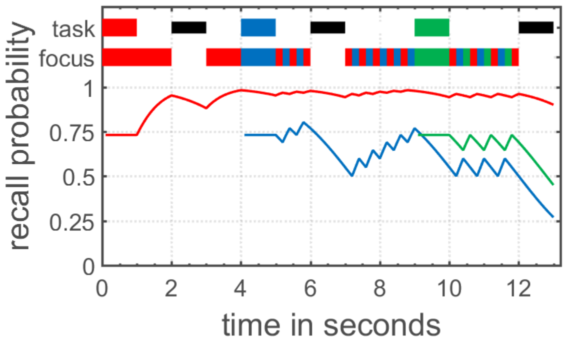

Figure 2. Visualization of the time-based resource sharing (TBRS) theory as implemented in the TBRS2 model by Gauvrit and Mathy [75]. At the top, the current task is displayed. A colored box represents a to- be-encoded memory item, a black box represents a distractor task, and a white box represents free time. Below, the focus of attention is shown. During free time, participants engage in refreshing of the already encoded memory item; during distractor tasks or encoding of other items, the already encoded memory items decay over time.

In recent years there have been formalizations of the TBRS model as an explanatory model [48] and as a simplified measurement model [75]. Such models may be of great interest for the field of intelligence research, not only because intelligence is strongly related to working memory [14,76,77], but because the field is still in debate about which specific cognitive processes within working memory, storage or executive processing, underlie its strong relationship with intelligence [11,12]. While the explanatory TBRS* model by Oberauer and Lewandowsky [48] is fairly complex and foremost an in-depth test for the experimental predictions of the TBRS theory, the TBRS2 implementation by Gauvrit and Mathy [75] provides a simplified version of the TBRS model and allows to estimate parameters that are directly linked to specific processes within the TBRS model. Such a model may provide person specific estimates of different processes in working memory, such as the encoding strength when an item is presented (i.e., the baseline β) or the speed of attentional refreshing (i.e., the refreshing rate r). These parameters may provide further information on which specific processes within working memory give rise to the strong relationship between working memory and intelligence.

As the mathematical implementations of the TBRS model have been developed only recently, there have not been any independent, systematic validation studies for the parameters of the model. Moreover, the psychometric properties of the model estimates (i.e., their reliability and validity) have not yet been assessed. Additionally, there is still a controversial debate in cognitive psychology whether decay actually is the core process limiting working memory capacity [78]. Although there are competing explanatory models of working memory questioning the role of decay as a limiting factor for working memory capacity [42,48], these models have not yet been translated into simple measurement models that allow estimating person-specific parameters of cognitive processes within working memory[1]. Until then, the TBRS2 model may provide a first step for including cognitive measurement models of working memory in intelligence research.

3.1.3. The Shrinking Spotlight Model of Selective Attention

The shrinking spotlight model of selective attention describes processing in the Eriksen flanker task, in which participants have to respond according to the orientation of a centrally presented target arrow while ignoring irrelevant arrows flanking the target stimulus [20,80]. The shrinking spotlight model is an extension of the drift diffusion model of sequential processing: It assumes that both target and flanker arrows provide perceptual evidence p for a particular response weighted by the amount of attention a allocated to each of these stimuli. The drift rate consists of the sum of weighted perceptual evidence across all stimuli at a given time. Over time, attention is assumed to zoom in on the central arrow, reflecting a narrowing of the focus of selective attention on the target stimulus. Thus, the target stimulus is weighted more strongly in comparison to the flanker stimuli and therefore affects the drift rate more strongly over time (see Figure 3). The initial width of attentional distribution is estimated in the attentional spotlight parameter sd_a, which reflects the standard deviation of a Gaussian distribution centered on the target stimulus, whereas the rate of attentional distribution reduction is estimated in the parameter r_d. In addition, the model also allows estimating the encoding and movement times in the t_{er} parameter, and the distance of symmetrical response thresholds from the starting point of evidence accumulation in the parameters A and $B = -A$.

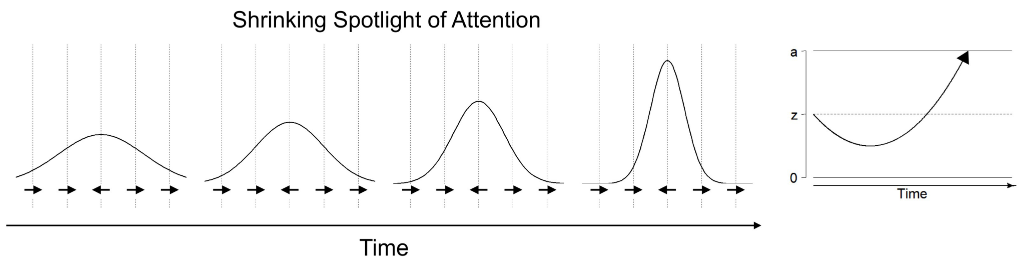

Figure 3. Illustration of the Shrinking Spotlight model for selective attention. The attentional focus narrows to the central arrow over time (**left part**). This results in a stronger weight of the critical information (i.e., the central stimulus) in the drift-rate of an associated diffusion process (**right part**).

The model has been shown to be able to account for data from a standard flanker task and experimental manipulations of task properties have been shown to specifically affect single model parameters [80]. Moreover, parameter recovery studies have shown that model parameters can be accurately recovered with as few as only 50 experimental trials [81]. However, simulation results have also shown that the model is not able to recover the attentional spotlight and the shrinking rate parameter accurately, because a wide initial spotlight with a high shrinking rate makes the same predictions as a narrow initial spotlight with a low shrinking rate [81]. Therefore, it has been recommended to calculate a composite measure of the duration of interference as the ratio of the two parameters, sd_a/d_r, to account for the trade-off during model estimation. Although there have not yet been any systematic analyses on the psychometric properties of parameter estimates, correlations of

[1] Oberauer and Lewandowsky [79] are working on an alternative measurement model that is more closely connected to interference models of working memory [42,45]. For a preprint, see: osf.io/vkhmu.

$r = 0.42 - 0.80$ between model parameters across different cue conditions of the Attention Network Test suggest at least moderate to good reliabilities, especially for the interference ratio with correlations of about $r \approx 0.80$ [82]. So far, the shrinking spotlight model has not yet been applied in intelligence research, but it would be promising to relate individual differences in the susceptibility to interference (as reflected in the interference parameter) to individual differences in intelligence test performance and working memory capacity to further explore the role of selective attention in cognitive abilities.

An alternative account of performance in the Eriksen flanker task is given by the dual-stage two phase model [83]. This model proposes two distinct processing stages: In the first processing stage, evidence accumulation is affected both by evidence accumulation towards the response associated with the target stimulus and by evidence accumulation towards the response associated with the flanker stimuli. At the same time, an attention-driven parallel evidence accumulation process selects a single stimulus for further processing. If this stimulus selection process terminates before response selection is finished, response selection enters a second stage with the drift rate being solely determined by the selected stimulus. As of yet, model comparison studies have not yet decided which of the two models provides the best account of selective attention phenomena [80,81,83,84]. Both models can be fit to data and subsequently be compared using the R package *flankr* [85].

3.2. Guidelines for Model Selection

When deciding which cognitive model to use for a specific research question, there are some conceptual and practical issues to be considered in order to select the appropriate model: First of all, the research question has to be specified. Second, the cognitive processes of interest that are to be related to general intelligence for this research question have to be identified. Third, an appropriate model providing a description of these cognitive processes has to be chosen. During this step, theoretical reasons for choosing one model over its alternatives should be considered. Fourth, experimental tasks congruent with the assumptions of the selected model should be selected to allow the valid estimation of model parameters. For an illustration of these decision steps see the upper part of Figure 4 (p. 10).

In general, discussing these issues during project planning aims to strengthen two important points for the conclusions from the modeling results. On the one hand, researchers should clarify which specific cognitive processes they are interested in and select a cognitive model accordingly. On the other hand, researchers should maximize the fit between the measurement or operationalization of a specific cognitive process (i.e., the task used) and the selected cognitive model.

For example, a group of researchers might be interested in which cognitive processes in simple decision tasks are related to intelligence. Such tasks may require participants to decide whether a number is odd or even, or whether a letter is a vowel or consonant. They decide to use the drift diffusion model to quantify the different cognitive processes associated with binary decision making. However, one of these tasks has an additional switching demand, requiring participants to switch between the number and the letter decision (for an example, see [86]). Because this task is a binary decision task, the drift-diffusion model may still provide suitable estimates for the cognitive processes in such a task [87,88]. However, this task arguably requires more than one decision: On the one hand the decision which task is to be carried out, and on the other hand the decision corresponding to the task. Thus, this task does not fully fit the conceptualization of the drift-diffusion model as there may not be a single decision process but two. Therefore, researchers should either think about using a different task that has a better fit to the basic assumptions of the drift-diffusion model or search for an alternative model that better fits the task they want to use.

This example reiterates the importance of an explicit and critical decision for a specific cognitive measurement model with respect to the measurement and operationalization that has already been pointed out before. As the developers of cognitive models often suggest a specific task suitable for parameter estimation e.g., [75], the initial model publication is usually a good starting point for finding prototypical tasks that match the model assumptions. For popular cognitive models such as the diffusion model there are review articles summarizing studies in which the diffusion model was successfully applied to data from several different tasks [55]. Although some of these prototypical tasks may not provide the most suitable measures for a specific research question, they nevertheless constitute a meaningful starting point.

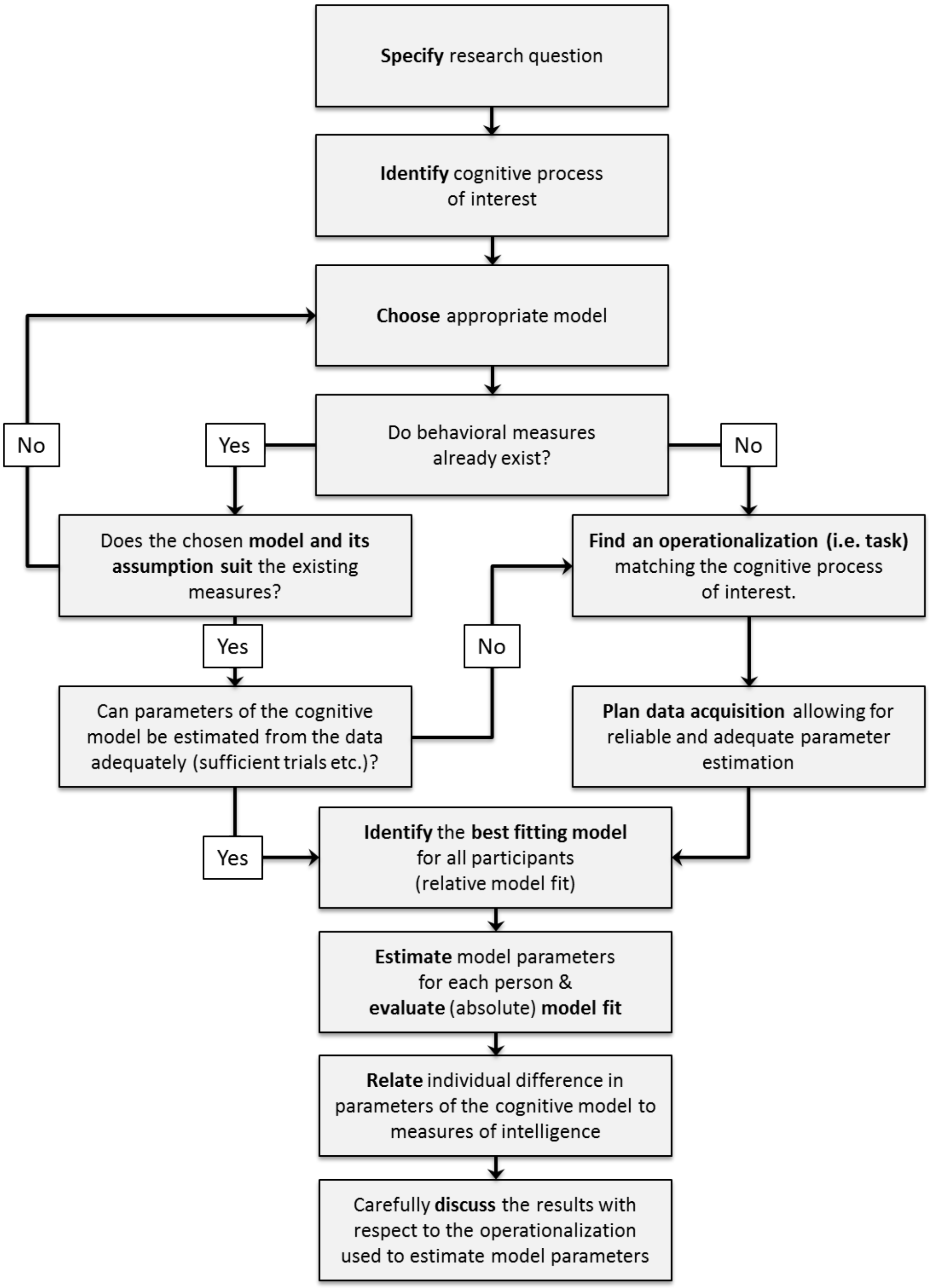

Figure 4. Flowchart illustrating the different planning and decision steps when using cognitive models in intelligence research.

4. Guidelines for Model Application

After identifying an appropriate model based on theoretical considerations as outlined in the previous section, we strongly recommend to further plan the application of mathematical models ahead of data collection to ensure the interpretability and trustworthiness of the estimated model parameters. Specifically, three basic steps should be pursued when applying a cognitive model to a specific research question (see lower part of Figure 4, p. 10):

1. Researchers should plan their data collection to meet requirements for reliable and stable parameter estimates.
2. Model fit should be carefully evaluated after fitting the model to the empirical data.
3. Model parameters should be adequately related to other individual differences variables of interest such as intelligence test performances.

In the following section, we will provide step-by-step instructions using examples from the application of diffusion models in intelligence research, which may serve as guidelines when using any kind of cognitive model in individual differences research.

4.1. Design and Data Collection

4.1.1. Reliability and Stability of Estimated Model Parameters

The reliable estimation of model parameters from empirical data usually requires more data points than would be needed if only applying a statistical model to the data. For illustration, compare the description of reaction time distributions in decision tasks by a Gaussian distribution to the description by a diffusion model. When describing performance in a binary choice task by a Gaussian distribution, 20–30 trials are usually sufficient to provide reliable estimates of means and standard errors of the distribution [89]. When describing performance by a diffusion model, however, many more trials are needed because model parameters are not calculated analytically, but are found by fitting them to empirical response time distributions in an iterative process. Hence, a small number of trials will result in an inadequate representation of the full response time distribution and will therefore impair the estimation of model parameters describing distributional elements beyond measures of central tendency [90].

For the basic DDM (with the four parameters drift rate, boundary separation, starting point, and non-decision time), simulation studies have shown that 100 trials are sufficient to produce relatively reliable estimates of drift rates and that no further increases in parameter reliabilities are gained by increasing trial numbers beyond 500 trials [90]. For other measurement models less prominently used in individual differences research, such systematic simulation studies have not yet been conducted. Therefore, we urge researchers interested in applying less frequently used models to run a simulation study before starting data collection to determine how many experimental trials are needed for a reliable parameter recovery. While a simulation does not guarantee reliable parameter estimates for an experiment in general, it rules out that low reliability is due to noisiness in the parameter estimation process.

4.1.2. Trait, Situation, and Task Characteristics of Model Parameters

In addition, it is important to consider to what degree individual differences in model parameters reflect individuals' personality traits or abilities, and to what degree they reflect task-specific characteristics, state-specific characteristics, and unsystematic measurement error. Imagine applying a model of verbal working memory to complex span data: Model parameters such as the individual rate of verbal refreshing or the ability to resist interference from distracting stimuli would reflect both individuals' *general abilities* in verbal refreshing and inhibition of interference as well as their abilities to maintain memory stimuli *in this specific task*. Depending on the research question, researchers may be more interested in the general ability to maintain information in working memory as reflected in those parameters across different working memory tasks, or they may be interested in the specific ability to maintain information in working memory in precisely this task.

Usually, intelligence research questions are more likely to concern abilities generalized across specific operationalizations and situations than abilities in specific operationalizations or situations. However, model parameters estimated in a specific task are always going to contain both trait-, state- and task-specific amounts of variance [60,91]. For example, a latent state-trait analysis of DDM parameters in elementary cognitive tasks revealed that only about 45 percent of the variance in task-specific drift rates was accounted for by the common trait, and that only about 30 to 35 percent of the variance in task-specific boundary separation and non-decision time parameters were accounted for by their respective common traits [59]. Therefore, if a research question using cognitive models in intelligence research concerns performance in certain cognitive processes that is generalizable across specific operations, it may be worthwhile to design a test battery consisting of three or more tasks to which the cognitive model can be applied. Averaged or latent performance in process parameters across tasks will then allow a more precise estimate of individuals' performance in model parameters that is independent of specific task or situation characteristics.

4.2. Evaluation of Model Fit

4.2.1. Relative Model Fit: Which Model Provides the Best Account for the Data?

After finishing data collection, but before relating model parameters to intelligence tests or other covariates, it is necessary to evaluate how well a chosen model describes the empirical data and to possibly adjust model specifications to increase model fit. For most cognitive models, these empirical data consist of single-trial accuracies and/or response times, but aggregate measures such as proportion correct for different conditions might also be entered into the analysis. Before the raw data are entered into any kind of model, they should be carefully inspected for extreme values or other distributional properties that violate model assumptions and that may impair or even systematically bias parameter estimation. Once fidelity in these raw values has been established, cognitive models can be fitted to these empirical data. For this purpose, it has to be decided how many and which model parameters will be estimated and which model parameters will be fixed, because they are not expected to be affected by task characteristics or are not of interest for the current research question. Moreover, if experimental tasks contain several conditions, it may be necessary to decide which (if any) parameters are allowed to vary between conditions. It may even be desirable to split data from different conditions into separate data sets for separate model estimations to be able to subsequently model these separately estimated model parameters as latent variables. For this purpose, it may be helpful to reflect on the relationship between model complexity and the stability of parameter estimates: The more parameters of a model are estimated, the more likely it is to provide an accurate account of the data. However, if too many model parameters are estimated relative to the number of experimental trials, the stability of parameter estimates will be impaired [90,92].

Therefore, we suggest fitting several models to the empirical data containing different combinations of estimated or fixed parameters that are consistent with the current research question, unless there are strong theoretical reasons to decide on a specific model instantiation a priori. These models can then be compared based on parsimonious fit indices such as the Akaike Information Criterion (AIC; [93]) or the Bayesian Information Criterion (BIC; [94]), which take into account both model fit and model parsimony, to identify the model making the best trade-off between model fit and model complexity. As mentioned before, this model comparison step may not be necessary when a priori deciding for a specific instantiation of the model.

However, this model comparison approach only addresses one element of model fit evaluation, *relative model fit*. By identifying the best-fitting specification of the model out of a number of alternative specifications, it is possible to identify the model providing the best description of the empirical data. However, this does not guarantee that the best-fitting model provides a *good* description of the data.

4.2.2. Absolute Model Fit: How Well Does the Selected Model Describe the Data?

Therefore, in the next step the *absolute model fit* has to be evaluated to decide if the model can be accepted for all data sets. Absolute model fit is typically ascertained by either (a) statistical tests of model fit, (b) goodness-of-fit (GOF) indices, or (c) graphical inspections of model fit.

Statistical tests of model fit quantify the discrepancy between the empirical data and model predictions by means of a test statistic that is then tested for significance. However, this null hypothesis-testing of model fit contains several problems, as the power of statistical tests is closely tied to the amount of data available. When only a few trials are available, statistical tests may not be capable of rejecting the null hypothesis due to a lack of power, whereas when the trial number is large, statistical tests tend to become overly sensitive and detect even irrelevant deviations between the empirical data and model predictions [95]. To overcome some of the problems associated with null hypothesis testing, it has been suggested to simulate a large number of data sets based on the estimated model parameters, fit the model to each of the simulated data sets, and derive the 95 percent or 90 percent quantile of the resulting distribution of p-values as a critical value for the statistical tests of the originally estimated models [96,97]. However, models will still be accepted with an unknown error probability.

Goodness-of-fit indices are much more common in individual differences research, where they are used to evaluate the model fit of structural equation models [98], than in cognitive modeling. GOF indices standardize test statistics and take into account both model complexity and the number of data points. Typically, GOF indices have a fixed value range from 0 to 1 with certain cut-off values that indicate acceptable or good model fit. GOF indices are less frequently used in cognitive modeling, probably because several GOF indices used in structural equation modeling require the comparison of the actual model to a minimally plausible baseline model, which cannot be easily specified for most cognitive models. However, it has been recently suggested to adapt the root mean square error of approximation for the evaluation of cognitive models that can be fitted with a χ^2-distribution, such as the diffusion model [95]. Note that simulation studies have shown that this approach is only advisable when trial numbers are sufficiently large.

Finally, a third and widespread approach to the evaluation of absolute model fit is to graphically compare the empirical data to model predictions. To graphically inspect model fit, empirical data can be plotted against or overlaid by model predictions separately for each participant or aggregated over participants. This process can be rather time-consuming in larger samples if each participant is inspected individually. Moreover, it is important to be aware of the fact that graphical evaluations of model fit are inherently subjective and may therefore lead to spurious conclusions [99]. Having two independent raters evaluate model fit and discuss their conclusions may therefore increase the objectivity of the evaluation process.

If individual data sets can be identified that do not provide a satisfying model fit, raw data should be inspected for coding errors or outliers that may need to be removed (e.g., extremely fast reaction times with decision behavior close to guessing in a decision task). If model fit remains unacceptable, individual data sets may then need to be removed from further analyses, as it cannot be ascertained that the model parameters characterize the cognitive processes in the task accurately.

4.3. Relating Model Parameters to Intelligence Test Performance

Finally, after reliable estimates for the best fitting model have been obtained, the model parameters should be related to measures of intelligence. While this seems straightforward, there are actually two major methodological concerns.

First, extreme values in either parameter estimates or cognitive abilities measures need to be addressed. If extreme values (univariate outliers) are detected in parameter estimates, it is imperative to inspect if any outliers, coding errors, or abnormal distributional properties of the participant's raw data may have contaminated parameter estimation. If this is the case and if these outliers only constitute only a small amount of the data, they should be removed or winsorized and the model fitting procedure should be repeated to see if this treatment has led to more reasonable parameter estimates. If parameter estimates are still extreme or if outliers in raw data cannot be dealt with (e.g., because this participant's distribution of raw data deviates from model assumptions), this participant should be removed from further analyses as model parameters most likely reflect other properties of cognitive processes for this participant than for the rest of the sample. A similar problem is raised by multivariate outliers that may need to be removed based on an inspection of scatterplots or the calculation of the Mahalanobis distance. It goes without saying that information about the number of data points and/or participants removed and a rationale of their removal needs to be included in any description of the modeling results.

Second, researchers usually obtain one or more person-specific estimates for each model parameter of interest across different tasks or experimental conditions, just like they do when using aggregated performance measures such as accuracies or mean reaction times. Then the relationship of these model parameters with intelligence test scores is estimated by means of correlations or structural equation modeling. However, this approach represents a sequential analysis plan that treats the estimated parameters as manifest variables when quantifying the relationship between parameters of the cognitive model and the intelligence measures.

Treating estimated model parameters as manifest variables ignores the uncertainty that these parameters inherit from estimation and leads to an underestimation of standard errors in the second analysis step [100]. In fact, this is the case both for behavioral aggregates, such as mean reaction times or proportion correct, and for model parameters that are estimated from behavioral data or calculated from aggregate performance measures. Although this does not necessarily affect the estimated size of the relation between parameters obtained from cognitive models and intelligence measures, a sequential analysis plan always leads to an overestimation of the statistical significance of the estimated relationships [101].

A solution to this problem is hierarchical modeling [102,103][2]. In hierarchical modeling approaches, parameters of a cognitive model can be estimated simultaneously not only for all participants but across various tasks. Additionally, relationships with third variables, such as intelligence, can be estimated in the same step. On the one hand, such models avoid underestimating the standard errors of the relationship between model parameters and third variables such as intelligence measures by simultaneously estimating the model parameters and their relationship to intelligence (for an example of hierarchical models of the worst performance rule, see [101]). On the other hand, by assuming that the distribution of model parameters across individuals follows a higher order distribution[3] (so called hyper-priors), hierarchical models do not estimate parameters for each individual independently, but instead estimate model parameter for each individual informed by the parameter estimates from all other individuals. Not only does this render the parameter estimation more robust, but it also allows obtaining reliable estimates for the parameters of a cognitive model for each individual with fewer trials (for an example, see the hierarchical diffusion model: [105]).

Although this modeling approach is structurally similar to hierarchical modeling in latent variable models (i.e., SEM), there are some important differences. While hierarchical latent variable models separate general from specific factors in between person variances e.g., [8], hierarchical models in the field of cognitive models distinguish between parameters estimated within a person and the distribution of parameters between persons. In this, hierarchical modeling of cognitive processes is closely related to multi-level modeling separating the within and between person level [106].

For instance, when applying the diffusion model, parameters of each individual can be estimated independently without assuming a specific distribution of estimated model parameters across

[2] These two references focus on Bayesian hierarchical modeling. While Bayesian parameter estimation might have additional advantages over frequentist estimation approaches [104], the benefits of hierarchical modeling apply to both Bayesian and frequentist methods.

[3] Typically a Gaussian distribution with a mean and standard deviation is assumed.

participants. While this achieves the highest flexibility in parameter estimation, this approach ignores possible information from the between person level. In contrast, hierarchical modeling assumes that the parameters from each individual stem from a distribution of parameters on the between person level, and thus parameters for each individual are estimated taking information from all other subjects into account. As stated before, this account has two important benefits: (1) hierarchal modeling renders the parameter estimation for each individual more efficient [105]; and (2) parameters and their relationship to third variables like intelligence can be estimated simultaneously, accounting for the uncertainty of parameter estimates and thus adequately reporting the significance of the relationship between parameter estimates and third variables [101].

A serious complication of hierarchical modeling is that these models typically have to be explicitly specified and translated into code for each application, and that software solutions for parameter estimation are still rare. Nevertheless, hierarchical models do provide the mathematically accurate and sound solution for estimating the relationship between estimated model parameters and intelligence measures. Still, the sequential estimation of model parameters and their relationship to intelligence test scores seems to yield results comparable to hierarchical approaches [101]. In conclusion, while sequential approaches may overestimate the statistical significance of the relationship between model parameters and covariates (biasing inference), they nevertheless provide reasonable and unbiased estimates of the effect size of this relationship. For the future, it would be desirable that the application of cognitive modeling in the field of intelligence research or individual differences in general leads to the development of further simple software solutions or R packages [107,108] that simplify the use of hierarchical models.

5. Interpretation of the Results

Regardless of how the relationship between parameters from a cognitive model and intelligence measures is estimated, ultimately this relationship has to be interpreted on a conceptual level. Although parameters of a cognitive model provide more specific information about the cognitive process underlying the behavioral responses, these parameters still have to be interpreted with respect to the operationalization of the cognitive process. For instance, the diffusion model can be estimated in a broad set of tasks, ranging from perceptual judgment tasks (e.g., a random-dot motion task), over elementary cognitive tasks (e.g., Posner or Sternberg task), to even more complex memory tasks. In all of these different tasks, the diffusion model estimates the same set of parameters (i.e., drift rates, boundary separations, and non-decision times). However, this alone does not imply that model parameters estimated in the different tasks can be interpreted the same way. Specifically, the drift rate estimated in a random-dot motion task may represent the speed of perceptual information accumulation towards one response alternative. In a memory recognition task, however, the drift rate would rather be interpreted as the signal-to-noise ratio of the representation in memory. Beyond a theoretical discussion of the similarity of different tasks, statistical methods such as factor analysis or structural equation models can be used to get further information on variance that is shared across different tasks, or that is specific to a task or a situation (see: [59]). However, all in all, the interpretation of parameters of a cognitive model always relies on the specific experimental tasks.

In general, a cognitive model always represents a structural description of the behavioral measures from a specific task. The semantic meaning of the parameters of a model, however, can only be obtained with respect to the context (i.e., the task or materials) they are estimated in. Consider the following equation: $v = x/t$. On its own, this equation is merely a structural description how v can be obtained from x and t. In contrast, if the context of the observations of x as a distance between two points, and t as the time taken to get from one point to the other is known, then v can reasonably interpreted as the average speed of travel. It is just the same with parameters from any cognitive model: Without the context of their estimation they are merely transformations or estimated simplifications of the observed variables. Adding the semantical context of the observations however allows to interpret the parameters in a meaningful way.

All in all, the matter of adequately interpreting parameters of a cognitive model relates to a broader issue, namely validity. On the one hand, there is the question of how far a cognitive model provides a valid description of the cognitive process underlying the behavioral responses in a task. On the other hand, there is the question of how far individual differences in these parameters can be generalized across different tasks and assumed to represent between person variation in a more general and task-unspecific cognitive process. These are hardly problems that can be solved within a single study, but there is a combined effort needed to establish which parameters of cognitive models provide meaningful representations of individual differences in specific aspects of cognitive processing.

For example, attempts to unite psychophysiological and neuroimaging research with cognitive modeling may be particularly informative about issues of validity, as they allow a direct test of the idea that process parameters reflect certain neural correlates. Several studies have already suggested a close link between diffusion model parameter and neural processing correlates in the EEG. In particular, the latency of the N2, which is a neural correlate of visual encoding time, has been shown to be associated with the non-decision time parameter of the diffusion model [109], and the buildup rate of a positive centroparietral positive potential has been suggested to directly reflect the rate of evidence accumulation captured in the drift rate parameter on a neural level [110,111].

Altogether, following certain guidelines and carefully discussing the underlying assumptions and the operationalization when using a cognitive model provides a more explicit approach to measuring individual differences in cognitive processes, and thus represents a decisive improvement compared to the prevailing methods. At the very least, such careful reflection might immunize against the category error that cognitive models are accurate reflections of a latent, unobservable cognitive process. Any (cognitive) model cannot be anything but a simplification of reality that, if successful, captures the most important aspects, but never the entirety of an ontological (cognitive) process. In the same way that a map provides a simplification of a city's layout that is useful for navigation without ever providing a detailed account of the whole city system, cognitive models may refine our understanding of how those unobservable cognitive processes operate and thereby facilitate the measurement of certain process properties. Or, as Box [112] put it: "All models are wrong, but some models are useful".

6. Conclusions

Altogether, incorporating cognitive models in intelligence research provides numerous advantages. On the one hand, cognitive models provide explicit theoretical descriptions of cognitive processes that may underlie individual differences in general intelligence. On the other hand, they allow to estimate person specific parameters for each individual that can be related to measures of intelligence. Therefore, cognitive models allow to relate theoretically founded measures of individual differences in parameters of cognitive processes to individual differences in general intelligence and to overcome the fuzzy theoretical interpretation of behavioral indicators such as reaction times or accuracies.

Beyond that, cognitive models may allow identifying the effects of experimental or pharmacological interventions and training interventions on specific cognitive processes. For example, the shrinking spotlight model of selective attention might be used to test if a training intervention aimed at improving selective attention actually affects interference parameters of the model or if the intervention only reduces non-decision times or response thresholds. In a similar vein, the drift diffusion model might be used to characterize experimental effects of a pharmacological intervention on mental speed by distinguishing an increase in the velocity of evidence accumulation from an increase in motor response times. Last but not least, cognitive process parameters could not only be related to general intelligence differences, but also to individual differences in neural measures related to cognitive abilities, and may thus provide a different and possibly more complete perspective on the neuro-cognitive processes giving rise to individual differences in general intelligence. Taken together, the application of cognitive models as elaborate measurement tools provides an exciting new avenue for research on the neuro-cognitive processes underlying intelligence.

This approach focuses on insights into the cognitive correlates of general intelligence and does not represent an actual theory of general intelligence. On the one hand, it shows that developing cognitive models for specific cognitive processes is possible. On the other hand, proper theories of general intelligence that provide a comprehensive and mechanistic description of general intelligence as a cognitive process are scarce. As long as *theories* of general intelligence are mainly concerned with its factorial structure (i.e., psychometric theories, [2,5–8]), developing a cognitive model of general intelligence in the sense of a process theory remains difficult. One recently published positive counterexample is process-overlap theory, which suggests that the positive manifold may arise from a set of various domain-specific and domain-general cognitive processes which are linked multiplicatively [4]. In conjunction with mathematical models of the cognitive processes involved in process-overlap theory, this conceptual idea might be used to develop a formal model from different cognitive models that are linked in the multiplicative way suggested in process-overlap theory. In this sense, integrating mathematical models of cognitive processes that are correlated with measures of intelligence may provide a first step towards a comprehensive process theory of general intelligence—something for which the field has been searching for a long time.

Author Contributions: G.T.F. and A.-L.S. both wrote the manuscript, and revised and commented on the parts of the other author respectively. Since both authors were equally involved in the structural and conceptual design of the manuscript, the first authorship was shared between the two.

Acknowledgments: We thank Debora Fieberg and Ben Riemenschneider for providing Matlab and R code for the illustration of the DDM and the TBRS2 model.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SOB-CS	Serial-Order in Box Model for Complex Span Tasks
DDM	Drift-Diffusion Model
LBA	Linear Ballistic Accumulator Model
LCA	Leaky Competing Accumulator Model
TBRS	Time-base resource sharing theory/model
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
GOF	Goodness-of-fit

References

1. De Boeck, P. Intelligence, Where to Look, Where to Go? *J. Intell.* **2013**, *1*, 5–24. [CrossRef]
2. Thurstone, L. *Primary Mental Abilities*; University of Chicago Press: Chicago, IL, USA, 1938.
3. Kievit, R.A.; Davis, S.W.; Griffiths, J.; Correia, M.M.; Henson, R.N. A watershed model of individual differences in fluid intelligence. *Neuropsychologia* **2016**, *91*, 186–198. [CrossRef] [PubMed]
4. Kovacs, K.; Conway, A.R.A. Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychol. Inquiry* **2016**, *27*, 151–177. [CrossRef]
5. Spearman, C. 'General intelligence', objectively determined and measured. *Am. J. Psychol.* **1904**, *15*, 201–293. [CrossRef]
6. Horn, J.L.; Cattell, R.B. Refinement and test of the theory of fluid and crystallized general intelligences. *J. Educ. Psychol.* **1966**, *57*, 253–270. [CrossRef] [PubMed]
7. Carroll, J.B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*; Cambridge University Press: Cambridge, UK, 1993.
8. McGrew, K. The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*; The Guilford Press: New York, NY, USA, 2005.

9. Jensen, A.R. *Clocking the Mind: Mental Chronometry and Individual Differences*; Elsevier: Amsterdam, The Netherlands, 2006.
10. Kyllonen, P.C.; Zu, J. Use of Response Time for Measuring Cognitive Ability. *J. Intell.* **2016**, *4*, 14. [CrossRef]
11. Colom, R.; Abad, F.J.; Ángeles Quiroga, M.; Shih, P.C.; Flores-Mendoza, C. Working memory and intelligence are highly related constructs, but why? *Intelligence* **2008**, *36*, 584–606. [CrossRef]
12. Conway, A.R.; Cowan, N.; Bunting, M.F.; Therriault, D.J.; Minkoff, S.R. A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence* **2002**, *30*, 163–183. [CrossRef]
13. Engle, R.W.; Tuholski, S.W.; Laughlin, J.E.; Conway, A.R. Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *J. Exp. Psychol. Gen.* **1999**, *128*, 309–331. [CrossRef] [PubMed]
14. Kyllonen, P.C.; Christal, R.E. Reasoning ability is (little more than) working-memory capacity? *Intelligence* **1990**, *14*, 389–433. [CrossRef]
15. Unsworth, N.; Engle, R.W. Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *J. Mem. Lang.* **2006**, *54*, 68–80. [CrossRef]
16. Unsworth, N.; Engle, R.W. The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychol. Rev.* **2007**, *114*, 104–132. [CrossRef] [PubMed]
17. Miyake, A.; Friedman, N.P.; Emerson, M.J.; Witzki, A.H.; Howerter, A. The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: A latent variable analysis. *Cognit. Psychol.* **2000**, *41*, 49–100. [CrossRef] [PubMed]
18. Wongupparaj, P.; Kumari, V.; Morris, R.G. The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence* **2015**, *53*, 166–180. [CrossRef]
19. Thomson, G.H. A hierarchy without a general factor. *Br. J. Psychol. 1904–1920* **1916**, *8*, 271–281. [CrossRef]
20. Eriksen, B.A.; Eriksen, C.W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.* **1974**, *16*, 143–149. [CrossRef]
21. Stroop, J.R. Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **1935**, *18*, 643–662. [CrossRef]
22. Kane, M.J.; Meier, M.E.; Smeekens, B.A.; Gross, G.M.; Chun, C.A.; Silvia, P.J.; Kwapil, T.R. Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *J. Exp. Psychol. Gen.* **2016**, *145*, 1017–1048. [CrossRef] [PubMed]
23. McVay, J.C.; Kane, M.J. Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *J. Exp. Psychol. Gen.* **2012**, *141*, 302–320. [CrossRef] [PubMed]
24. Donders, F. On the speed of mental processes. *Acta Psychol.* **1969**, *30*, 412–431. [CrossRef]
25. Friston, K.J.; Price, C.J.; Fletcher, P.; Moore, C.; Frackowiak, R.S.; Dolan, R.J. The trouble with cognitive subtraction. *NeuroImage* **1996**, *4*, 97–104. [CrossRef] [PubMed]
26. Schubert, A.L.; Hagemann, D.; Voss, A.; Schankin, A.; Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **2015**, *51*, 28–46. [CrossRef]
27. Cronbach, L.J.; Furby, L. How we should measure change: Or should we? *Psychol. Bull.* **1970**, *74*, 68–80. [CrossRef]
28. Hedge, C.; Powell, G.; Sumner, P. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* **2017**, *50*, 1166–1186. [CrossRef] [PubMed]
29. Borsboom, D. *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*; Cambridge University Press: Cambridge, UK, 2005.
30. Lord, F.; Novick, M.; Birnbaum, A. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Boston, MA, USA, 1968.
31. Schmidt, F.L.; Hunter, J.E. Theory testing and measurement error. *Intelligence* **1999**, *27*, 183–198. [CrossRef]
32. Borsboom, D. Latent variable theory. *Meas. Interdiscip. Res. Perspect.* **2008**, *6*, 25–53. [CrossRef]
33. Borsboom, D.; Mellenbergh, G.J. True scores, latent variables and constructs: A comment on Schmidt and Hunter. *Intelligence* **2002**, *30*, 505–514. [CrossRef]
34. Schwarz, W. The ex-Wald distribution as a descriptive model of response times. *Behav. Res. Methods Instrum. Comput.* **2001**, *33*, 457–469. [CrossRef] [PubMed]

35. Schwarz, W. On the Convolution of inverse Gaussian and exponential Random Variables. *Commun. Stat. Theory Methods* **2002**, *31*, 2113–2121. [CrossRef]
36. Miller, R.; Scherbaum, S.; Heck, D.W.; Goschke, T.; Enge, S. On the Relation Between the (Censored) Shifted Wald and the Wiener Distribution as Measurement Models for Choice Response Times. *Appl. Psychol. Meas.* **2018**, *42*, 116–135. [CrossRef] [PubMed]
37. Keats, J.A.; Lord, F.M. A theoretical distribution for mental test scores. *Psychometrika* **1962**, *27*, 59–72. [CrossRef]
38. Wilcox, R.R. Estimating true score in the compound binomial error model. *Psychometrika* **1978**, *43*, 245–258. [CrossRef]
39. Matzke, D.; Wagenmakers, E.J. Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychon. Bull. Rev.* **2009**, *16*, 798–817. [CrossRef] [PubMed]
40. Farrell, S.; Lewandowsky, S. *Computational Modeling of Cognition and Behavior*; Cambridge University Press: Cambridge, UK, 2018.
41. Heathcote, A.; Brown, S.; Wagenmakers, E. An Introduction to Good Practices in Cognitive Modeling. In *An Introduction to Model-Based Cognitive Neuroscience*; Forstmann, B., Wagenmakers, E., Eds.; Springer: New York, NY, USA, 2015.
42. Oberauer, K.; Lin, H.Y. An interference model of visual working memory. *Psychol. Rev.* **2017**, *124*, 21–59. [CrossRef] [PubMed]
43. Zhang, W.; Luck, S.J. Discrete fixed-resolution representations in visual working memory. *Nature* **2008**, *453*, 233. [CrossRef] [PubMed]
44. Banks, W.P. Signal detection theory and human memory. *Psychol. Bull.* **1970**, *74*, 81–99. [CrossRef]
45. Oberauer, K.; Lewandowsky, S.; Farrell, S.; Jarrold, C.; Greaves, M. Modeling working memory: An interference model of complex span. *Psychon. Bull. Rev.* **2012**, *19*, 779–819. [CrossRef] [PubMed]
46. Bröder, A.; Schütz, J. Recognition ROCs are curvilinear—Or are they? On premature arguments against the two-high-threshold model of recognition. *J. Exp. Psychol. Learn. Mem. Cognit.* **2009**, *35*, 587–606. [CrossRef] [PubMed]
47. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **1978**, *85*, 59–108. [CrossRef]
48. Oberauer, K.; Lewandowsky, S. Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychon. Bull. Rev.* **2011**, *18*, 10–45. [CrossRef] [PubMed]
49. Carpenter, P.A.; Just, M.A.; Shell, P. What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychol. Rev.* **1990**, *97*, 404–431. [CrossRef] [PubMed]
50. Schmiedek, F.; Oberauer, K.; Wilhelm, O.; Süß, H.M.; Wittmann, W.W. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.* **2007**, *136*, 414–429. [CrossRef] [PubMed]
51. Ratcliff, R.; Schmiedek, F.; McKoon, G. A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence* **2008**, *36*, 10–17. [CrossRef] [PubMed]
52. Schmitz, F.; Wilhelm, O. Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence. *J. Intell.* **2016**, *4*, 13. [CrossRef]
53. Ratcliff, R.; Tuerlinckx, F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.* **2002**, *9*, 438–481. [CrossRef] [PubMed]
54. Van Ravenzwaaij, D.; Oberauer, K. How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *J. Math. Psychol.* **2009**, *53*, 463–473. [CrossRef]
55. Ratcliff, R.; McKoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **2008**, *20*, 873–922. [CrossRef] [PubMed]
56. Voss, A.; Rothermund, K.; Voss, J. Interpreting the parameters of the diffusion model: An empirical validation. *Mem. Cognit.* **2004**, *32*, 1206–1220. [CrossRef] [PubMed]
57. Lerche, V.; Voss, A. Experimental validation of the diffusion model based on a slow response time paradigm. *Psychol. Res.* **2017**. [CrossRef] [PubMed]
58. Lerche, V.; Voss, A. Retest reliability of the parameters of the Ratcliff diffusion model. *Psychol. Res.* **2017**, *81*, 629–652. [CrossRef] [PubMed]

59. Schubert, A.L.; Frischkorn, G.T.; Hagemann, D.; Voss, A. Trait Characteristics of Diffusion Model Parameters. *J. Intell.* **2016**, *4*, 7. [CrossRef]
60. Steyer, R.; Schmitt, M.; Eid, M. Latent state–trait theory and research in personality and individual differences. *Eur. J. Personal.* **1999**, *13*, 389–408. [CrossRef]
61. Longstreth, L.E. Jensen's reaction-time investigations of intelligence: A critique. *Intelligence* **1984**, *8*, 139–160. [CrossRef]
62. Ratcliff, R.; Thapar, A.; McKoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cognit. Psychol.* **2010**, *60*, 127–157. [CrossRef] [PubMed]
63. Ratcliff, R.; Thapar, A.; McKoon, G. Effects of aging and IQ on item and associative memory. *J. Exp. Psychol.* **2011**, *140*, 464–487. [CrossRef] [PubMed]
64. Schulz-Zhecheva, Y.; Voelkle, M.C.; Beauducel, A.; Biscaldi, M.; Klein, C. Predicting Fluid Intelligence by Components of Reaction Time Distributions from Simple Choice Reaction Time Tasks. *J. Intell.* **2016**, *4*, 8. [CrossRef]
65. Wagenmakers, E.J.; Van Der Maas, H.L.J.; Grasman, R.P.P.P. An EZ-diffusion model for response time and accuracy. *Psychon. Bull. Rev.* **2007**, *14*, 3–22. [CrossRef] [PubMed]
66. Wagenmakers, E.J.; van der Maas, H.L.J.; Dolan, C.V.; Grasman, R.P.P.P. EZ does it! Extensions of the EZ-diffusion model. *Psychon. Bull. Rev.* **2008**, *15*, 1229–1235. [CrossRef] [PubMed]
67. Voss, A.; Voss, J. Fast-dm: A free program for efficient diffusion model analysis. *Behav. Res. Methods* **2007**, *39*, 767–775. [CrossRef] [PubMed]
68. Brown, S.D.; Heathcote, A. The simplest complete model of choice response time: Linear ballistic accumulation. *Cognit. Psychol.* **2008**, *57*, 153–178. [CrossRef] [PubMed]
69. Usher, M.; McClelland, J.L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.* **2001**, *108*, 550–592. [CrossRef] [PubMed]
70. Barrouillet, P.; Bernardin, S.; Camos, V. Time Constraints and Resource Sharing in Adults' Working Memory Spans. *J. Exp. Psychol. Gen.* **2004**, *133*, 83–100. [CrossRef] [PubMed]
71. Barrouillet, P.; Bernardin, S.; Portrat, S.; Vergauwe, E.; Camos, V. Time and cognitive load in working memory. *J. Exp. Psychol. Learn. Mem. Cognit.* **2007**, *33*, 570–585. [CrossRef] [PubMed]
72. Vergauwe, E.; Barrouillet, P.; Camos, V. Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *J. Exp. Psychol. Learn. Mem. Cognit.* **2009**, *35*, 1012–1028. [CrossRef] [PubMed]
73. Vergauwe, E.; Camos, V.; Barrouillet, P. The impact of storage on processing: How is information maintained in working memory? *J. Exp. Psychol. Learn. Mem. Cognit.* **2014**, *40*, 1072–1095. [CrossRef] [PubMed]
74. Barrouillet, P.; Portrat, S.; Camos, V. On the law relating processing to storage in working memory. *Psychol. Rev.* **2011**, *118*, 175–192. [CrossRef] [PubMed]
75. Gauvrit, N.; Mathy, F. Mathematical transcription of the 'time–based resource sharing' theory of working memory. *Br. J. Math. Stat. Psychol.* **2018**, *71*, 146–166. [CrossRef] [PubMed]
76. Colom, R.; Rebollo, I.; Palacios, A.; Juan-Espinosa, M.; Kyllonen, P.C. Working memory is (almost) perfectly predicted by g. *Intelligence* **2004**, *32*, 277–296. [CrossRef]
77. Gignac, G.E. Working memory and fluid intelligence are both identical to g? Reanalyses and critical evaluation. *Psychol. Sci.* **2007**, *49*, 187–207.
78. Oberauer, K.; Farrell, S.; Jarrold, C.; Lewandowsky, S. What limits working memory capacity? *Psychol. Bull.* **2016**, *142*, 758–799. [CrossRef] [PubMed]
79. Oberauer, K.; Lewandowsky, S. Simple Measurement Models for Complex Working-Memory Tasks. Available online: https://osf.io/vkhmu/ (accessed on 17 July 2018).
80. White, C.N.; Ratcliff, R.; Starns, J.J. Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognit. Psychol.* **2011**, *63*, 210–238. [CrossRef] [PubMed]
81. White, C.N.; Servant, M.; Logan, G.D. Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychon. Bull. Rev.* **2018**, *25*, 286–301. [CrossRef] [PubMed]
82. White, C.N.; Curl, R. A Spotlight Diffusion Model Analysis of the Attentional Networks Task. 2017. Available online: https://osf.io/h9b8v/ (accessed on 17 July 2018).
83. Huebner, R.; Steinhauser, M.; Lehle, C. A dual-stage two-phase model of selective attention. *Psychol. Rev.* **2010**, *117*, 759–784. [CrossRef] [PubMed]

84. Huebner, R.; Tobel, L. Does attentional selectivity in the flanker task improve discretely or gradually? *Front. Psychol.* **2012**, *3*, 434.
85. Grange, J.A. flankr: An R package implementing computational models of attentional selectivity. *Behav. Res. Methods* **2016**, *48*, 528–541. [CrossRef] [PubMed]
86. Rogers, R.D.; Monsell, S. Costs of a predictible switch between simple cognitive tasks. *J. Exp. Psychol. Gen.* **1995**, *124*, 207–231. [CrossRef]
87. Schmitz, F.; Voss, A. Decomposing task-switching costs with the diffusion model. *J. Exp. Psychol. Hum. Percept. Perform.* **2012**, *38*, 222–250. [CrossRef] [PubMed]
88. Schmitz, F.; Voss, A. Components of task switching: A closer look at task switching and cue switching. *Acta Psychol.* **2014**, *151*, 184–196. [CrossRef] [PubMed]
89. Miller, J.; Ulrich, R. Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychon. Bull. Rev.* **2013**, *20*, 819–858. [CrossRef] [PubMed]
90. Lerche, V.; Voss, A.; Nagler, M. How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behav. Res. Methods* **2017**, *49*, 513–537. [CrossRef] [PubMed]
91. Eid, M. A multitrait-multimethod model with minimal assumptions. *Psychometrika* **2000**, *65*, 241–261. [CrossRef]
92. Van Ravenzwaaij, D.; Donkin, C.; Vandekerckhove, J. The EZ diffusion model provides a powerful test of simple empirical effects. *Psychon. Bull. Rev.* **2017**, *24*, 547–556. [CrossRef] [PubMed]
93. Akaike, H. Information theory and an extension of the maximum likelihood principle. In Proceedings of the 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, 2–8 September 1971; Petrov, B., Csáki, F., Eds.; Akadémiai Kiadó: Budapest, Hungary, 1973.
94. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [CrossRef]
95. Schubert, A.L.; Hagemann, D.; Voss, A.; Bergmann, K. Evaluating the model fit of diffusion models with the root mean square error of approximation. *J. Math. Psychol.* **2017**, *77*, 29–45. [CrossRef]
96. Clauset, A.; Shalizi, C.R.; Newman, M.E.J. Power-Law Distributions in Empirical Data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
97. Voss, A.; Nagler, M.; Lerche, V. Diffusion Models in Experimental Psychology. *Exp. Psychol.* **2013**, *60*, 385–402. [CrossRef] [PubMed]
98. Jackson, D.L.; Gillaspy, J.A.J.; Purc-Stephenson, R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychol. Methods* **2009**, *14*, 6–23. [CrossRef] [PubMed]
99. D'Agostino, R.B. Graphical analyses. In *Goodness-of-Fit Techniques*; D'Agostino, R.B., Stephens, M.A., Eds.; Marcel Dekker: New York, NY, USA, 1986; pp. 7–62.
100. Skrondal, A.; Laake, P. Regression among factor scores. *Psychometrika* **2001**, *66*, 563–575. [CrossRef]
101. Frischkorn, G.T.; Schubert, A.L.; Neubauer, A.B.; Hagemann, D. The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis. *J. Intell.* **2016**, *4*, 9. [CrossRef]
102. Lee, M.D. How cognitive modeling can benefit from hierarchical Bayesian models. *J. Math. Psychol.* **2011**, *55*, 1–7. [CrossRef]
103. Lee, M.D.; Wagenmakers, E.J. *Bayesian Cognitive Modeling: A Practical Course*; Cambridge University Press: Cambridge, UK, 2013.
104. Kruschke, J.K.; Liddell, T.M. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychon. Bull. Rev.* **2017**, *25*, 178–206. [CrossRef] [PubMed]
105. Vandekerckhove, J.; Tuerlinckx, F.; Lee, M.D. Hierarchical diffusion models for two-choice response times. *Psychol. Methods* **2011**, *16*, 44–62. [CrossRef] [PubMed]
106. Hamaker, E.L.; Dolan, C.V.; Molenaar, P.C.M. Statistical Modeling of the Individual: Rationale and Application of Multivariate Stationary Time Series Analysis. *Multivar. Behav. Res.* **2005**, *40*, 207–233. [CrossRef] [PubMed]
107. Heck, D.W.; Arnold, N.R.; Arnold, D. TreeBUGS: An R Package for Hierarchical Multinomial-Processing-Tree Modeling. *Behav. Res. Methods* **2017**, *50*, 264–284. [CrossRef] [PubMed]
108. Bürkner, P.C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **2017**, *80*, 1–28. [CrossRef]
109. Nunez, M.D.; Gosai, A.; Vandekerckhove, J.; Srinivasan, R. The latency of a visual evoked potential tracks the onset of decision making. *bioRxiv* **2018**. [CrossRef]

110. Kelly, S.P.; O'Connell, R.G. Internal and External Influences on the Rate of Sensory Evidence Accumulation in the Human Brain. *J. Neurosci.* **2013**, *33*, 19434–19441. [CrossRef] [PubMed]
111. O'Connell, R.G.; Dockree, P.M.; Kelly, S.P. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.* **2012**, *15*, 1729–1735. [CrossRef] [PubMed]
112. Box, G.E.P. Science and Statistics. *J. Am. Stat. Assoc.* **1976**, *71*, 791–799. [CrossRef]

Journal of
Intelligence

MDPI

Article

Trait Characteristics of Diffusion Model Parameters

Anna-Lena Schubert *, Gidon T. Frischkorn , Dirk Hagemann and Andreas Voss

Institute of Psychology, Heidelberg University, Hauptstr. 47-51, D-69117 Heidelberg, Germany; gidon.frischkorn@psychologie.uni-heidelberg.de (G.T.F.); dirk.hagemann@psychologie.uni-heidelberg.de (D.H.); andreas.voss@psychologie.uni-heidelberg.de (A.V.)

* Correspondence: anna-lena.schubert@psychologie.uni-heidelberg.de; Tel.: +49 6221 54-7746

Academic Editor: Oliver Wilhelm
Received: 31 March 2016; Accepted: 8 July 2016; Published: 18 July 2016

Abstract: Cognitive modeling of response time distributions has seen a huge rise in popularity in individual differences research. In particular, several studies have shown that individual differences in the drift rate parameter of the diffusion model, which reflects the speed of information uptake, are substantially related to individual differences in intelligence. However, if diffusion model parameters are to reflect trait-like properties of cognitive processes, they have to qualify as trait-like variables themselves, i.e., they have to be stable across time and consistent over different situations. To assess their trait characteristics, we conducted a latent state-trait analysis of diffusion model parameters estimated from three response time tasks that 114 participants completed at two laboratory sessions eight months apart. Drift rate, boundary separation, and non-decision time parameters showed a great temporal stability over a period of eight months. However, the coefficients of consistency and reliability were only low to moderate and highest for drift rate parameters. These results show that the consistent variance of diffusion model parameters across tasks can be regarded as temporally stable ability parameters. Moreover, they illustrate the need for using broader batteries of response time tasks in future studies on the relationship between diffusion model parameters and intelligence.

Keywords: mental speed; diffusion model; latent state-trait theory; response times; drift rate; boundary separation; non-decision time; temporal stability

1. Introduction

Being *quick on the uptake* or *being quick-witted* are popular idioms when describing smart individuals. Decades of research on the relationship between general intelligence and behavioral response times established the close link between mental abilities and the speed of information processing. A recent review of 172 studies reported an average correlation of $r = -.24$ between mental abilities and different measures of information processing speed [1]. More specifically, the correlations between mean response times in elementary cognitive tasks and general intelligence ranged from $r = -.25$ to $r = -.40$. In general, composite measures of response times tend to show higher correlation with general intelligence than single response time measures. For example, canonical correlations between test batteries of response time tasks and general intelligence ranged from $C = .55$ to $C = .72$ [2–4]. This suggests that it is the variance shared by different response time tasks—general mental speed—that is closely related to general intelligence.

Typically, mental speed is assessed by calculating the mean or median of each participant's intra-individual response time distribution in an experimental task. Because mean response times should exhibit ideal psychometric properties under assumptions of classical test theory, they are usually preferred over other parameters describing an individual's response time distribution such as the standard deviation, skewness, or kurtosis. However, there are a couple of findings suggesting that inter-individual differences in various parameters of intra-individual response time distributions may be of particular interest for explaining individual differences in mental abilities. First, previous research has indicated that the intra-individual standard deviation of response times is sometimes more strongly related to general intelligence than mean or median response times [5,6]. However, a recent meta-analysis of 24 studies found no consistent difference in the size of correlations between these measures [7]. Second, the worst performance rule [8] describes the phenomenon that when individual response times are ranked from fastest to slowest, the slowest response times are more predictive of general intelligence than the mean or best response times [9,10]. Third, using means as a measure of central tendency to summarize the information contained in individual response time distributions may not be the ideal choice, as response time distributions are strongly positively skewed, which contradicts the assumption of a Gaussian distribution.

Moreover, when only mean response times are used as measures of mental speed, information contained in the shape of response time distributions is inevitably lost. Response time distributions can be described as combined Gaussian and exponential distributions. Shifts in both of these elements of response time distributions can lead to a similar increase in mean response times, just like a shift in the response time distribution or an increase in skewness can result in higher mean response times [11,12]. Taken together, these phenomena suggest that it may be worthwhile to consider the complete distribution of response times when analyzing the relationship between mental speed and mental abilities.

Beyond that, the distinction between decision times and movement times has proven to be another critical issue for the study of individual differences in mental chronometry. Some experimental setups require participants to rest their finger(s) on a home button that they are instructed to release as soon as the stimulus is presented and their decision is made. Subsequently, they have to press one of several response keys. Home-button setups are supposed to allow dissociating between the time required for the stimulus to be perceived, encoded, and processed (decision time, DT), and the time required for response execution (motor time, MT). Although DT and MT tend to load on two orthogonal factors [13], moderate correlations between DT and MT are reported in many studies [6]. The view that some degree of process contamination exists in both of these measures is supported by two experiments showing that participants release the home key immediately after detecting the stimulus, but before finalizing their decision, in anticipation of their response [14]. These results suggest that a clear-cut distinction between decision times and movement times cannot be easily achieved by a home-button setup and that other methods have to be employed to obtain a process-pure measurement of the speed of information processing.

1.1. The Diffusion Model: A Process Model of Speeded Binary Decision Making

Mathematical models of response times can overcome the aforementioned problems (i.e., providing adequate parameters for the description of a response time distribution, analyzing the complete distribution, allowing a more valid distinction between the time required for decision processes and for non-decision processes such as movement times), because they provide a process-based account of decision making that uses a participant's whole response time distribution to estimate parameters reflecting various elements of the decision process. The most prominent mathematical model of binary response time tasks is the diffusion model, which is a random-walk model that assumes a continuous information accumulation during a binary decision until one of two decision thresholds is reached [15]. This information accumulation process can be described by a Wiener diffusion process consisting of a constant systematic component, the drift, and normally distributed random noise.

The basic diffusion model contains four parameters (see Figure 1): The drift rate (v) reflects the strength and direction of the systematic influence on the diffusion process and is a direct performance measure for the speed of information uptake. Boundary separation (a) reflects the amount of information considered for a decision, which is for example influenced by a participant's cautiousness or by instructions stressing speed over accuracy and vice versa, e.g., [16,17]. The starting point (z) reflects a priori biases in decision making that can be influenced by asymmetric pay-off matrices [16].

Finally, the non-decision time (t_0) reflects the time required for all sorts of decision-unrelated processes such as encoding or motor programming and execution.

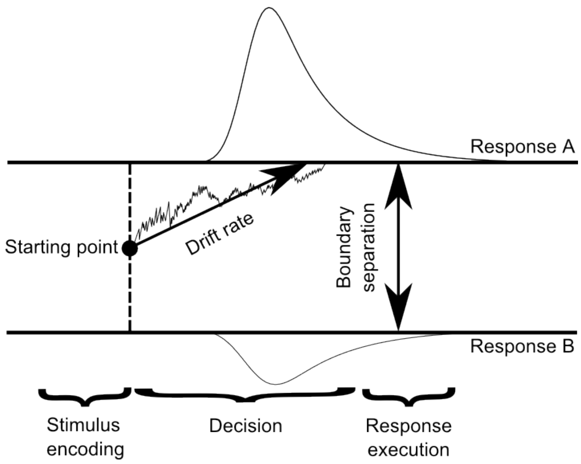

Figure 1. A simplified illustration of the basic diffusion model. Information accumulation begins at the starting point z and continues with a mean drift rate v (affected by random noise) until one of two thresholds is hit. Boundary separation a represents the amount of information that has to be accumulated before a decision is made. Outside of the information accumulation process, non-decision time t_0 (not shown here) quantifies the time of non-decision processes such as stimulus encoding and response execution. This figure was inspired by the illustration of the diffusion model in Voss et al. [18].

For each participant, a set of parameters is estimated by permuting parameter values until the predicted response time distribution closely resembles the empirical response time distributions. One advantage of the diffusion model is that it maps different cognitive processes to different model parameters and provides process-pure measures of these parameters that have been empirically validated [16]. Hence, the diffusion model takes into account the whole distribution of response times and allows separating the speed of information uptake—reflected in the drift rate parameter—from the speed of the motor response—reflected in the non-decision time parameter.

The diffusion model has seen a recent rise in popularity thanks to the publication of software solutions that allow fitting the model without extensive programming knowledge [19–22]. Applications in individual differences research include studies on individual differences in attention [23], in impulsivity, e.g., [24], in mental abilities, e.g., [25–28], in numeracy [29], and in word recognition [30].

1.2. Correlations between Diffusion Model Parameters and Mental Abilities

First studies on the relationship between diffusion model parameters and mental abilities have supported the notion that the diffusion model may help to identify the speed of specific cognitive processes and their specific associations with intelligence. This research is based on the assumption that individual differences in these model parameters reflect trait-like properties of cognitive processes. Four studies analyzing the relation between diffusion model parameters and mental abilities reported substantial correlations between drift rate and mental abilities ranging from r = .18 to .90 [25–28]. Ratcliff et al. [25] asked participants in three age groups (college age, 60–74 years old, 75–90 years old) to complete the vocabulary and matrix reasoning subtests of the Wechsler intelligence test and analyzed correlations with a latent drift rate factor from a numerosity discrimination, recognition memory, and lexical decision response time task. Correlations ranged from r = .60 to .90 for the vocabulary subtest, and from r = .36 to .85 for the matrix reasoning subtest. Moreover, the same participants also completed an item and associative recognition task. Manifest correlations between drift rate parameters estimated from response time distributions of these tasks ranged from r = .18 to .67 with the matrix reasoning subtest, and from r = .28 to .68 with the verbal subtest [26]. Schmiedek et al. [27] reported a latent correlation of r = .79 between reasoning ability and drift rates in eight response time tasks (including verbal, numerical, and spatial tasks) in a student sample, and Schubert et al. [28] reported a correlation of r = .50 between general intelligence and a drift rate factor derived from three response time tasks (choice response task, recognition task, semantic discrimination task). Taken together, these results indicate that smarter individuals have a greater speed of information uptake as reflected in the drift rate parameter.

Previous studies have reported less consistent associations between non-decision time and mental abilities. Ratcliff et al. [25] found latent correlations ranging from $r = -.14$ to .40 for the vocabulary subtest, and from $r = -.04$ to $-.50$ for the matrix reasoning subtest, and Ratcliff et al. [26] reported manifest correlations ranging from $r = -.29$ to .37 between non-decision times in item recognition tasks and associative recognition tasks and intelligence, leading them to conclude that non-decision times were not reliably associated with intelligence. Schmiedek et al. [27], however, reported a small but significant positive correlation between a latent non-decision time factor and reasoning ability, r = .25. Schubert et al. [28] also found evidence for an association between the non-decision time parameter and general intelligence, but in the direction that more intelligent individuals had shorter non-decision times, $r = -.42$. More research is needed to conclude whether intelligence is consistently associated with the non-decision time parameter. Because the non-decision time parameter captures to some degree movement times, it would be consistent with previous research if it was not related to intelligence. However, the non-decision time parameter does not only reflect motor speed, but also the time required for encoding and for memory-related processes [31] and may thus be related to general intelligence.

Finally, some studies have found associations between boundary separation and mental abilities. Ratcliff et al. [25] found latent correlations ranging from $r = -.02$ to .52 for the vocabulary subtest, and from r = .15 to .37 for the matrix reasoning subtest, and Ratcliff et al. [26] reported manifest correlations ranging from $r = -.32$ to .14 between boundary separation parameters in item and associative recognition tasks and intelligence. Schmiedek et al. [27] reported a negative latent correlation between a boundary separation factor and reasoning ability, $r = -.48$. Again, the results seem largely inconsistent, because greater intelligence has been found to be associated both with a smaller and a larger boundary separation, i.e., with less and more decision cautiousness.

Taken together, these first studies strongly support the view that more intelligent individuals have a higher drift rate across a broad variety of response time tasks and participant samples. In comparison, associations between intelligence and non-decision time/boundary separation parameters were smaller and largely inconsistent within and across studies. The number of studies relating diffusion model parameters to mental abilities is still too small to allow identifying variables moderating the size and direction of these associations such as sample characteristics or task difficulties.

1.3. Diffusion Model Parameters as Personality Traits

Before a claim can be made that diffusion model parameters reflect trait-like properties of cognitive processes, it has to be shown that they qualify as trait-like variables themselves. Several authors have previously suggested an integration of item response theory (IRT) and the diffusion model into a latent variable model, which takes into account that the mean drift rate consists of a person part and an item part [32–34]. It has been shown that diffusion IRT models can account for psychometric responses on tests of bipolar traits [33] as well as for ability tests under the assumption that abilities have a natural zero point [34]. Moreover, Vandekerckhove [35] has suggested a cognitive latent variable model based on Bayesian hierarchical modeling for the simultaneous analysis of response time data and personality or ability test data. This framework allows estimating latent task abilities, which are reflected in diffusion model parameters across different tasks and which can be related to external covariates. However, none of these approaches explicitly takes into consideration the core assumption of many personality theories, which consists of the temporal stability and trans-situational consistency of personality traits [36–38]. Showing that diffusion model parameters can be accounted for with a latent variable model is a necessary, but not a sufficient condition to conclude that they can be treated as trait-like variables, as state variables can also be modeled as latent variables [38]. Hence, diffusion model parameters only qualify as trait-like variables if it can be shown that they are stable across time and consistent over different tasks/situations.

Moreover, because diffusion model parameters can be affected by task properties [16], it is important to identify to what extent specific parameter values reflect person properties common to different tasks and to what extent they reflect specific person-task interactions. This knowledge would facilitate the planning of future studies on the relationship between mental speed and intelligence by helping to determine how many response time tasks are needed for a reliable assessment of common model parameters. Moreover, estimates of the consistency of task-specific model parameters would allow evaluating the associations between these model parameters and intelligence. From the viewpoint of classical test theory (CTT), the effects of the situation and the effects of the person–situation interaction act as nuisance variables when estimating the diffusion model parameters, i.e., these effects add to the error variance. In turn, the variance portions that are not due to individual differences in a latent trait attenuate the correlations between diffusion model parameters and intelligence. Thus, the estimates of consistency inform us about the maximum correlation that may be expected in empirical investigations of the relationship between diffusion model parameters and intelligence (the coefficient of consistency indicates an upper bound of the correlation between the diffusion model parameters and intelligence). Therefore, knowing the consistencies helps to interpret the empirical correlations that have been reported in previous research and that may be reported in future research.

Previous research has suggested that response times may be experimentally influenced by fatigue and performance-dependent rewards [39,40]. However, it is as of yet unclear whether such situational factors affect the relative performance in response time tasks when not experimentally induced and how this might be reflected in diffusion model parameters. Previous research on the temporal stability of diffusion model parameters is scarce, but first studies suggest a moderate temporal stability over a period of one week. One week test-retest correlations of drift rate, threshold separation, and non-decision time parameters ranged from r = .48 to .86 (mean r = .66) in lexical decisions tasks [30,41], from r = .35 to .77 (mean r = .56) in a recognition memory task [41], and from r = .30 to .79 (mean r = .67) in an associative priming task [41].

However, a period of one week may be suited to estimate the reliability with test-retest correlations, but it does not convey much information about the temporal stability and trans-situational consistency of model parameters in a broader sense. Because intelligence is known to show a great temporal stability over longer periods of time, e.g., [13,42], diffusion model parameters should show a similar temporal stability if considered to reflect processes giving rise to individual differences in general intelligence. Moreover, if diffusion model parameters are to be considered as trait-like properties of cognitive processes, not only the temporal stability of parameters in specific tasks, but the temporal stability of model parameters across tasks—i.e., of hierarchical or latent model parameters—is of particular interest for individual differences research.

To evaluate whether diffusion model parameters (namely: drift rate, boundary separation, and non-decision time) qualify as trait-like variables, we asked participants to complete three response time tasks at two laboratory sessions approximately eight months apart. The tasks were so-called elementary cognitive tasks (ECTs) that are the most widely used tasks in individual differences research on the relationship between mental speed and mental abilities [1]. ECTs are tasks with minimal cognitive demands that minimize unwanted sources of variance such as strategy use and learning effects. The first task we used was a visual choice response time task in which participants had to decide in which of four squares a cross appeared. Previous research has shown that an increase of stimulus alternatives leads to a linear increase in response times [43]. This linear increase indicates that evidence is accumulated continuously until a decision point is reached and that this process takes longer the more stimulus alternatives are presented, either because more evidence has to be considered or because the process gets noisier. We know that other cognitive models such as the linear ballistic accumulator (LBA [44]) or the leaky competing accumulator (LCA [45]) model may be better suited to model response times in this task, but we are aware of at least one study that previously used the diffusion model to model behavioral data of this task and came to similar results as a study that used the LCA model [28,46]. The second task we used was the Sternberg memory scanning task [47], in which participants have to decide whether a probe item was part of a previously presented memory set. As Ratcliff [15] has shown, performance in this task can be adequately described by the diffusion model under the assumption of parallel diffusion processes for each memory set item. The third task we used was the Posner letter matching task [48], in which participants have to decide whether two letters have (a) the same physical and (b) the same name identity. While participants decide whether the letters are identical, they may either accumulate information simultaneously from both letters regarding their similarity until reaching a threshold, or they may first encode one of the letters and then apply a decision process to the second one. In the first case, one common diffusion process or two parallel ones might be occurring, whereas in the latter case only a single diffusion process should occur reflecting the comparison process to the previously encoded stimulus. Moreover, evaluating the name identity of letters may require additional processing demands due to the access of long-term memory, which should be reflected in the non-decision time [31].

We used latent state-trait (LST) models to quantify the amount of variance in model parameters that can be attributed to a common trait, to situational influences, to specific experimental tasks, and to measurement error [49,50]. If diffusion model parameters qualify as trait-like properties of cognitive processes affecting response times in a variety of tasks , they should show at least moderate consistencies and low occasion-specificities.

2. Experimental Section

2.1. Participants

We recruited N = 134 participants (81 females, 53 males, M_{age} = 37.1, SD_{age} = 13.8) from different educational and occupational backgrounds. Of these, N = 114 (66 females, 48 males, M_{age} = 36.9, SD_{age} = 13.5) attended both the first and the second experimental session that were approximately eight months apart. Participants who did not attend the second laboratory session tended to have smaller drift rates, average $d = -0.46$, greater boundary separation parameters, average $d = 0.44$, and negligible differences in non-decision time parameters, average $d = -0.06$. We only included participants who attended both experimental sessions in our analyses. All participants had normal or corrected to normal vision. As a reward for their participation, they received 100€ and feedback about their personal results.

2.2. Measures

Response Time Tasks

Visual choice response time task. We used a choice response time (CRT) task with either two (CR2) or four (CR4) response alternatives. Four white squares were presented in a row on a black screen. Participants' middle and index fingers rested on four keys directly underneath the squares. After a delay of 1000–1500 ms, a cross appeared in one of the four squares and participants had to press the corresponding key as fast as possible. In the two-choice response time condition, the choice space was reduced to two squares in which the cross could appear for 50 subsequent trials. After completing a block of 50 trials, participants were informed that the cross could now only appear in a different combination of squares (outer left and left squares, outer right and right squares, inner squares, outer squares). In the four-choice response time condition, the cross could appear in any of the four squares.

Both conditions began with ten practice trials with immediate feedback followed by 200 test trials without feedback. The order of conditions was counterbalanced across participants.

Sternberg memory scanning task. Participants were shown a memory set consisting of one (set size one, S1), three (set size three, S3), or five (set size five, S5) digits from 0 to 9 on a black computer screen. Subsequently, participants were shown a probe digit and had to decide whether the probe was contained in the previously presented memory set by pressing one of two keys. This was the case in 50% of the trials. The position of keys indicating whether the probe item was part of the memory set was counterbalanced across participants. Each of the three conditions began with ten practice trials with immediate feedback followed by 100 test trials without feedback. The order of conditions was counterbalanced across participants.

Posner letter matching task. Participants were shown two letters and had to decide whether they were identical. In the physical identity (PI) condition, participants had to decide whether they were physically identical, and in the the name identity (NI) condition, they had to decide whether the two presented letters had the same name. The position of keys indicating whether the letters were identical was counterbalanced across participants. Both conditions began with ten practice trials with immediate feedback followed by 300 test trials without feedback. All participants started with the PI condition at the first laboratory session, whereas all participants started with the NI condition at the second laboratory session.

2.3. Procedure

The two experimental sessions were approximately eight months apart. We administered the CRT task first, followed by the Sternberg memory scanning task and the Posner letter matching task. The order of tasks was the same for all participants at both experimental sessions. An EEG was recorded while participants completed the tasks (data are not reported here). Each session took approximately three hours.

2.4. Data Analysis

2.4.1. Response Time Data

We discarded any RTs faster than 100 ms or slower than 3000 ms. In a second step, we discarded any trials with logarithmized RTs exceeding $\pm$ 3 SDs of the mean of each condition on an intra-individual level. Diffusion model parameters were estimated with fast-dm-30 [19] for each participant, each condition, and each experimental session separately. We analyzed the data in such a way that responses to the upper threshold reflected correct decisions and that responses to the lower threshold reflected incorrect decisions. For the CR4 task, this implied that responses to any of the three incorrect locations were aggregated into one RT distribution of incorrect trials. The starting point z was fixed to $a/2$, and the difference in speed of response execution d and the trial-to-trial variability parameters of the drift and the starting point were fixed to 0. Thus, four parameters were estimated for each participant, each condition, and each experimental session: The drift rate v, the boundary separation a, the non-decision time t_0, and the variability of the non-decision time st_0. Model parameters were estimated with the Kolmogorov-Smirnov statistic. Model fit was evaluated graphically by plotting predicted values against empirical values for the 25, 50, and 75 quantiles of the RT distribution separately for each task and each laboratory session (see Figure A1, Appendix A).

2.4.2. Statistical Analysis

We used latent state-trait models to assess the relationship between diffusion model parameters estimated for the three response time tasks across the two experimental sessions. We specified a structural equation model with a common trait T, a state residual SR_i for each of the two measurement occasions i, and a method factor M_j for each of the three experimental tasks j. The path coefficients of the common trait T as well as the path coefficients of the state residuals SR_i on the latent states S_i were fixed to 1. Moreover, the path coefficients of the method factors M_j were also fixed to 1. Finally, the path coefficients of the latent states S_i were estimated with the exception of the path coefficient loading on the CR2 task, which was fixed to 1. The variances of error residuals were set equal across measurement occasions. In addition, we also specified LST models separately for diffusion model parameter and each experimental task. These structural equation models were specified as above with the exceptions that method factors M_j reflected experimental conditions within the task and that all paths were fixed to 1. We did not analyze the mean structure in any of the models and fixed the intercepts to 0, because the diffusion model parameters were z-standardized prior to being entered into structural equation models. Note that the assumption of measurement invariance is not necessary to explore the relative temporal stability of latent model parameter states [38]. All in all, we calculated four LST models for each diffusion model parameter (v, a, t_0): One model across all experimental tasks and three separate models for each of the experimental tasks (CRT task, Sternberg memory scanning task, Posner letter matching task).

Moreover, we calculated several LST parameters to assess the reliability, consistency, occasion-specificity, and method-specificity of the model parameters [38]. The coefficient of consistency is computed as $\sigma^2(T)/\sigma^2(Y_{ij})$ and reflects the proportion of variance of the manifest variable Y_{ij} that can be accounted for by individual differences in the latent trait T. The coefficient of occasion-specificity is computed as $\sigma^2(SR_i)/\sigma^2(Y_{ij})$ and reflects the proportion of variance that is due to situational effects SR_i. Similarly, the coefficient of method-specificity is computed as $\sigma^2(M_j)/\sigma^2(Y_{ij})$ and reflects the proportion of variance that can be accounted for by a specific method M_j. Taken together, these different sources of systematic variation contribute to the reliability of a manifest variable Y_{ij}, which can thus be computed as $[\sigma^2(T) + \sigma^2(SR_i) + \sigma^2(M_j)]/\sigma^2(Y_{ij})$. For the purposes of our analyses, we defined LST parameters $\leq$.30 as small, between .30 and .60 as moderate, and $\geq$.60 as great, as consistencies of well-known intelligence and personality tests all exceeded this threshold value [51,52]. Note that these cut-offs should only be treated as heuristics. For the evaluation of LST parameters, we sighted the literature carefully for previous evaluations of these parameters. In their formulation of LST theory, Steyer et al. [49] did not offer any guidelines for the interpretation of LST parameters. Hence, we used previous applications of LST theory in individual differences research on mental abilities and personality traits as benchmarks. Danner et al. [51] calculated the LST parameters of intelligence tests and reported consistencies ranging from .67 to .72 and reliabilities ranging from .86 to .83, which they evaluated as "great". In comparison, method-specificities of intelligence tests ranged from .13 to .24 and were evaluated as "small". Deinzer et al. [52] evaluated the trait characteristics of different personality questionnaires (Freiburg Personality Inventory (FPI), NEO Five-Factor Inventory (NEO-FFI), Eysenck Personality Questionnaire (EPI)) and reported consistencies ranging from .73 to .94 for the FPI, from .62 to .92 for the NEO-FFI, and from .72 to .83 for the E and N scales of the EPI. In comparison, occasion-specificities ranged from .00 to .17 for the FPI, from .00 to .22 for the NEO-FFI, and from 0 to .16 for the E and N scales of the EPI. These results led the authors to conclude that situational and/or interactional influences explained a significant proportion of variance of these personality questionnaires and that test scores depended both on latent traits and (albeit to a lesser degree) on situational and/or interactional influences.

Structural equation models were estimated with MPlus 7 [53] (the data and analysis files for all structural equation models are provided in the supplementary materials online). Model fit was evaluated with the chi-square test, the comparative fit index (CFI) and root-mean-square error of approximation (RMSEA). According to Browne and Cudeck [54] and Hu and Bentler [55], we considered CFI values > .90 and RMSEA values < .08 to indicate acceptable model fit, and CFI values > .95 and RMSEA values < .06 to indicate good model fit. Missing values were accounted for using the ML algorithm implemented in MPlus.

3. Results and Discussion

3.1. Descriptive Data

Table 1 shows the mean RTs and accuracies for each condition of the three response time tasks at both experimental sessions. All in all, accuracies were relatively high due to the low complexity of the response time tasks. Moreover, RTs tended to increase with increasing information-processing demands within each task.

Table 1. Mean accuracies (ACC), mean RTs (RT), and mean diffusion model parameters (v, a, t_0, and st_0) across conditions in the three response time tasks at both measurement occasions (SDs in parantheses).

	ACC	RT	v	a	t_0	st_0
Session 1						
CR2	1.00 (.01)	383.45 (58.08)	5.55 (1.31)	1.15 (0.26)	0.27 (0.04)	0.08 (0.04)
CR4	.98 (.02)	479.92 (89.30)	4.68 (1.30)	1.18 (0.34)	0.34 (0.06)	0.15 (0.09)
S1	.98 (.02)	585.07 (108.54)	3.48 (1.15)	1.63 (0.98)	0.35 (0.08)	0.13 (0.11)
S3	.98 (.02)	719.53 (161.38)	3.20 (1.04)	1.63 (0.79)	0.45 (0.09)	0.16 (0.11)
S5	.96 (.03)	878.86 (232.06)	2.55 (0.79)	1.73 (0.52)	0.53 (0.13)	0.21 (0.19)
PI	.98 (.02)	614.90 (88.35)	4.00 (0.94)	1.27 (0.25)	0.45 (0.05)	0.14 (0.06)
NI	.97 (.02)	699.66 (112.81)	2.97 (0.70)	1.46 (0.35)	0.45 (0.05)	0.14 (0.07)
Session 2						
	ACC	RT	v	a	t_0	st_0
CR2	1.00 (.01)	381.26 (61.00)	5.58 (1.56)	1.14 (0.27)	0.27 (0.03)	0.08 (0.05)
CR4	.98 (.02)	467.36 (85.75)	4.72 (1.11)	1.14 (0.32)	0.34 (0.04)	0.14 (0.06)
S1	.98 (.02)	584.02 (135.64)	3.65 (1.35)	1.38 (0.41)	0.36 (0.07)	0.13 (0.10)
S3	.98 (.03)	706.61 (176.81)	3.24 (1.00)	1.43 (0.35)	0.47 (0.10)	0.16 (0.11)
S5	.95 (.09)	850.98 (223.18)	2.52 (1.00)	1.54 (0.48)	0.53 (0.13)	0.19 (0.15)
PI	.98 (.02)	605.19 (102.41)	4.04 (1.06)	1.33 (0.36)	0.42 (0.05)	0.12 (0.06)
NI	.97 (.2)	704.38 (126.36)	3.10 (0.77)	1.49 (0.38)	0.45 (0.06)	0.15 (0.08)

3.2. Diffusion Model Analysis

Results from one participant in the set size one condition of the Sternberg memory scanning task and results from one participant in the name identity condition of the Posner letter matching task had to be discarded, because the predicted RTs deviated strongly from the empirical RTs across all four quantiles (Figure A1 in Appendix A displays the relation of empiric vs. predicted response times for the remaining data).

Drift rates tended to decrease and boundary separation and non-decision parameters tended to increase with increasing information-processing demands within each task. Drift rates of experimental tasks were comparable across experimental sessions, all $|ds| \leq 0.13$, except for the name identity condition of the Posner letter matching task. Here, drift rates were slightly greater at the second than at the first laboratory session, d = 0.24. Boundary separation parameters were also comparable across sessions, all $|ds| \leq 0.19$, except for the Sternberg memory scanning task. Here, boundary separation parameters decreased from the first to the second session, all $|ds| \geq 0.27$. Finally, non-decision time parameters were also comparable across sessions, all $|ds| \leq 0.16$, except for the set size one condition of the Sternberg memory scanning task, in which non-decision time parameters increased across sessions, d = 0.21, and the physical identity condition of the Posner letter matching task, in which non-decision time parameters decreased across sessions, d = 0.48.

3.2.1. Drift Rate

Manifest correlations of drift rate parameters are shown in Table A1 in Appendix B. The LST model for drift rate across experimental tasks provided a good fit for the data, $\chi^2(94) = 136.44$, p = .003, CFI = .95, RMSEA = .06 [.04; .09]. However, the variances of the latent state residuals and of the method factor for the Posner letter matching task were non-significant ($VAR(SR_1)$ = 0.03, p = .660; $VAR(SR_2)$ = 0.03, p = .643; $VAR(P_v)$ = 0.00, p = .992). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(97)$ = 143.94, p = .001, CFI = .94, RMSEA = .07 [.04; .09]. Although the χ^2 test indicated significant differences between the implied and empirical covariance structure, the goodness-of-fit indices CFI and RMSEA indicated acceptable model fit [54,55]. See Figure 2 for the modified model.

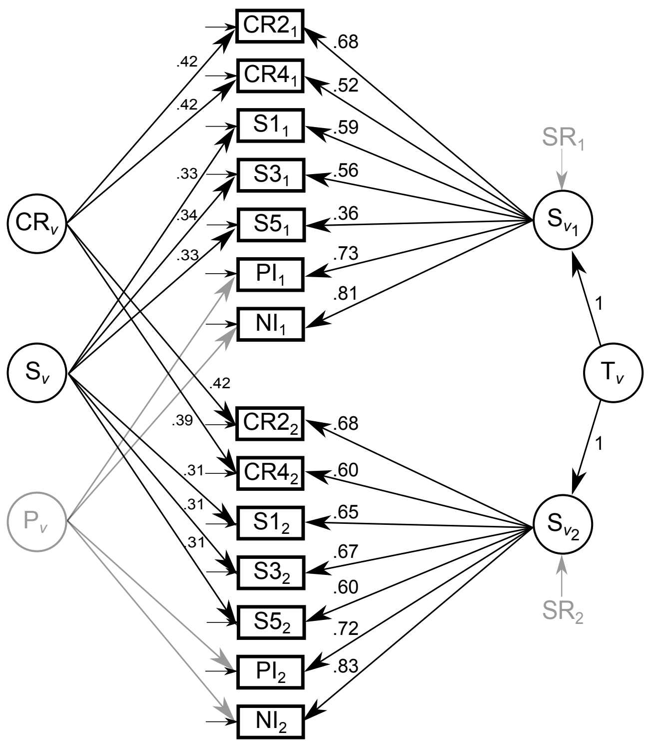

Figure 2. The latent state-trait model of drift rate parameters consists of a common trait T, a state residual SR_i for each of the two measurement occasions i, and a method factor M_j for each of the three experimental tasks. CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity. Latent variables displayed in gray were non-significant.

Subsequently, we calculated the coefficients of reliability, consistency, occasion-specificity, and method-specificity based on LST theory (see Table 2). Drift rate parameters estimated in the CRT and Sternberg memory scanning task showed moderate consistencies and low method-specificities. Drift rate parameters estimated in the Posner letter matching task showed the highest consistencies and no method-specificities. Overall, the temporal stabilities of the latent drift rate states were great as state residuals were not significant and occasion-specificities were therefore zero, but the reliabilities of the manifest drift rate parameters were only moderate, with coefficients of reliability ranging from .38 to .69.

Table 2. Latent state-trait theory parameters of diffusion model parameters. Occ. Spec. = Occasion-specificity; Meth. Spec. = Method-specificity.

Session	Consistency		Occ. Spec.		Meth. Spec.		Reliability	
	1	2	1	2	1	2	1	2
Drift rate parameters								
CR2	.46	.46	.00	.00	.17	.17	.63	.63
CR4	.28	.36	.00	.00	.17	.15	.45	.51
S1	.34	.42	.00	.00	.11	.10	.45	.52
S3	.31	.44	.00	.00	.12	.09	.43	.54
S5	.28	.36	.00	.00	.11	.10	.38	.45
PI	.53	.52	.00	.00	.00	.00	.53	.52
NI	.66	.69	.00	.00	.00	.00	.66	.69
Boundary separation parameters								
CR2	.14	.14	.00	.00	.20	.20	.35	.35
CR4	.38	.33	.00	.00	.17	.19	.55	.52
S1	.16	.32	.00	.00	.13	.10	.29	.42
S3	.06	.30	.00	.00	.13	.10	.20	.40
S5	.21	.15	.00	.00	.11	.12	.32	.27
PI	.42	.54	.00	.00	.00	.00	.42	.54
NI	.64	.62	.00	.00	.00	.00	.64	.62
Non-decision time parameters								
CR2	.19	.19	.00	.00	.24	.24	.43	.43
CR4	.36	.31	.00	.00	.24	.26	.60	.57
S1	.14	.31	.00	.00	.00	.00	.14	.31
S3	.36	.45	.00	.00	.00	.00	.36	.45
S5	.43	.43	.00	.00	.00	.00	.43	.43
PI	.60	.54	.00	.00	.00	.00	.60	.54
NI	.41	.34	.00	.00	.00	.00	.41	.34

CRT task. The LST model provided a good fit for the data, $\chi^2(7) = 6.90$, $p = .440$, CFI = 1, RMSEA = .00 [.00; .11]. However, the variances of the latent state residuals and of the method factor for the four choice condition were negative and/or non-significant ($VAR(SR_1) = -0.03$, $p = .659$; $VAR(SR_2) = 0.05$, $p = .413$; $VAR(CR4_v) = 0.01$, $p = .948$). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(10) = 7.67$, $p = .661$, CFI = 1, RMSEA = .01 [.00; .08]. See the upper part of Figure 3A (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

Sternberg memory scanning task. The LST model provided an acceptable fit for the data, $\chi^2(18) = 26.53$, $p = .088$, CFI = .96, RMSEA = .06 [.00; .11]. However, the variances of the latent state residuals and of the method factor for the set size 3 condition were non-significant ($VAR(SR_1) = 0.02$, $p = .748$; $VAR(SR_2) = 0.90$, $p = .099$; $VAR(S3_v) = 0.01$, $p = .851$). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(21) = 31.38$, $p = .068$, CFI = .96, RMSEA = .07 [.00; .11]. See the middle part of Figure 3A (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

Posner letter matching task. The LST model provided a good fit for the data, $\chi^2(7) = 7.44$, $p = .385$, CFI = 1, RMSEA = .02 [.00; .12]. However, the variances of the latent state residuals and of the method factor for the physical identity condition were non-significant ($VAR(SR_1) = 0.06$, $p = .301$; $VAR(SR_2) = 0.12$, $p = .063$; $VAR(PI_v) = 0.07$, $p = .322$). Hence, these variances were fixed to zero. Afterwards, model fit was still acceptable, $\chi^2(10) = 16.08$, $p = .097$, CFI = .97, RMSEA = .07 [.00; .14]. See the lower part of Figure 3A (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

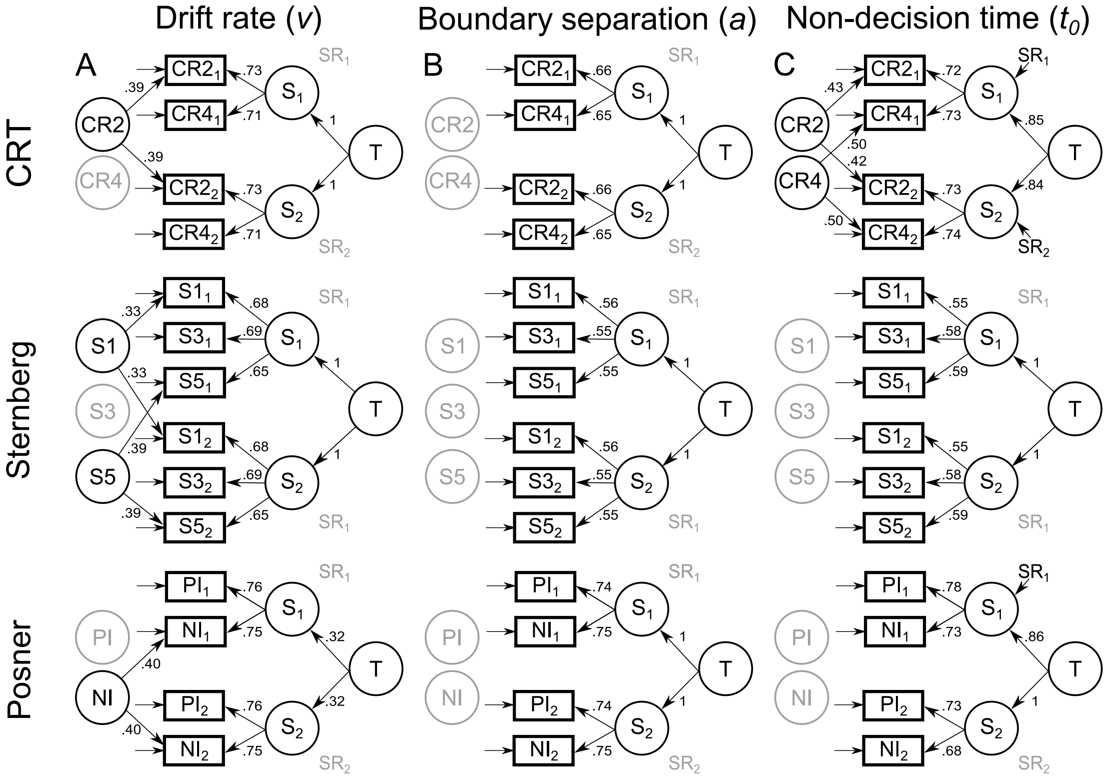

Figure 3. Seperate LST models for the three parameters of the diffusion model: drift rate v, boundary separation a, and non-decision time t_0—estimated for each of the three tasks.

3.2.2. Boundary Separation

Manifest correlations between boundary-separation parameters across tasks and measurements are shown in Table A2 in Appendix B. The LST model for boundary separation across the three experimental tasks provided a good fit for the data, $\chi^2(94)$ = 124.78, p = .019, CFI = .94, RMSEA = .05 [.02; .08]. Because the variances of the latent state residuals and of the method factor for the Posner letter matching task were negative and/or non-significant again ($VAR(SR_1) = -0.04$, p = .181; $VAR(SR_2) = 0.10$, p = .156; $VAR(P_a) = 0.08$, p = .227), we also fixed these variances to zero. These modifications did not impair model fit, $\chi^2(97)$ = 134.26, p = .007, CFI = .93, RMSEA = .06 [.03; .08]. Again, the χ^2 test indicated significant differences between the implied and empirical covariance structure, but the CFI and the RMSEA indicated an acceptable model fit [54,55]. See Figure 4 for the modified model.

Coefficients of reliability, consistency, occasion-specificity, and method-specificity are shown in the middle part of Table 2. Consistencies of boundary separation parameters were low to moderate and highest in the Posner letter matching task. Methods-specificities were low in the SRT, CRT, and Sternberg memory scanning task, whereas method factors explained no variance in boundary separation parameters in the Posner letter matching task. While the temporal stabilities were as great for the latent boundary separation states as for the latent drift rate states, the reliabilities of manifest boundary separation parameters were notably lower ranging from .27 to .64.

Table 3. LST parameters for the LST models by task (see Figure 3). Cond. = Condition; Occ. Spec. = Occasion-specificity; Meth. Spec. = Method-specificity; Rel. = Reliability; Boundary sep. = Boundary separation; Non-dec. time = Non-decision time.

Task	dm Parameter	Cond.	MP	Cons.	O. Spec.	M. Spec	Rel.
CRT	Drift rate v	CR2	1	.53	.00	.15	.68
		CR4	1	.51	.00	.00	.51
		CR2	2	.53	.00	.15	.68
		CR4	2	.51	.00	.00	.51
	Boundary sep. a	CR2	1	.43	.00	.00	.43
		CR4	1	.43	.00	.00	.43
		CR2	2	.43	.00	.00	.43
		CR4	2	.43	.00	.00	.43
	Non-dec. time t_0	CR2	1	.38	.14	.18	.70
		CR4	1	.39	.15	.25	.79
		CR2	2	.37	.16	.18	.71
		CR4	2	.38	.16	.25	.79
Sternberg	Drift rate v	S1	1	.46	.00	.11	.57
		S3	1	.47	.00	.00	.47
		S5	1	.42	.00	.15	.57
		S1	2	.46	.00	.11	.57
		S3	2	.47	.00	.00	.47
		S5	2	.42	.00	.15	.57
	Boundary sep. a	S1	1	.31	.00	.00	.31
		S3	1	.30	.00	.00	.30
		S5	1	.30	.00	.00	.30
		S1	2	.31	.00	.00	.31
		S3	2	.30	.00	.00	.30
		S5	2	.30	.00	.00	.30
	Non-dec. time t_0	S1	1	.31	.00	.00	.31
		S3	1	.34	.00	.00	.34
		S5	1	.35	.00	.00	.35
		S1	2	.31	.00	.00	.31
		S3	2	.34	.00	.00	.34
		S5	2	.35	.00	.00	.35
Posner	Drift rate v	PI	1	.58	.00	.00	.58
		NI	1	.57	.00	.16	.72
		PI	2	.58	.00	.00	.58
		NI	2	.57	.00	.16	.72
	Boundary sep. a	PI	1	.55	.00	.00	.55
		NI	1	.57	.00	.00	.57
		PI	2	.55	.00	.00	.55
		NI	2	.57	.00	.00	.57
	Non-dec. time t_0	PI	1	.44	.16	.00	.60
		NI	1	.39	.14	.00	.54
		PI	2	.52	.00	.00	.52
		NI	2	.46	.00	.00	.46

CRT task. The LST model provided a good fit for the data, $\chi^2(7) = 2.43$, $p = .933$, CFI = 1, RMSEA = .00 [.00; .03]. However, the variances of the latent state residuals and of the method factors were non-significant ($VAR(SR_1) = 0.04$, $p = .536$; $VAR(SR_2) = 0.11$, $p = .134$; $VAR(CR2_v) = 0.08$, $p = .303$; $VAR(CR4_v) = 0.15$, $p = .079$). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(11) = 8.38$, $p = .679$, CFI = 1, RMSEA = .00 [.00; .08]. See the upper part of Figure 3B (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

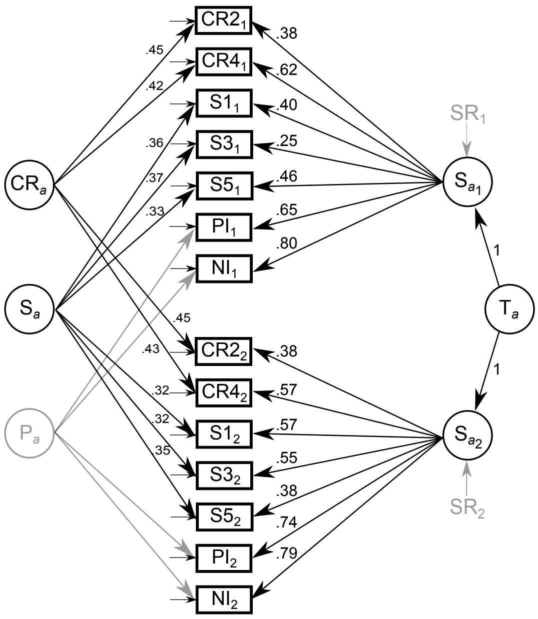

Figure 4. The latent state-trait model of boundary separation parameters consists of a common trait T, a state residual SR_i for each of the two measurement occasions i, and a method factor M_j for each of the three experimental tasks. CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity. Latent variables displayed in gray were non-significant.

Sternberg memory scanning task. The LST model provided an acceptable fit for the data, $\chi^2(18) = 24.78$, $p = .131$, CFI = .94, RMSEA = .06 [.00; .11]. However, the variances of the latent state residuals and of the method factors were non-significant and/or negative ($VAR(SR_1) = -0.04$, $p = .465$; $VAR(SR_2) = 0.11$, $p = .100$; $VAR(S1_v) = 0.08$, $p = .330$; $VAR(S3_v) = -0.09$, $p = .247$; $VAR(S5_v) = 0.05$, $p = .529$). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(23) = 31.51$, $p = .111$, CFI = .93, RMSEA = .06 [.00; .10]. See the middle part of Figure 3B (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

Posner letter matching task. The LST model provided a good fit for the data, $\chi^2(7) = 5.52$, $p = .597$, CFI = 1, RMSEA = .00 [.00; .10]. However, the variances of the latent state residuals and of the method factors were negative and/or non-significant ($VAR(SR_1) = -0.03$, $p = .652$; $VAR(SR_2) = 0.05$, $p = .429$; $VAR(PI_v) = -0.11$, $p = .071$, $VAR(NI_v) = 0.13$, $p = .079$). Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(11) = 11.55$, $p = .398$, CFI = 1, RMSEA = .02 [.00; .10]. See the lower part of Figure 3B (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

3.2.3. Non-Decision Time

Manifest correlations for non-decision time parameters across tasks and measurements are shown in Table A3 in Appendix B. The LST model for non-decision time provided a mediocre fit for the data, $\chi^2(94) = 274.01$, $p < .001$, CFI = .76, RMSEA = .13 [.11; .15]. Because the variances of the latent state residuals and of the method factor for the Sternberg and the Posner letter matching task were non-significant or negative ($VAR(SR_1) = 0.02$, $p = .667$; $VAR(SR_2) = -0.02$, $p = .718$; $VAR(S_{t0}) = -0.06$, $p = .053$; $VAR(P_{t0}) = 0.03$, $p = .542$), we also fixed these variances to zero. These modifications did not impair model fit, $\chi^2(98) = 292.27$, $p < .001$, CFI = .76, RMSEA = .13 [.11; .15]. This time, both the χ^2 test and the CFI and RMSEA indicated significant differences between the implied and empirical covariance structure [54,55]. See Figure 5 for the modified model.

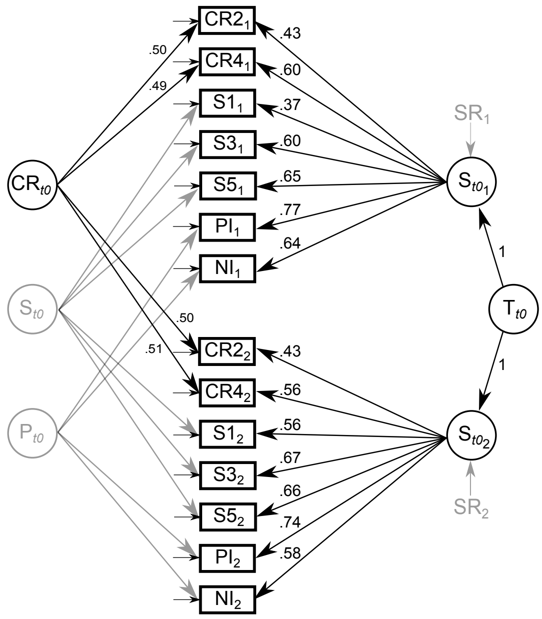

Figure 5. The latent state-trait model of non-decision time parameters consists of a common trait T, a state residual SR_i for each of the two measurement occasions i, and a method factor M_j for each of the three experimental tasks. CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity. Latent variables displayed in gray were non-significant.

Coefficients of reliability, consistency, occasion-specificity, and method-specificity are shown in the lower part of Table 2. Consistencies of non-decision time parameters were overall low to moderate and highest in the Posner letter matching task. We observed low method-specificities only in the SRT/CRT task; for the other two tasks, we found no effect of task-specific characteristics. Again, the temporal stabilities of the latent states were great, but reliabilities of the manifest non-decision time parameters were low to moderate, ranging from .14 to .60.

CRT task. The LST model provided a good fit for the data, $\chi^2(7) = 2.21$, $p = .947$, CFI = 1, RMSEA = .00 [.00; .01]. See the upper part of Figure 3C (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

Sternberg memory scanning task. The LST model did not provide an acceptable fit for the data, $\chi^2(18) = 62.32$, $p < .001$, CFI = .78, RMSEA = .15 [.11; .19]. The variances of the latent state residuals and of the method factors were non-significant and/or negative ($VAR(SR_1) = -0.06$, $p = .316$; $VAR(SR_2) = 0.05$, $p = .473$; $VAR(S1_v) = -0.07$, $p = .377$; $VAR(S3_v) = -0.24$, $p = .002$; $VAR(S5_v) = -0.17$, $p = .016$). Hence, these variances were fixed to zero. These modifications did not change model fit to a great degree, $\chi^2(23) = 83.49$, $p < .001$, CFI = .67, RMSEA = .15 [.12; .19]. See the middle part of Figure 3C (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

Posner letter matching task. The LST model provided a good fit for the data, $\chi^2(7) = 1.77$, $p = .971$, CFI = 1, RMSEA = .00 [.00; .10]. However, the variances of the latent state residual reflecting the second measurement occasion and of the method factor for the name identity condition were negative and/or non-significant ($VAR(SR_2) = 0.13$, $p = .086$; $VAR(NI_v) = -0.10$, $p = .145$. Hence, these variances were fixed to zero. These modifications did not impair model fit, $\chi^2(9) = 8.13$, $p = .521$, CFI = 1, RMSEA = .00 [.00; .10]. Now, however, the method factor for the physical identity condition was non-significant, $VAR(PI_v) = 0.11$, $p = .110$, and was subsequently fixed to zero. The final model still provided a good fit for the data, $\chi^2(10) = 10.99$, $p = .359$, CFI = .99, RMSEA = .03 [.00; .11]. See the lower part of Figure 3C (p. 12) for the final model and Table 3 (p. 13) for the associated LST parameters.

3.3. Discussion

The present study evaluated whether diffusion model parameters (drift rate, boundary separation, non-decision time) qualify as trait-like variables that may be considered temporally stable person properties of cognitive processes. For this purpose, we used LST models to assess the amount of variance in model parameters that can be attributed to a common trait. If diffusion model parameters from elementary cognitive tasks are to be treated as trait-like variables in individual differences research, they have to be stable across time and consistent over situations [38].

First, we evaluated the temporal stability of the latent drift rate, boundary separation, and non-decision time states across a time period of eight months in three tasks. The variance of model parameters consistent across tasks was not affected by situational influences or influences of person-situation interactions such as fatigue, motivation, or familiarization with the testing environment, as reflected in the result that the state residuals in all LST models could be fixed to zero without impairing model fit. These results are consistent with previous research on the reliability of diffusion model parameters reporting moderate to good test-retest correlations over a period of one week [30,41]. They support the notion that the variances of diffusion model parameters that are consistent across tasks can be considered trait-like properties of cognitive processes. Moreover, because their temporal stability is comparable to the temporal stability of intelligence tests [13,42,51], these results are consistent with the view that individual differences in diffusion model parameters reflect elementary person properties that may give rise to general intelligence.

Second, we evaluated the consistency of diffusion model parameters across three elementary cognitive tasks, which showed a great variability. Only manifest (note that in this context "manifest" refers to the estimates of the latent diffusion model parameters that were entered into structural equation models as manifest variables) drift rate variables showed a moderate consistency, whereas the coefficients of consistency of boundary separation and non-decision time parameters were rather low. On average, 44 percent of the variance of each manifest drift rate parameter was accounted for by the common trait. Based on these results, individual differences in drift rate can be considered as a temporally stable and trans-situationally consistent trait across different tasks, as they are likely to reflect an ability of a person. Nevertheless, the consistencies of drift rate parameters were still lower than the consistencies typically observed in intelligence tests [51], which is why we suggest administering more than one response time task in future studies on individual differences in drift rates to reliably capture the common variance across tasks.

In comparison, on average only 32 to 36 percent of the variance of each manifest boundary separation and non-decision time parameter was accounted for by the common trait across the three tasks. Moreover, the LST model of non-decision times did not even provide a good account for the data due to the complex factorial structure with several low covariances between manifest parameters. These results suggest that while latent variables of boundary separation and non-decision times showed a great temporal stability, the manifest parameters are likely to reflect mostly narrower and more task-specific skills as well as measurement error. Hence, an individual does not have *one* boundary separation or *one* non-decision time. Instead, an individual has task-specific boundary separation and non-decision time parameters that are only weakly correlated across tasks. This result is most intuitive for the non-decision time parameter, which reflects the speed of different non-decisional processes such as encoding, memory retrieval, or response execution [31]. As such, it is not a process-pure parameter. Individual differences in non-decision time may thus reflect mostly individual differences in encoding speed in one kind of task, and individual differences in response execution in another kind of task.

Third, we repeated these analyses separately for each task and each model parameter. While the results of these task-specific analyses were largely consistent with the results of the LST models across tasks, we observed that consistencies for all model parameters tended to increase when evaluated at a task-specific level. This finding supports the notion that model parameters consist to a considerable degree of task-specific variance. Interestingly, we found that even at a task-specific level a common trait model was not suited to describe the non-decision times in the Sternberg memory scanning task as reflected in the bad model fits. Correlations between non-decision times in the set size three and five conditions were substantially greater than correlations between non-decision times in the set size one condition and any of the other conditions, as can be seen in Table A3 in Appendix B. We believe that the additional short-term memory access demands of any set size larger than one are reflected in the non-decision time parameter [28,31], which explains why non-decision time parameters from these conditions are less strongly related to a non-decision time parameter from a condition without short-term memory access. Again, this result suggests that non-decision time parameters should not be treated as trait-like variables, as they contain different sources of variation. On a side note, we would like to point out that the model specifications of the task-specific analyses allow a weak test of relative measurement invariance, as all paths were fixed to one and error variances for each condition were constrained to be equal across sessions. This weak test of measurement invariance suggests that at least the assumption of relative measurement invariance holds for the majority of diffusion model parameters, as indicated by the acceptable model fits under these model specifications except for the LST model of non-decision times in the Sternberg memory scanning task.

4. Conclusions

All in all, our analysis of the psychometric properties of diffusion model parameters may help understanding why only drift rate parameters have been consistently positively associated with mental abilities, whereas the associations between boundary separation and non-decision time parameters have been inconsistent and sometimes even in opposite directions [25–28]. First, drift rate parameters showed overall slightly higher consistencies than boundary separation and non-decision time parameters. Second and more importantly, drift rate parameters had fewer extremely small consistencies than boundary separation and non-decision time parameters (smallest consistencies: .26 (drift rate), .09 (boundary separation), .18 (non-decision time)). Future studies should thus consider using broader batteries of response time tasks to capture the small amount of temporally stable common variance in boundary separation and non-decision time parameters across different tasks and to minimize the effect of tasks with extremely low consistencies, or focus on studying task-specific associations with mental abilities. However, when analyzing only the association between drift rates and mental abilities, a small test battery consisting of only a few response time tasks will be sufficient to reliably estimate the common drift rate trait. Moreover, we caution against treating the non-decision time parameter as a trait-like variable, as it is not a process-pure parameter and contains substantial task-specific sources of variation.

There are several limitations to the present study. First, we only used elementary cognitive tasks, which are relatively simple response time tasks with low error rates. Whether cognitively more demanding tasks such as recognition memory or lexical decision tasks yield comparable results is an open question. Second, we applied the diffusion model to the CRT task, although it is not clear whether the decision process in this task reflects a diffusion process or whether it requires only a spatial identification of the stimulus position. There is some evidence that the decision process reflects at least an evidence accumulation process, as Leite and Ratcliff [46] successfully fitted sequential sampling models to data from a similar task. Moreover, factor loadings of the structural equation models and the resulting consistencies in Figures 2–5 and Table 2 do not suggest that the CRT task stands out from the rest of the tasks. Nevertheless, we do not presume that the diffusion model captures the *true* diffusion process, but rather that it provides a simplifying description related to the true process. As such, it describes the data in terms of parameters that are likely related to a substantial (but unknown) degree to the true parameters characterizing evidence accumulation, boundary separation, and non-decision time. This important distinction between the true process and its description in terms of model parameters has important implications for the interpretation of the resulting parameters estimates. In particular, other process models may be better suited to describe the cognitive processes involved in the CRT task than the diffusion model. It would be interesting to explore the convergent validity of diffusion model parameters with conceptually related parameters from sequential sampling models in the CRT task in future studies. Third, we only used the Kolmogorov-Smirnov statistic to estimate diffusion model parameters and previous research has suggested that other estimation algorithms such as the χ^2 statistic and the maximum likelihood estimation may yield different test-retest correlations [41]. Fourth, we did not evaluate the temporal stability of manifest variables on a latent level, because an odd-even split of trials within each condition was not feasible without affecting the stability of parameter estimates due to relatively low trial numbers. Hence, we do not know to what degree manifest variables reflect task-specific skills and to what extent they reflect measurement error.

Taken together, our results show that the consistent variance of diffusion model parameters across tasks can be regarded as temporally stable ability parameters. We have shown that only the drift rate parameter can be regarded as a trait-like variable that is stable across time and consistent over different tasks [38]. Because the diffusion model allows an elegant, process-pure measurement of the speed of information uptake with trait-like characteristics, we believe that the mathematical modeling of response times provides a promising avenue for identifying the cognitive processes giving rise to individual differences in general intelligence.

Author Contributions: Dirk Hagemann, Anna-Lena Schubert, and Andreas Voss designed the experiment; Anna-Lena Schubert performed the experiment; Anna-Lena Schubert and Gidon T. Frischkorn analyzed the data; Anna-Lena Schubert wrote the paper; Gidon T. Frischkorn, Dirk Hagemann, and Andreas Voss gave conceptual and technical advice. All authors discussed the results and commented on the paper.

Conflicts of Interest: All authors declare no conflict of interest.

Appendix A. QQ-Plot Evaluating the Fit of Diffusion Model Parameters

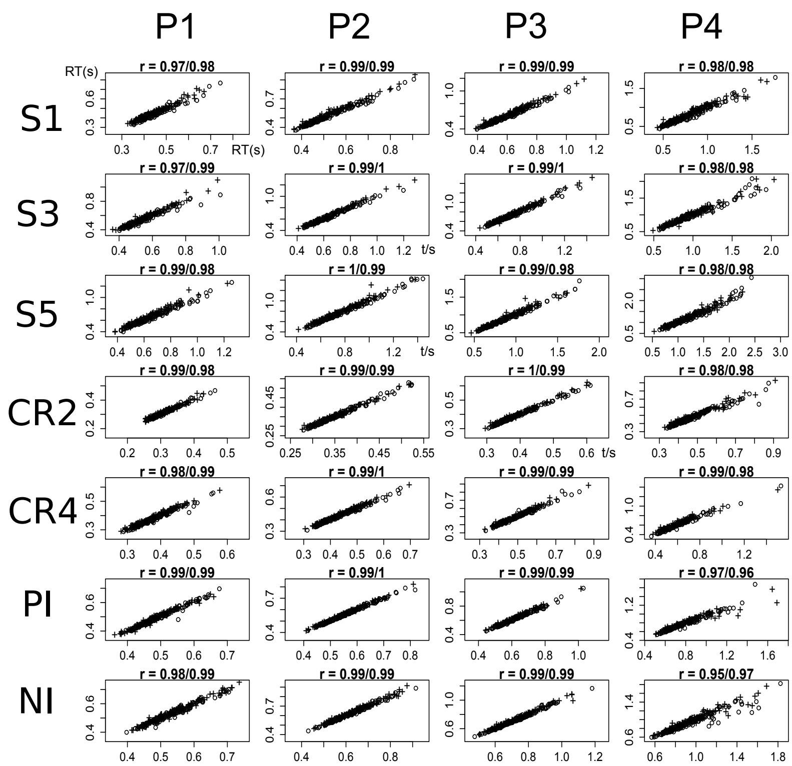

Figure A1. Correlations between empirical and predicted mean response times in seconds across four percentiles (P1 to P4) after the removal of outliers in all tasks. Dots represent mean response times at the first laboratory session and crosses represent mean response times at the second laboratory session.

Appendix B. Correlation Tables for Diffusion Model Parameters Across Measurement Points

Table A1. Product–moment correlations between drift rate parameters at the first and second laboratory session.

		Session 1							Session 2						
		CR2	**CR4**	**S1**	**S3**	**S5**	**PI**	**NI**	**CR2**	**CR4**	**S1**	**S3**	**S5**	**PI**	**NI**
	CR2	1	0.43	0.28	0.35	0.25	0.51	0.55	0.60	0.47	0.51	0.45	0.47	0.48	0.56
	CR4		1	0.26	0.37	0.21	0.31	0.45	0.39	0.44	0.42	0.43	0.33	0.35	0.41
	S1			1	0.41	0.30	0.37	0.37	0.37	0.22	0.53	0.42	0.32	0.23	0.23
Ses. 1	S3				1	0.31	0.44	0.40	0.30	0.15	0.43	0.52	0.35	0.39	0.40
	S5					1	0.38	0.45	0.25	0.26	0.30	0.53	0.53	0.32	0.44
	PI						1	0.56	0.35	0.36	0.56	0.53	0.42	0.60	0.57
	NI							1	0.39	0.37	0.46	0.56	0.53	0.49	0.71
	CR2								1	0.64	0.52	0.43	0.38	0.37	0.30
	CR4									1	0.36	0.30	0.32	0.35	0.31
	S1										1	0.60	0.50	0.49	0.51
Ses. 2	S3											1	0.59	0.59	0.66
	S5												1	0.47	0.60
	PI													1	0.70
	NI														1

Note: Ses.1/2 = measures at measurement point one or two; CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity.

Table A2. Product–moment correlations between boundary separation parameters at the first and second laboratory session.

		Session 1							Session 2						
		CR2	**CR4**	**S1**	**S3**	**S5**	**PI**	**NI**	**CR2**	**CR4**	**S1**	**S3**	**S5**	**PI**	**NI**
	CR2	1	0.39	0.06	0.15	0.26	0.22	0.17	0.46	0.33	0.40	0.26	0.15	0.18	0.13
	CR4		1	0.02	0.12	0.24	0.42	0.54	0.41	0.51	0.46	0.30	0.22	0.48	0.49
	S1			1	0.09	0.11	0.07	0.23	0.18	0.19	0.20	0.14	0.05	0.16	0.07
Ses. 1	S3				1	0.15	0.09	0.20	0.24	0.07	0.23	0.22	0.22	0.23	0.15
	S5					1	0.29	0.45	0.25	0.35	0.32	0.47	0.34	0.24	0.36
	PI						1	0.56	0.27	0.33	0.43	0.36	0.22	0.47	0.53
	NI							1	0.32	0.33	0.39	0.55	0.24	0.57	0.60
	CR2								1	0.50	0.48	0.28	0.20	0.29	0.35
	CR4									1	0.52	0.33	0.26	0.49	0.48
	S1										1	0.55	0.33	0.45	0.44
Ses. 2	S3											1	0.45	0.38	0.47
	S5												1	0.22	0.39
	PI													1	0.57
	NI														1

Note: Ses.1/2 = measures at measurement point one or two; CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity.

Table A3. Product–moment correlations between non-decision time parameters at the first and second laboratory session.

		Session 1							Session 2						
		CR2	CR4	S1	S3	S5	PI	NI	CR2	CR4	S1	S3	S5	PI	NI
Ses. 1	CR2	1	0.48	0.25	0.31	0.28	0.39	0.41	0.26	0.28	0.22	0.21	0.24	0.33	0.27
	CR4		1	0.17	0.33	0.37	0.48	0.44	0.17	0.26	0.33	0.43	0.28	0.34	0.40
	S1			1	0.16	0.11	0.28	0.21	0.32	0.27	0.44	0.22	0.13	0.31	0.22
	S3				1	0.61	0.42	0.29	0.30	0.18	0.33	0.55	0.53	0.23	0.20
	S5					1	0.54	0.34	−0.01	0.03	0.23	0.61	0.54	0.29	0.34
	PI						1	0.59	0.21	0.38	0.40	0.56	0.45	0.63	0.63
	NI							1	0.33	0.36	0.31	0.35	0.20	0.55	0.56
Ses. 2	CR2								1	0.99	0.12	−0.03	−0.06	−0.09	−0.22
	CR4									1	0.15	0 .01	−0.02	−0.04	−0.14
	S1										1	0.49	0.35	0.28	0.37
	S3											1	0.53	0.29	0.34
	S5												1	0.27	.30
	PI													1	0.66
	NI														1

Note.: Ses.1/2 = measures at measurement point one or two; CR2/4 = choice response time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size five; PI = physical identity; NI = name identity.

References

1. Sheppard, L.D.; Vernon, P.A. Intelligence and speed of information-processing: A review of 50 years of research. *Personal. Individ. Differ.* **2008**, *44*, 535–551.
2. Kranzler, J.H.; Jensen, A.R. The nature of psychometric *g*: Unitary process or a number of independent processes? *Intelligence* **1991**, *15*, 397–422.
3. Miller, L.T.; Vernon, P.A. Intelligence, reaction time, and working memory in 4- to 6-year-old children. *Intelligence* **1996**, *22*, 155–190.
4. Saccuzzo, D.P.; Larson, G.E.; Rimland, B. *Speed of Information Processing and Individual Differences In Intelligence*; Research Report NPRDC TR 86-23; Navy Personnel Research and Development Center: San Diego, CA, USA, 1986.
5. Jensen, A.R. The importance of intraindividual variation in reaction time. *Personal. Individ. Differ.* **1992**, *13*, 869–881.
6. Jensen, A.R. *Clocking the Mind: Mental Chronometry and Individual Differences*; Elsevier: Amsterdam, The Netherlands, 2006.
7. Doebler, P.; Scheffler, B. The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learn. Individ. Differ.* **2015**, doi:10.1016/j.lindif.2015.02.009.
8. Larson, G.E.; Alderton, D.L. Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence* **1990**, *14*, 309–325.
9. Coyle, T.R. A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence* **2003**, *31*, 567–587.
10. Fernandez, S.; Fagot, D.; Dirk, J.; de Ribaupierre, A. Generalization of the worst performance rule across the lifespan. *Intelligence* **2014**, *42*, 31–43.
11. Heathcote, A.; Popiel, S.J.; Mewhort, D.J. Analysis of response time distributions: An example using the Stroop task. *Psychol. Bull.* **1991**, *109*, 340–347.
12. Spieler, D.H.; Balota, D.A.; Faust, M.E. Levels of selective attention revealed through analyses of response time distributions. *J. Exp. Psychol. Hum. Percept. Perform.* **2000**, *26*, 506–526.
13. Carroll, J.B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 1993.
14. Smith, G.A.; Carew, M. Decision time unmasked: Individuals adopt different strategies. *Aust. J. Psychol.* **1987**, *39*, 339–351.
15. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **1978**, *85*, 59–108.
16. Voss, A.; Rothermund, K.; Voss, J. Interpreting the parameters of the diffusion model: An empirical validation. *Memory Cogn.* **2004**, *32*, 1206–1220.
17. Ratcliff, R.; Spieler, D.; McKoon, G. Explicitly modeling the effects of aging on response time. *Psychon. Bull. Rev.* **2000**, *7*, 1–25.
18. Voss, A.; Nagler, M.; Lerche, V. Diffusion models in experimental psychology: A practical introduction. *Exp. Psychol.* **2013**, *60*, 385–402.
19. Voss, A.; Voss, J. Fast-dm: A free program for efficient diffusion model analysis. *Behav. Res. Methods* **2007**, *39*, 767–775.
20. Vandekerckhove, J.; Tuerlinckx, F. Fitting the Ratcliff diffusion model to experimental data. *Psychon. Bull. Rev.* **2007**, *14*, 1011–1026.
21. Vandekerckhove, J.; Tuerlinckx, F. Diffusion model analysis with MATLAB: A DMAT primer. *Behav. Res. Methods* **2008**, *40*, 61–72.
22. Wagenmakers, E.J.; van der Maas, H.L.J.; Grasman, R.P.P.P. An EZ-diffusion model for response time and accuracy. *Psychon. Bull. Rev.* **2007**, *14*, 3–22.
23. Nunez, M.D.; Srinivasan, R.; Vandekerckhove, J. Individual differences in attention influence perceptual decision making. *Front. Psychol.* **2015**, *8*, 18.
24. Stahl, C.; Voss, A.; Schmitz, F.; Nuszbaum, M.; Tuscher, O.; Lieb, K.; Klauer, K.C. Behavioral components of impulsivity. *J. Exp. Psychol. Gen.* **2014**, *143*, 850–886.
25. Ratcliff, R.; Thapar, A.; Mckoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cogn. Psychol.* **2010**, *60*, 127–157.

26. Ratcliff, R.; Thapar, A.; Mckoon, G. Effects of aging and IQ on item and associative memory. *J. Exp. Psychol. Gen.* **2011**, *140*, 464–487.
27. Schmiedek, F.; Oberauer, K.; Wilhelm, O.; Suss, H.M.; Wittmann, W.W. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J. Exp. Psychol. Gen.* **2007**, *136*, 414–429.
28. Schubert, A.L.; Hagemann, D.; Voss, A.; Schankin, A.; Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **2015**, *51*, 28–46.
29. Ratcliff, R.; Thompson, C.A.; Mckoon, G. Modeling individual differences in response time and accuracy in numeracy. *Cognition* **2015**, *137*, 115–136.
30. Yap, M.J.; Balota, D.A.; Sibley, D.E.; Ratcliff, R. Individual differences in visual word recognition: Insights from the English Lexicon Project. *J. Exp. Psychol. Hum. Percept. Perform.* **2012**, *38*, 53–79.
31. Ratcliff, R.; Mckoon, G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Comput.* **2008**, *20*, 873–922.
32. Molenaar, D.; Tuerlinckx, F.; van der Maas, H.L.J. Fitting Diffusion Item Response Theory Models for Responses and Response Times Using the R Package diffIRT. *J. Stat. Softw.* **2015**, *66*, doi:10.18637/jss.v066.i04.
33. Tuerlinckx, F.; Boeck, P.D. Two interpretations of the discrimination parameter. *Psychometrika* **2005**, *70*, 629–650.
34. Van der Maas, H.L.J.; Molenaar, D.; Maris, G.; Kievit, R.A.; Borsboom, D. Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychol. Rev.* **2011**, *118*, 339–356.
35. Vandekerckhove, J. A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *J. Math. Psychol.* **2014**, *60*, 58–71.
36. Allport, G.W. *Personality, a Psychological Interpretation*; Holt: New York, NY, USA, 1937.
37. Cattell, R. *Personality: A Systematic Theoretical and Factual Study*; McGraw-Hill: New York, NY, USA, 1950.
38. Steyer, R.; Schmitt, M.; Eid, M. Latent state–trait theory and research in personality and individual differences. *Eur. J. Personal.* **1999**, *13*, 389–408.
39. Langner, R.; Steinborn, M.B.; Chatterjee, A.; Sturm, W.; Willmes, K. Mental fatigue and temporal preparation in simple reaction-time performance. *Acta Psychol.* **2010**, *133*, 64–72.
40. Doll, T.J. Motivation, reaction time, and the contents of active verbal memory. *J. Exp. Psychol.* **1971**, *87*, 29–36.
41. Lerche, V.; Voss, A. Retest reliability of the parameters of the ratcliff diffusion model. *Psychol. Res.* **2016**, *2016*, 1–24.
42. Larsen, L.; Hartmann, P.; Nyborg, H. The stability of general intelligence from early adulthood to middle-age. *Intelligence* **2008**, *36*, 29–34.
43. Hick, W.E. On the rate of gain of information. *Q. J. Exp. Psychol.* **1952**, *4*, 11–26.
44. Brown, S.D.; Heathcote, A. The simplest complete model of choice response time: Linear ballistic accumulation. *Cogn. Psychol.* **2008**, *57*, 153–178.
45. Usher, M.; McClelland, J.L. The time course of perceptual choice: The leaky, competing accumulator model. *Psychol. Rev.* **2001**, *108*, 550–592.
46. Leite, F.P.; Ratcliff, R. Modeling reaction time and accuracy of multiple-alternative decisions. *Atten. Percept. Psychophys.* **2010**, *72*, 246–273.
47. Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *Am. Sci.* **1969**, *57*, 421–457.
48. Posner, M.I.; Mitchell, R.F. Chronometric analysis of classification. *Psychol. Rev.* **1967**, *74*, 392–409.
49. Steyer, R.; Ferring, D.; Schmitt, M.J. States and traits in psychological assessment. *Eur. J. Psychol. Assess.* **1992**, *8*, 79–98.
50. Steyer, R.; Mayer, A.; Geiser, C.; Cole, D.A. A theory of states and traits–revised. *Annu. Rev. Clin. Psychol.* **2015**, *11*, 71–98.
51. Danner, D.; Hagemann, D.; Schankin, A.; Hager, M.; Funke, J. Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence* **2011**, *39*, 323–334.
52. Deinzer, R.; Steyer, R.; Eid, M.; Notz, P.; Schwenkmezger, P.; Ostendorf, F.; Neubauer, A. Situational effects in trait assessment: The FPI, NEOFFI, and EPI questionnaires. *Eur. J. Personal.* **1995**, *9*, 1–23.

53. Muthén, L.K.; Muthén, B.O. *Mplus User's Guide*, 7th ed.; Muthén & Muthén: Los Angeles, CA, USA, 1998–2012.
54. Browne, M.W.; Cudeck, R. Alternative Ways of Assessing Model Fit. *Sociol. Methods Res.* **1992**, *21*, 230–258.
55. Hu, L.T.; Bentler, P.M. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model.* **1999**, *6*, 1–55.

Journal of
Intelligence

MDPI

Article

The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis

Gidon T. Frischkorn *, Anna-Lena Schubert, Andreas B. Neubauer and Dirk Hagemann

Institute of Psychology, Heidelberg University, Hauptstr. 47-51, 69117 Heidelberg, Germany;
anna-lena.schubert@psychologie.uni-heidelberg.de (A.-L.S.);
andreas.neubauer@psychologie.uni-heidelberg.de (A.B.N.);
dirk.hagemann@psychologie.uni-heidelberg.de (D.H.)

* Correspondence: gidon.frischkorn@psychologie.uni-heidelberg.de; Tel.: +49-6221 54-7746

Academic Editor: Oliver Wilhelm
Received: 15 March 2016; Accepted: 15 July 2016; Published: 20 July 2016

Abstract: Worst performance in cognitive processing tasks shows larger relationships to general intelligence than mean or best performance. This so called Worst Performance Rule (WPR) is of major theoretical interest for the field of intelligence research, especially for research on mental speed. In previous research, the increases in correlations between task performance and general intelligence from best to worst performance were mostly described and not tested statistically. We conceptualized the WPR as moderation, since the magnitude of the relation between general intelligence and performance in a cognitive processing task depends on the performance band or percentile of performance. On the one hand, this approach allows testing the WPR for statistical significance and on the other hand, it may simplify the investigation of possible constructs that may influence the WPR. The application of two possible implementations of this approach is shown and compared to results of a traditional worst performance analysis. The results mostly replicate the WPR. Beyond that, a comparison of results on the level of unstandardized relationships (e.g., covariances or unstandardized regression weights) to results on the level of standardized relationships (i.e., correlations) indicates that increases in the inter-individual standard deviation from best to worst performance may play a crucial role for the WPR. Altogether, conceptualizing the WPR as moderation provides a new and straightforward way to conduct Worst Performance Analysis and may help to incorporate the WPR more prominently into empirical practice of intelligence research.

Keywords: mental speed; worst performance rule; moderation; general intelligence

1. Introduction

A wealth of research has reported results that support a consistent and moderate to mediocre relationship between mental speed and general intelligence (for a review, see [1]). Across a variety of different tasks measuring mental speed, Sheppard and Vernon [1] reported an average correlation of $r = -.24$ between response times and measures of general intelligence. This indicates that faster speed of information processing is associated with greater general intelligence. Beyond that, this relationship between mental speed and intelligence is not only present in behavioral measures of mental speed but in neural measures of mental speed, such as latencies of event-related potentials, as well [2]. However, most of these results rely on mean scores for mental speed.

Recent empirical results have suggested that inter-individual differences in mean performance on tasks measuring mental speed or memory capacity may not be best suited to predict general intelligence. In fact, it seems that within such tasks worst performance is more indicative for general intelligence [3]. In detail, performance in various processing speed or memory tasks was ranked from fastest to slowest reaction times (RTs) or best to worst memory recall. Then, means of different RT bands or best and worst memory performance were correlated with a measure of general intelligence. The absolute size of these correlations mostly increased from best to worst performance, suggesting that worst performance is more closely related to general intelligence than mean or best performance. This phenomenon is called the worst performance rule (WPR) [4]. Although a few studies replicated this phenomenon [5–10], it has yet to be acknowledged adequately in the field of intelligence research.

In previous research, the analysis of the WPR, or so called worst performance analysis (WPA), mostly described increases in correlations between performance in cognitive processing tasks and general intelligence from best to worst performance instead of providing an adequate statistical test for this phenomenon (for a detailed review, see [3]). The present work conceptualizes the WPR as moderation and introduces new approaches to analyze and test the WPR for statistical significance. These analyses try to overcome the rather descriptive approach of formerly published WPA and offer new possibilities to search for empirical foundations of the WPR. The here presented analysis may hence constitute a useful step towards an accurate test for the WPR that may in turn help to better distinguish between different theoretical explanations for the WPR in future empirical research.

1.1. The Phenomenon of the Worst Performance Rule

The WPR was first explicitly described by Larson and Alderton [4] who showed that correlations of general intelligence with RTs in a simple reaction time paradigm increased from best ($r_{BP} = -.20$) to worst performance ($r_{WP} = -.37$). This initiated a number of conceptually associated studies [5–11] with different related tasks of which all but one [11] reproduced the basic phenomenon of the WPR.

Interestingly, the WPR did not only occur in speeded processing tasks but in non-speeded tasks as well [7,8]. In a multi-trial word recall task, the number of words recalled in worst performance trials correlated more strongly with general intelligence ($r_{WP} = .38$) than in best performance trials ($r_{BP} = .13$). The effect size *Cohen's* q [12] for this difference in correlations was $q = .27$. Moreover, even the number of memorizing strategies in worst performance trials showed a higher correlation with general intelligence in worst performance trials ($r = .24$) than in best performance trials ($r = .12$, $q = .11$). However, only minimal recall performance (i.e., worst performance) predicted general intelligence, and strategy use showed no incremental validity [7].

With these results, the WPR questions some of the core assumptions within theories of intelligence [3]. In particular, the WPR shows that it may not be the inter-individual difference in average performance that depicts differences in intelligence best, rather inter-individual differences in worst performance seem to be most predictive for general intelligence. Although the standard deviation of intra-individual reaction time distributions (RTsd) has been discussed as an additional and supposedly more valid predictor for general intelligence (e.g., the oscillation theory [13–15]), a recent meta-analysis has shown that mean RT and the RTsd are equally valid predictors for general intelligence [16]. In addition, the magnitude of the WPR relies on the g-loading of a task [3]. This suggests that processes fundamental to the WPR may well be processes fundamental to g [17]. Therefore, the WPR is an interesting phenomenon that needs to be studied further.

Theoretically, two main approaches to understand the WPR have been suggested so far. Either worst performance more strongly reflects the speed of information accumulation [18,19] or worst performance trials occur when a person has lapses in attention resulting in longer reaction times [20,21]. Both approaches acknowledge that worst performance trials may contain information on processes that are not adequately represented in mean performance. In fact, the phenomenon of the WPR may be one of the possibilities to shed light on interactions between different cognitive processes (e.g., information accumulation and attentional control), because the WPR does not necessarily represent a cognitive process on its own, but it may occur in the interplay of different cognitive processes.

Beyond that, some methodological explanations have been suggested that may explain the WPR [3]. Specifically, Coyle [3] discussed five possible methodological explanations: (1) the role of outliers; (2) variance compression in best performance; (3) skewness of the intra-individual performance distribution; (4) differences in measurement reliability between best and worst performance; and (5) trial novelty as confound of worst performance. Altogether, Coyle [3] concluded that none of these explanations sufficiently explains the phenomenon of the WPR.

In sum, the WPR seems to be a rather robust phenomenon regarding the relationship between mental speed and intelligence and is of major theoretical interest for research on intelligence, specifically for insights on processes fundamental to *g*.

1.2. Analyzing the Worst Performance Rule

The studies that investigated the WPR so far commonly used a stepwise procedure. First, intra-individual performance in a cognitive processing task was ranked within each person. Then the intra-individual distribution was separated either into performance bands consisting of a specific number of trials per band [4,6], or into percentiles (e.g., [9,11]). Finally, the mean or median for each performance band or percentile was correlated with a measure of general intelligence. The size of these correlations usually increased with ascending performance bands. Yet, this is merely a description of the WPR. An actual test for the significance of the WPR has hardly been reported.

Infrequently the increases in correlations were tested for statistical significance with a rank-correlation between number of performance band and the correlation in the respective performance band [4,5]. This tests whether correlations increase consistently across performance bands or percentiles (quantified in the size of R^2). However, this does neither quantify the slope nor the shape of the increases across performance bands or percentiles and therefore the rank-correlation does not allow to compare the magnitude and shape of the WPR across different conditions or tasks. Additionally, this test treats the estimated correlations as manifest and enters them as discrete values into a new analysis, namely the rank-correlation, that does not account for the uncertainty of the estimation. Thus, the significance of this test may be over-estimated [22].

Alternatively, correlations of best, mean and worst performance bands may be compared with a Fisher's *Z*-test [12,23]. However, this was rarely implemented in studies on the WPR, presumably because the test for differences in correlations lacks statistical power [12]. Moreover, the effect size of the difference in correlations between best, mean and worst performance was hardly discussed. Due to the low power of the Fisher's *Z*-test and rather small WPR effects (Mean $q = .14$ [17]), these tests will most likely be non-significant, no matter how consistent the increases of correlations across RT bands may be. Nevertheless, there have been efforts to overcome some shortcomings of the *Z*-Test [24,25] and a recent study by Rammsayer and Troche [10] presented significant results from *Z*-Tests comparing correlations between RT in the fastest and slowest RT band with general intelligence. Nevertheless, this approach does not directly account for the shape of increases as well and assumes a linear increase in correlations across performance bands.

Concluding, there have been some attempts to test the WPR for statistical significance, nevertheless the WPR was rather described than tested. For more elaborate insights into the WPR a quantification and statistical test for the WPR is much-needed.

WPR as Moderation

The core of the WPR is moderation. Specifically, the relationship between performance in a cognitive processing task and general intelligence depends on the consecutive number of the performance band or percentile. This means that the size of the relationship between *g* and performance in a cognitive processing task is moderated by the number of the performance band in which the relationship is quantified. This essentially represents an interaction or moderation effect [26,27].

The conceptual approach of WPR as moderation offers an interesting way to test the WPR. Unlike testing a number of correlations against each other, the WPR effect may be modeled as increase in correlations from best to worst performance bands. This increase can be represented as a regression predicting the size of the relationship between *g* and performance by number of performance band. This regression can easily be tested for significance with its coefficients depicting the magnitude of the WPR in the slope of the regression. Finally, this test may be more powerful than the test for differences between correlations and thus even small effects may be detected, even in smaller samples [12].

Furthermore, by comparing results from increases in unstandardized and standardized estimates for the relationship between performance across performance bands and g some of the methodological explanations of the WPR (e.g., variance compression in best performance) can be explored in more detail. Specifically, increases in standardized estimates (e.g., correlations or standardized regression weights) control for increases in inter-individual variance from best to worst performance. In contrast, increases in unstandardized estimates (e.g., covariance or unstandardized regression weights) do not control for these increases in inter-individual variance across performance bands. If differences in variance across performance bands do not affect the WPR, as suggested by Coyle [3], then there should be no difference between WPA on the level of standardized versus unstandardized estimates.

1.3. Possibilities to Analyze the WPR as a Moderation

The regression of covariance or correlation between g and performance (PF) in a cognitive processing tasks on the number of performance band or percentile can be implemented in two ways. First, the relationship between g and performance in a cognitive processing task across performance bands can be estimated and then the estimated relationships across performance bands are predicted by the consecutive number of performance band in a second step. Alternatively, this relationship and its increases across performance bands can be estimated within one step. The first approach represents a sequential regression procedure, whereas the second approach requires Multi-Level Modeling for estimating both the covariances and their increases across performance bands in one step.

Whether this analysis is carried out at the level of standardized or unstandardized parameters, is regulated by the way the dependent variable is entered into the analysis. Entering RT or cognitive performance from best to worst performance bands as absolute values will yield an analysis on the level of unstandardized estimates. In contrast, when RT or performance is z-standardized within each performance band, the analysis is performed on the level of standardized estimates. For convenience and ease of interpretation, we recommend that the measure for g is z-standardized in both the analysis with unstandardized and standardized estimates prior to the analysis.

Both, the sequential regression and the multi-level modeling (MLM) approach will be outlined for unstandardized and standardized estimates in detail in the following sections.[1] We start with presenting the two approaches for unstandardized estimates and then present the two approaches for standardized estimates.

1.3.1. New WPA Approaches with Unstandardized Estimates

Sequential Regression. The sequential regression approach is basically an extension of the traditional worst performance analysis. In a first step mean or median performance PF of each participant i within each performance band B is predicted by general intelligence g:

$$\mathrm{PF}_B = b_B \cdot g + b_{0B} + \epsilon_i \tag{1}$$

This yields different unstandardized regression weights b_B for each performance band B, representing the relationship of PF with g within each performance band. The intercept of these

[1] Please note that both approaches can be implemented with all common approaches that separate the intra-individual performance distribution (i.e., performance bands, percentiles, or, quantiles). The only prerequisite is that the number of performance band or percentile is coded so that the variable contains ascending integer values from fast to worst performance percentile, bands or quantiles. Furthermore this variable is ideally centered to a meaningful value in order to gain interpretable results [28].

regressions b_{0B} in contrast represents the performance of a person with $g = 0$, therefore g should be centered prior to this step [28].

In a second step the unstandardized regression weights b_B across performance bands B are predicted by the number of performance band B (i.e., the consecutive number of performance bands: 1 for the first and best performance band, 2 for the next best, and so on). This represents the moderation of the relationship between g and performance by performance bands. To approximate the increases of unstandardized regression weights across performance bands adequately it may be reasonable to implement non-linear parameters within this regression. In correspondence to the shape of increases of mean RT across performance bands (see Equation (4) in the section of the Multi-level approach), we implemented a polynomial function of third order. Not only does this function approximate the increases in unstandardized regression weights reasonably well ($R^2 \geq .99$, see Figure 1, p. 12), but it also implements the moderation of the RT–g relationship by all performance band variables that describe the shape of increases in mean RT across performance bands (Equation (4)). This solution suited the present data very well. Beyond that, this may still be a good description for the WPR in performance measures that show normal distribution at the intra-individual level in general due to their characteristic shape of increases of mean performance across performance bands. Therefore, the second regression was specified as follows:

$$b_B = b_L \cdot B + b_Q \cdot B^2 + b_C \cdot B^3 + b_{00} + \epsilon_b \tag{2}$$

For this second regression, there are two parameters that quantify the significance and magnitude of the WPR. The variance explained by the regression (i.e., R^2) represents in how far the unstandardized regression weights increase across performance bands and thus the significance of the WPR. As a high R^2 only indicates how *consistent* the increases in unstandardized regression weights across RT bands are, an additional measure is needed to quantify the *magnitude* of the WPR.

The shape and magnitude of the increases across performance bands are determined by the size of the slope parameters within the regression (b_L, b_Q b_C)[2]. The intercept (b_{00}) of the regression represents the regression weight within the centred performance band (i.e., $B = 0$). As the interpretation of the three slope parameters within this regression is rather complex, we propose a difference between b_B in the worst performance (WP) percentile and in the best performance (BP) percentile in reference to mean performance (M_{PF}) as a measure for the effect size (ES) of the WPR:

$$\mathrm{ES}_{\mathrm{WPR}} = \frac{b_{B(\mathrm{WP})} - b_{B(\mathrm{BP})}}{M_{PF}} \tag{3}$$

This effect size basically corresponds to the effect size Cohen's q comparing the correlation between best performance and g to the correlation between worst performance and g. It quantifies the magnitude of increase in unstandardized regression weights as percentage of mean performance in the respective task (for an example see p. 12). Still, it is a simplification with respect to the full shape of the increases in unstandardized regression weights across performance bands. As it does only reflect the absolute difference between best and worst performance bands, it does not represent the non-linear shape of increases between best and worst performance bands. Thus, increases in unstandardized regression weights should always be plotted by performance bands, so that the shape of increases can be evaluated as well. Nevertheless, the proposed effect size may be a good heuristic to evaluate in how far increases in unstandardized regression weights are larger in one task or condition compared to another.

Altogether, the sequential regression approach quantifies the WPR by predicting the magnitude of the relationship between g and performance across performance bands by band number. This provides

---

[2] The indices of these regression weights refer to the linear (L), quadratic (Q), and cubic (C) trend across performance bands.

a set of regression parameters that can be tested for statistical significance on the one hand, and an estimate for the consistency of increases across percentiles on the other hand.

Mutli-level moderation. The multi-level approach is essentially equal to the sequential regression approach, except it estimates all parameters of the sequential regression approach within one step. Accounting for the data structure of performance bands nested in participants, the multi-level approach combines Equations (1) and (2) by entering Equation (2) into Equation (1).

In addition, the intercept varies across performance bands for obvious reasons. The mean performance within each performance band evidently decreases with ascending performance bands. Specifically performance in the first band will be best or fastest, whereas performance in the last band will be worst or slowest. Therefore the intercept b_0 from Equation (1) should be able to change across performance bands as well.

Similar to the increase of unstandardized regression weights across performance bands, the intercept does not increase linearly across performance bands either. In fact, a linear increase of intercepts across performance bands would correspond to an equal distribution of performance at the intra-individual level. Usually we would assume intra-individual performance to be normally distributed, or in case of reaction times right-skewed (e.g., ex-Gaussian or Wald distributed). The intercepts within these distributions usually show non-linear increases across performance bands. These increases are again quite well approximated by a polynomial function of third order. For this, b_0 from Equation (1) will be predicted by percentile:

$$b_{0B} = b_B \cdot B + b_{B^2} \cdot B^2 + b_{B^3} \cdot B^3 + b_{B0} + \epsilon_b \tag{4}$$

Entering this equation into Equation (1) together with Equation (2) yields a prediction of the performance PF in each performance band B for participants i:

$$\begin{aligned} \mathrm{PF}_{iB} = {} & (b_L \cdot B + b_Q \cdot B^2 + b_C \cdot B^3 + b_{00}) \cdot g_i \\ & + (b_B \cdot B + b_{B^2} \cdot B^2 + b_{B^3} \cdot B^3 + b_{B0}) + \epsilon_{iB} \end{aligned} \tag{5}$$

The term in the first line represents Equation (2) and the term in the second line represents Equation (4). Note that now performance within performance bands PF_{iB} is the dependent variable and that all regression parameters are summarized within one Equation.

As multi-level modeling (MLM) allows to separate effects on level 1 (within a person) and level 2 (between people), we may additionally implement random effects for predictors on level 1. This means that the level 1 parameters (i.e., regression weights of performance bands) may vary across level 2 units (i.e., participants). Specifically, this reflects inter-individual differences in the increases of mean RT across performance bands that basically correspond to inter-individual differences in the intra-individual distribution of performance or RTs. This seemed reasonable to us and therefore the whole Equation (4) was estimated with random effects[3]. This results in a MLM equation with correct notation of:

$$\begin{aligned} \mathrm{PF}_{iB} = {} & (\gamma_0 + u_{i0}) + (\gamma_1 + u_{i1}) \cdot B + (\gamma_2 + u_{i2}) \cdot B^2 + (\gamma_3 + u_{i3}) \cdot B^3 + \\ & \gamma_4 \cdot g_i + \gamma_5 \cdot (g_i \times B) + \gamma_6 \cdot (g_i \times B^2) + \gamma_7 \cdot (g_i \times B^3) + \epsilon_{iB} \end{aligned} \tag{6}$$

The performance PF in each performance band B of each participant i is composed of a random intercept $(\gamma_0 + u_{i0})$ and random effects of performance band $(\gamma_{1-3} + u_{i1-3})$. This first line of Equation (6) essentially is Equation (4) . Additionally, PF_{iB} is predicted by a fixed effect of g (γ_4), representing the relationship of PF_{iB} and g for $B = 0$ and cross level interactions between g and

[3] Within MLM, random effects are estimated with a fixed effect γ equal for all level 2 units and a variance u_i across level 2 units (with $u_i = N(0, \sigma_i^2)$).

B (γ_{5-7}), representing the increases of the relationship between PF_{iB} and g across performance bands. This second line of Equation (6) basically represents Equation (2), with the difference that the interactions between performance band B and g are explicitly stated in Equation (6).

This Equation of the MLM allows to estimate the interaction between performance band and g on the relation between g and performance across performance bands. However, because the dependent variable within this approach is the performance within each performance band B, the overall explained variance R^2 of this regression does not refer to the same explained variance as in the second step of the sequential regression approach. In contrast, this approach treats the unstandardized regression weights between g and performance across performance bands as estimates, whereas the sequential regression approach enters these coefficients as manifest variables. Consequently, the sequential regression approach will underestimate the standard errors of coefficients in the second step and thus overestimates their statistical significance [22]. In this sense, the MLM approach results in more accurate estimates of the standard errors from a statistical perspective, because it does not underestimate the standard errors of the respective coefficients. Hence, the MLM approach judges the significance of coefficients more accurately than the sequential regression approach.

The interpretation of the results of the MLM approach is arguably more complex. There is no direct measure for the effect size of the WPR, because unlike the R^2 in the second step of the sequential regression approach, the R^2 of the MLM approach does not refer to the consistency of the increases of unstandardized regression weighs across performance bands. Instead it refers to the variance explained in the performance (PF_{iB}) across performance bands. Still, the effect size introduced in Equation (3) can be computed in the MLM approach as well. For this, the unstandardized regression weight b_B predicting PF by g across performance bands B can be estimated with γ_4 to γ_7:

$$\hat{b}_B = \gamma_4 + \gamma_5 \cdot P + \gamma_6 \cdot P^2 + \gamma_7 \cdot P^3 \tag{7}$$

To calculate the effect size as stated in Equation (3) the regression weights for the best and worst performance bands can be estimated. The mean performance can be estimated with the fixed slope (γ_0), when the performance band variable was centered. With these variables, the proposed effect size can then be calculated.

1.3.2. New WPA Approaches for Standardized Regression Weights

To implement these two WPA approaches on the level of standardized regression weights, the performance within each performance band has to be z-standardized on the inter-individual standard deviation (SD) of the respective performance band. Although we thereby lose information on the absolute increases in performance across performance bands (e.g., increasing RTs from best to worst performance bands) the covariance structure between performance across performance bands and g remains the same only that it is now controlled for increasing variances from best to worst performance. Furthermore, it is necessary that g is z-standardized for the analyses on the level of standardized estimates. However, we recommend to do that for both the analyses on the level of unstandardized and standardized estimates.

Sequential Regression. For the sequential regression approach with standardized estimates only the first step differs considerably from analysing unstandardized estimates. Specifically, we no longer predict the absolute performance PF of each participant i within each performance band B, but the z-standardized performance $z(\mathrm{PF})$ within each performance band B by general intelligence g:

$$z(\mathrm{PF}_B) = \beta_B \cdot g + \epsilon_i \tag{8}$$

This results in standardized regression weights β_B for each performance band quantifying the standardized relationship (i.e., correlation) between performance in each performance band with g. Please note that there is no longer any intercept for this regression, because the intercept is always zero when using z-standardized measures. The standardized regression weights across performance bands

β_B can again be predicted by the number of performance band in a second step that implements the moderation of the relationship between performance PF and g by performance band:

$$\beta_B = b_L \cdot B + b_{00} + \epsilon_b \tag{9}$$

According to the common assumption that correlations increase linearly from best to worst performance [3], we implemented only a linear increases in standardized regression weights across performance bands.[4] Nevertheless, it is possible to implement non-linear increases in this approach as well. For this, additional regression weights specifying quadratic or cubic trends can be entered into Equation (9), just like in Equation (2).

Comparable to the sequential regression approach on the level of standardized regression weights, there are two parameters that quantify the significance of the WPR. On the one hand, the R^2 of this regression quantifies the consistency of increases in performance bands. On the other hand, the regression weight b_L quantifies the size of increases across performance bands. The intercept b_{00} quantifies the standardized relation for the centered performance band.

To quantify the magnitude of the WPR on the level of standardized estimates it is best to compute the effect size Cohen's q from Equation (9). To do so, we calculate the estimated standardized regression weight for the best performance band β_{BP} and the estimated standardized regression weight for the worst performance band β_{WP}. These can then be transformed into Z-values with a Fisher Z-transformation and the difference between Z_{WP} and Z_{BP} yields Cohen's q [12].

Multi-level moderation. Again, the Multi-level approach is essentially equal to the sequential regression approach apart from the fact that it estimates both steps of the sequential regression approach in one step. For this, Equation (9) is entered into Equation (8), resulting in:

$$\begin{aligned} z(\text{PF}_{iB}) &= (b_L \cdot B + b_{00}) \cdot g + \epsilon_{iB} \\ &= b_{00} \cdot g + b_L \cdot (B \times g) + \epsilon_{iB} \end{aligned} \tag{10}$$

In contrast to the MLM approach on the level of unstandardized regression weights, it is not necessary to estimate a fixed effect of the increases in performance across performance bands (see Equation (4)), because the z-standardization of performance in each performance band resulted in a mean performance of zero within each performance band. However, a random effect for this effect can still be estimated. This effect reflects that there may not be full differential stability in performance across performance bands. For example, one person can show above average performance in best performance bands and only average performance in worst performance bands, whereas for another person the position in comparison to other participants stays the same across performance bands. This results in a full MLM equation with correct notation of:

$$z(\text{PF}_{iB}) = \gamma_1 \cdot g + \gamma_2 \cdot (B \times g) + (\gamma_3 + u_{i3}) \cdot B + \epsilon_{iB} \tag{11}$$

with the fixed effect $\gamma_3 = 0$ this results in:

$$z(\text{PF}_{iB}) = \gamma_1 \cdot g + \gamma_2 \cdot (B \times g) + (u_{i3}) \cdot B + \epsilon_{iB}$$

Within this approach γ_1 represents the relationship between performance and g in the centred performance band and γ_2 represents the linear increase in this relationship across performance bands. The random effect u_{i3} represents the variance in the relative position across performance bands for

[4] This is also reflected in often non-linear increase of variance across performance bands, especially for RT. In the process of standardization the MLM equation for unstandardized regression weights basically gets divided by this non-linear increase and this leaves only the linear part of the increases in regression weights as a good approximation for the WPR across RT bands.

participants. In detail, this variance would be zero, if performance across performance bands is perfectly correlated.

As in the sequential regression approach, the increase in standardized regression weights can be computed with γ_1 and γ_2. Thus we can estimate the relationship between performance and g in the best and worst performance band and estimate the effect size Cohen's q as difference between these two estimates on the level of Z-scores. Specifically, the estimated standardized regression weight within each performance band B can be estimated via:

$$\hat{\beta}_B = \gamma_2 \cdot B + \gamma_1 \tag{12}$$

Once more, the MLM approach treats the standardized regression weights β_B across performance bands as estimated, whereas the sequential regression approach treats them as manifest. Thus the MLM approach is generally, for unstandardized and standardized estimates, the statistically more sound approach because it will lead to less attenuated standard errors of increases in regression weights and thus does not inflate α-error probability of these increases.

1.4. Aims of the Empirical Example

The application of these newly introduced methods for WPA with empirical data has two main objectives. First, this will allow to compare the newly introduced methods for WPA to the traditional approach for WPA. Second, this comparison will allow to determine advantages and problems of the newly introduced methods, and may thereby convey which method and which level of analysis (unstandardized versus standardized) is adequate for a powerful analysis of the WPR.

Providing a powerful test and a quantification of the WPR would help researchers to determine processes that underlie the WPR and thus gain deeper knowledge on processes basic for g. Specifically, the newly introduced methods will not be capable of distinguishing between different theoretical explanations of the WPR. Nevertheless, they may present a more accurate analysis and test for the WPR and thereby provide researchers with a method that gives more robust results in studies that aim at testing different theoretical explanations of the WPR. This may help in finding processes underlying the WPR and result in a better understanding of processes fundamental to g.

To facilitate the use of the new approaches for WPA we provide commented R code for both approaches in the supplementary material. Additionally, the data of the empirical example are given in the supplementaries, so that the results can be reproduced and both approaches for unstandardized and standardized estimates can be studied in more detail.

2. Experimental Section

2.1. Participants

Data for this example were taken from a study over three measurement occasions with a cognitive abilities and personality assessment on the second measurement occasion. For this study, 134 participants from the area around Heidelberg, Germany were recruited. Participants' age ranged from 18 to 61 years (M_{age} = 37.12, SD_{age} = 13.75), 60.4% were female, and they had different educational and occupational backgrounds.

For the present analysis we used data from the first and second measurement occasion. Some participants dropped out and one participant was excluded due to extreme scores (for a detailed description of the outlier analysis see the statistical analyses section on page 10). This resulted in a sample of 121 participants (58.7% female) aged from 18 to 61 years (M_{age} = 36.64, SD_{age} = 15.65) that were included in this analysis.

2.2. Measures

Sternberg Memory Span Task. The cognitive processing task analyzed in the present study was a computerized version of the Sternberg Memory Span Task [29] also used by Schubert et al. [2]. In this task participants were shown a memory set consisting of one to five numbers from 0 to 9 on a black computer screen. Subsequently, participants were shown a probe number and had to decide whether the probe was or was not contained in the afore presented number set by pressing one of two keys. The position of keys indicating whether the probe item was part of the memory set or not was counterbalanced across participants.

Three experimental conditions with different memory set sizes (1, 3, and 5 numbers) were administered. All three blocks started with ten practice trials with feedback, followed by 100 test trials without feedback. The order of the three memory set size conditions was counterbalanced across participants. Between blocks participants were offered a short break.

Each trial started with a fixation cross presented for 1000 to 1500 ms. Then, numbers were presented sequentially for 1000 ms. Between numbers a blank screen was presented for 400 to 600 ms. After the last number of the memory set was presented, a black screen with a question mark was shown for 1800 to 2200 ms, followed by a probe item showing a single digit. Participants then had to indicate whether the number was part of the memory set or not by pressing the corresponding key on a standard computer keyboard.[5] After the response the probe item remained on screen for 1000 ms, followed by an inter-trial interval of 1000 to 1500 ms. The stimuli were presented on a 17 inch LED computer screen and the experiment was programmed in E-Prime 2.0 Professional.

Berlin Intelligence Structure Test (BIS). Within the cognitive abilities and personality assessment, participants completed the Berlin Intelligence Structure Test (BIS [30]). The assessment was carried out according to the standardized instructions. The assessment ran in groups of up to four persons and started with the BIS assessment, followed by a personality questionnaire (NEO-FFI), the Raven Advanced Progressive Matrices, and a demographic questionnaire. For this study only the BIS results were analyzed.[6]

BIS results were evaluated in correspondence with the evaluation instructions from the manual. First, raw scores were determined for all tasks and subsequently the raw scores were transformed into standardized scores. From these scores one score for general intelligence (g) was calculated.

2.3. Statistical Analyses

Outlier Analysis. Before running all analyses, we carefully examined the data for uni- and multivariate outliers in a three-step procedure. First we discarded all RTs with incorrect responses. For all correct response RTs we checked for intra-individual outliers in reaction times: initially, reaction times lower than 100 ms and higher than 3000 ms were excluded for all participants. Then, we computed mean and standard deviation for the logarithmized reaction time within each participant and each experimental condition, and excluded reaction times below and above three standard deviations from the mean of logarithmized reaction times.

Secondly, participants with univariate outliers in reaction time and intelligence test scores were excluded from the data analysis when mean reaction time or intelligence test scores showed an absolute difference larger than three standard deviations from the sample mean. Finally, multivariate outliers

---

5 Although often specialized response boxes are used for the registration of responses in such tasks because latencies are a lot smaller on these specialized devices (1 to 3 ms) compared to a standard keyboard (12 to 36 ms), we used a standard keyboard for economic reasons. However, the same keyboard and computer set-up was used for all participants and thus it is unlikely that the use of a standard keyboard systematically distorted the RT data.

6 In many former studies investigating the WPR the Raven Advanced Progressive Matrices (RAPM) were used as measure for g. While the RAPM may be the best single measurement to approximate g, estimating g with a more heteogeneous set of tasks (in our case the BIS) gains a better estimate for g [31]. Furthermore, results with the RAPM as measure for g were similar to the results reported in the manuscript.

on the combination of mean RT in each condition and intelligence test score were excluded when they had a Mahalanobis distance larger than 13.816, corresponding to $\chi^2(2)_{p<.001}$. In an iterative process, this procedure was repeated until no further participants were detected as multivariate outliers. Within this procedure, one person was identified as a uni-variate outlier on general intelligence, processing speed, creativity, and verbal abilities ($z_s < -3$), as well as a multivariate outlier on the combination of IQ and RT in experimental conditions with memory set size 1 and 5 (Mahalanobis Distance = 14.08–16.71).

Statistical Analysis. Both the sequential regression approach and the MLM approach were calculated with R [32] and conducted separately for the three experimental conditions of the Sternberg memory span task. For the sequential regression analysis, regressions were estimated in a stepwise procedure. First, performance within each RT band was predicted by general intelligence (see Equations (1) and (8)), and second, the unstandardized and standardized regression weights from step one were predicted by RT band (see Equations (2) and (9)).

The analysis of the MLM approach of the Worst Performance Rule was conducted using the *nlme* package in R [33]. In a stepwise procedure all parameters from Equations (6) and (11) were added to a Random Intercept Model that served as baseline model. As following models were nested, models with additional parameters were required to show significant increase in Log likelihood to be considered a better data description. In addition, decreases in Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used as indicators for model fit.

Fixed effects were tested for significant deviation from zero using a one-sided t-test. Further, random slopes were tested for significance with a Likelihood-Ratio test. Additionally, random effects were estimated with an unstructured G-Matrix, not only estimating the variances of each random effect but additionally estimating the covariances between all random effects.

For both approaches, nine RT bands that contained 9 to 11 RTs were constructed, so that each RT band contained approximately 11.1% of RTs of the intra-individual RT distribution. Although many other studies used percentiles or RT bands with five RTs in each band, we decided to construct nine RT bands in order to be able to center the RT band variable to a meaningful value (i.e., the fifth RT band). If two reaction times at the border of an RT band were equal, they were assigned to different RT bands. RT band number (i.e., performance band variable B) was centered to the fifth RT band of the intra-individual RT distribution, in order to obtain meaningfully interpretable estimates for the fixed intercepts and fixed slopes [28]. Specifically, the intercept represents approximately the median RT of an average intelligent person and the fixed effect of g represents the unstandardized regression weight from g on RT in the fifth RT band. After centering B to the fifth RT band, B^2 and B^3 were derived from the centred B-variable. The general intelligence score (g) from the BIS was z-standardized within the sample.

3. Results and Discussion

3.1. Descriptives

Descriptive statistics for the Sternberg memory span task and BIS results are given in Table 1. In line with earlier results, RT increased with larger memory set size, $F(2, 240) = 277.5$, $p < .05$, $\epsilon = .69$, $\omega^2 = .70$. The BIS scores were representative compared to the standardization sample, $t(120) = -.71$, $p > .05$, Cohen's $d = -.06$.[7] There were no significant differences in general intelligence between women and men, $t(119) = .95$, $p > .05$, Cohen's $d = .18$.

[7] Please note that the standardization sample of the BIS consisted of adolescents and young adults with higher education. Thus, the present sample may be somewhat above average in cognitive abilities compared to an average intelligent population.

Table 1. Descriptives for the Sternberg Task and the BIS.

Measure	Mean	Median	SD	Min	Max	Rel.
S1 (RT)	587.9	564.8	111.0	396.2	969.5	.96 [a]
S3 (RT)	726.7	694.0	166.8	479.0	1372.0	.98 [a]
S5 (RT)	889.6	828.0	241.7	539.4	1638.0	.97 [a]
BIS [c]	99.6	99.7	5.7	86.3	114.8	.94 [b]

Note: Rel. = reliability; [a] Estimated via Odd-Even correlations—for this trials were separated into odd and even trials by trial-number; [b] Estimated via Cronbach's α; [c] Standardized scores of the BIS are set to have a mean of 100 and a standard deviation of 10.

3.2. Results from New WPA Approaches with Unstandardized Regression Weights

3.2.1. Results of the Sequential Regression Approach

In the first step of the sequential regression approach, general intelligence predicted reaction time within RT bands across all three conditions, $F\text{s}(1, 119) \geq 6.7$, $p\text{s} < .05$ and $R^2\text{s} = .05$–$.22$. Specifically, these results convey that inter-individual differences in mean reaction time in each RT band across all three conditions were predicted by general intelligence. In the second step, the number of RT band predicted the unstandardized regression weights across RT bands in all three conditions, $F\text{s}(3, 5) \geq 127.2$, $p\text{s} < .05$ and $R^2\text{s} \geq .98$ (for an illustration of the regressions estimated in the second step, see Figure 1).

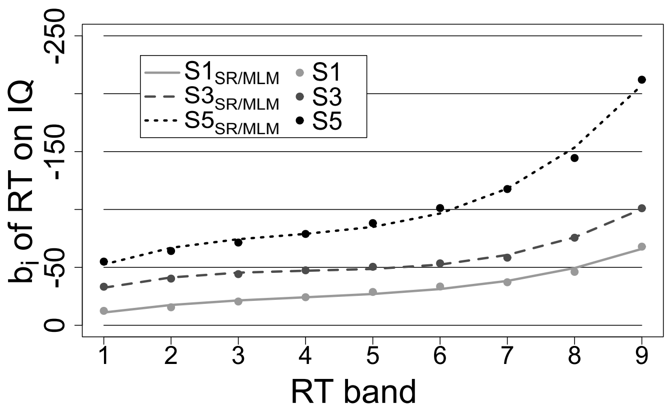

Figure 1. Depicts the prediction of the unstandardized regression weights (dots) across RT bands by the sequential regression and the MLM approach (lines). Note that point estimates for the SR and MLM approach were equal. Thus there were no separate lines for the two approaches.

All estimated parameters differed significantly from zero (see Table 2). As indicated by the effect size proposed in Equation (3), increases in the unstandardized regression weight b_B did not differ between set size 1 and 3, but tended to be larger for set size 5 ($\text{ES}_{S1} = .09$, $\text{ES}_{S3} = .09$ and $\text{ES}_{S5} = .17$). Specifically, this means that the increases in the unstandardized regression weight from best to worst performance correspond to 9% to 17% of the mean reaction time in the corresponding condition. For example, in the S1 condition the mean performance was 587.9 ms. With an ES = .09, the difference of the unstandardized regression weight between best and worst performance thus was about 52.9 ms. This means that the difference in RT between an individual one SD above average in IQ and an individual average in IQ increased for 52.9 ms from best to worst performance RT band.

Table 2. Estimated parameters for the sequential regression approach with unstandardized regression weights.

Parameters		Condition					
		S1		S3		S5	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
b_{00}	(Int.)	−27.00 *	1.25	−48.62 *	0.85	−84.93 *	2.86
b_{WPR1}	(B)	−3.32 *	0.84	−2.26 *	0.57	−8.11 *	1.92
b_{WPR2}	(B^2)	−0.72 *	0.14	−1.11 *	0.10	−2.81 *	0.32
b_{WPR3}	(B^3)	−0.22 *	0.07	−0.39 *	0.04	−0.70 *	0.15
R^2		.98		1.00		.99	

Note: S.E. = standard error of the respective estimate; * Indicates p < .05; The terms in brackets indicted the predictor corresponding to the respective parameter.

3.2.2. Results of the MLM Approach

For all three conditions the Log-likelihood (LL) ratio tests indicated best fit for the full MLM with all parameters included (see Table 3). Although successive LL-ratio tests did not always indicate significant improvement in fit, for S1, the comparison of model 6 to 10 showed improved model fit, $\chi^2(4) = 11.1$, $p < .05$, for S3, the comparison of model 7 to 10 indicated improved model fit, $\chi^2(3) = 8.1$, $p < .05$, and for S5, the comparison of model 7 to 10 indicated better model fit, $\chi^2(3) = 28.5$, $p < .05$ (see Table 3).

Table 3. Estimates for model fit of the MLM across all three conditions for the WPA with unstandardized regression weights.

Cond	Model		DF	AIC	BIC	LogLik	L. Ratio	p
S1	Base		3	14,352.4	14,367.3	−7173.2	–	–
	1	$(+\gamma_1)$	4	13,099.3	13,119.2	−6545.6	1255.1	.000
	2	$(+\gamma_2)$	5	12,891.7	12,916.7	−6440.8	209.6	.000
	3	$(+\gamma_3)$	6	12,775.8	12,805.8	−6381.9	117.9	.000
	4	$(+u_{i1})$	8	11,657.3	11,697.2	−5820.6	1122.6	.000
	5	$(+u_{i2})$	11	11,247.5	11,302.4	−5612.7	415.8	.000
	6	$(+u_{i3})$	15	10,940.2	11,015.1	−5455.1	315.3	.000
	7	$(+\gamma_4)$	16	10,939.3	11,019.2	−5453.6	2.9	.090
	8	$(+\gamma_5)$	17	10,937.7	11,022.6	−5451.9	3.6	.059
	9	$(+\gamma_6)$	18	10,938.9	11,028.8	−5451.5	0.8	.369
	10	$(+\gamma_7)$	19	10,937.0	11031.9	−5449.5	3.9	.049
S3	Base		3	14,825.1	14,840.1	−7409.5	–	–
	1	$(+\gamma_1)$	4	13,630.7	13,650.6	−6811.3	1196.4	.000
	2	$(+\gamma_2)$	5	13,465.1	13,490.1	−6727.6	167.6	.000
	3	$(+\gamma_3)$	6	13,343.2	13,373.2	−6665.6	123.9	.000
	4	$(+u_{i1})$	8	12,329.8	12,369.7	−6156.9	1017.4	.000
	5	$(+u_{i2})$	11	11,736.6	11,791.5	−5857.3	599.2	.000
	6	$(+u_{i3})$	15	11,260.5	11,335.4	−5615.2	484.2	.000
	7	$(+\gamma_4)$	16	11,255.3	11,335.2	−5611.6	7.2	.007
	8	$(+\gamma_5)$	17	11,255.9	11,340.8	−5610.9	1.4	.236
	9	$(+\gamma_6)$	18	11,257.8	11,347.7	−5610.9	0.0	.847
	10	$(+\gamma_7)$	19	11,253.2	11,348.1	−5607.6	6.7	.010
S5	Base		3	15,501.8	15,516.8	−7747.9	–	–
	1	$(+\gamma_1)$	4	14,245.4	14,265.4	−7118.7	1258.4	.000
	2	$(+\gamma_2)$	5	14,038.7	14,063.7	−7014.4	208.7	.000
	3	$(+\gamma_3)$	6	13,905.5	13,935.5	−6946.8	135.2	.000
	4	$(+u_{i1})$	8	12,740.3	12,780.2	−6362.1	1169.3	.000
	5	$(+u_{i2})$	11	12,265.7	12,320.6	−6121.9	480.6	.000
	6	$(+u_{i3})$	15	11,974.5	12,049.4	−5972.3	299.2	.000
	7	$(+\gamma_4)$	16	11,971.0	12,050.9	−5969.5	5.5	.019
	8	$(+\gamma_5)$	17	11,970.6	12,055.4	−5968.3	2.4	.119
	9	$(+\gamma_6)$	18	11,965.3	12,055.1	−5964.6	7.3	.007
	10	$(+\gamma_7)$	19	11,948.5	12,043.4	−5955.3	18.8	.000

Note: Cond = condition, LogLik = Log Likelihood, L. Ratio = Log Likelihood Ratio in comparison to the model one line above, Base = Baseline model, Expressions in parentheses denote variables added to the model.

The AIC decreased across successive models, except within the S3 conditions for models 7 to 9. For models 6 to 10 in the S1 and S3 conditions, these decreases were below the critical difference of 10, which is often used as a cut off criterion for significant differences in model fit [34]. The BIC indicated best fit for the full MLM only in the S5 condition. For S1, the BIC was lowest for model 6, without the prediction of PF_{iB} by g and interactions between RT band and g (i.e., without γ_4 to γ_7). And for S3, BIC was lowest for model 7, without interactions between RT band and g (i.e., without γ_5 to γ_7). However, as the prediction of PF_{iB} by g and interactions of g with RT band were the core of the present analysis and because the LL-ratio tests for model fit indicated better fit for the more complex models, we retained the full MLM for all three conditions.

The parameters estimating the WPR from the retained MLMs were numerically equivalent to those from the sequential regression approach (see Table 4). All other parameters for the MLMs in all three conditions can be reproduced with the syntax and data given in the supplementary material online. As parameters from the MLM were equal to the parameters from the sequential regression approach, the estimated effect sizes were equal for the MLM likewise ($ES_{S1} = .09$, $ES_{S3} = .09$ and $ES_{S5} = .17$).

Table 4. Parameters estimating the WPR within the MLM approach with unstandardized regression weights.

Parameters		Condition					
		S1		S3		S5	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
γ_4	(g)	−27.00 *	8.79	−48.62 *	12.02	−84.93 *	18.4
γ_5	($g \times P$)	−3.32 *	1.42	−2.26	1.30	−8.11 *	2.53
γ_6	($g \times P^2$)	−0.72 *	0.34	−1.11 *	0.51	−2.81 *	0.53
γ_7	($g \times P^3$)	−0.22 *	0.11	−0.39 *	0.15	−0.70 *	0.16

Note: S.E. = standard error of the respective parameter; * Indicates $p < .05$; The expression in brackets indicates the predictor corresponding to each parameter.

The standard errors for the coefficients were considerably larger in the MLM than in the sequential regression approach. As mentioned earlier, this is because the MLM approach treats the covariances between PF_{iB} and g across RT bands as estimated, whereas the sequential regression approach treats them as observed. Thus, the sequential regression approach underestimates the standard errors of the coefficient and the MLM approach estimates the standard errors more accurately (see p. 7). This is why one parameter (γ_5) in the S3 condition did not differ from zero to a statistically meaningful extent in the MLM approach, although parameters showed significant differences from zero in the sequential regression approach.

Altogether, these results showed that both approaches, the sequential regression approach and the MLM approach, estimate the WPR via an interaction between RT band and g. In addition to RT band and g itself, this interaction predicts the performance in each RT band. Beyond that, results showed that increases in unstandardized regression weights between performance in each RT band and g are almost perfectly predicted by RT band ($R^2 \geq .99$).

3.3. Results from New WPA Approaches with Standardized Regression Weights

3.3.1. Sequential Regression Approach

On the level of standardized regression weights, in step one, z-standardized RT was predicted by g across all performance bands in all three experimental conditions, $F_s(1, 119) \geq 6.7$, $p_s < .05$, and $R^2 = .05$–$.22$. In step two, results showed that standardized regression weights β_B increased in absolute size in the S1 and S5 condition, whereas standardized regression weights decreases in size in the S3 condition (see Table 5 for the estimated parameters). Additionally, the results from step two showed that standardized regression weights were overall smallest in the S1 condition and increased with larger memory set size, which is reflected in increasing size of the intercepts from S1 to S5 (see Table 5).

Table 5. Estimated parameters from WPA with z-standardized RTs as DVs.

Condition	Parameter		Seq. Regression		MLM	
	Seq. Reg.	MLM	Estimate	S.E.	Estimate	S.E.
S1	b_{00}	γ_1	−.266 *	.004	−.266 *	.082
	b_L	γ_2	−.006 *	.001	−.006	.010
S3	b_{00}	γ_1	−.336 *	.005	−.336 *	.083
	b_L	γ_2	.008 *	.001	.008	.007
S5	b_{00}	γ_1	−.409 *	.005	−.409 *	.081
	b_L	γ_2	−.008 *	.002	−.008	.007

Note: Seq. Regression and Seq. Reg. = sequential regression approach; S.E. = standard error of the respective parameter; * Indicates $p < .05$.

With respect to the consistency and magnitude of increases across RT bands, the results from the sequential regression with standardized estimates suggested that increases of standardized regression weights across RT bands were less consistent, $R^2 = .66$–$.67$, than increases of unstandardized regression weights. Furthermore, the magnitude of increases was slightly smaller in the S1 condition ($q = .05$) than in the S5 condition ($q = .08$). In the S3 condition standardized regression weights actually decreased in size from best to worst performance bands ($q = -.07$). Thus, the S3 condition contradicted the WPR on the level of standardized regression weights.

3.3.2. MLM Approach

For all three conditions, the Log-likelihood-ratio tests indicated best model fit for a MLM model without the interaction of $g \times B$ (i.e., model 2), indicating that there are no increases in the relationship between g and RT across RT bands (see Table 6 for model fit). AIC and BIC likewise indicated best model fit for a MLM model without the interaction of $g \times B$, although ΔAIC is below the critical value of 10 in the S1 condition and ΔBIC is below the critical value of 10 for the S1 and S3 condition. Altogether these results indicate that there are significant relationships between g and RT, however this relationship does not vary across RT bands, contrary to the predictions made by the WPR.

Table 6. Estimates for model fit of the MLM with standardized regression weights across all three conditions.

Cond	Model	DF	AIC	BIC	LogLik	L. Ratio	p
S1	Base	3	1313.0	1328.0	−653.5	–	–
	1 $(+u_{i3})$	5	384.1	409.0	−187.0	933.0	.000
	2 $(+\gamma_1)$	6	376.0	406.0	−182.0	10.0	.000
	3 $(+\gamma_2)$	7	377.7	412.6	−181.8	0.4	.551
S3	Base	3	588.2	603.2	−291.1	–	–
	1 $(+u_{i3})$	5	−399.1	−374.2	204.6	991.3	.000
	2 $(+\gamma_1)$	6	−412.5	−382.6	212.3	15.4	.000
	3 $(+\gamma_2)$	7	−411.7	−376.7	212.8	1.1	.287
S5	Base	3	534.5	549.5	−264.2	–	–
	1 $(+u_{i3})$	5	−297.4	−272.5	153.7	835.9	.000
	2 $(+\gamma_1)$	6	−319.5	−289.6	165.8	24.1	.000
	3 $(+\gamma_2)$	7	−318.8	−283.9	166.4	1.3	.249

Note: Cond = condition, LogLik = Log Likelihood, L. Ratio = Log Likelihood Ratio in comparison to the model one line above, Base = Baseline model, Expressions in parentheses denote the variable added to the model.

However, if we take a look at the estimated parameters of model 3 with the interaction of $g \times B$ (see Table 5), the results show that the numerical estimates for the relationship in $B = 0$ between g and RT (i.e., b_{00} or γ_1) and the increases across RT bands (i.e., b_L or γ_2) are equal for the sequential regression approach and the MLM approach in all three conditions. The difference in the significance of these parameters is again due to larger standard errors in the MLM approach. Just as in the analyses with unstandardized regression weights, this can be explained by the fact that the sequential regression approach treats the standardized regression weights across RT bands as manifest and the MLM approach treats them as estimated. Thus, the MLM approach takes into account that there is uncertainty in the estimation of standardized regression weights across RT bands and estimates standard errors of the parameters accordingly. All in all, the MLM approach is therefore more accurate and the results suggest that there is no WPR on the level of standardized regression weights.

3.4. Results from a Traditional WPA

To compare the results of the two introduced methods with results from a traditional worst performance analysis, we performed the latter as well. For this, we calculated mean RTs for the 9 RT bands within each participant in all experimental conditions. Then, we computed correlations of the mean RTs across RT bands with the BIS score.

For the S1 and S5 condition, the WPR was replicated with slightly more consistently increasing correlations across RT bands in the S5 condition (see Table 7). Furthermore, the effect size Cohen's q for the difference between correlations in best and worst performance RT band was higher in the S5 condition ($q_{S5} = .10$) than in the S1 condition ($q_{S1} = .04$). In the S3 condition, correlations were positively associated with RT band number and decreased in absolute value with ascending RT bands. The S3 condition thus did not replicate the WPR. However, the differences in correlations between best and worst performance bands did not differ significantly from zero for all three conditions, $Z_{S1} = .41$, $p = .68$, $Z_{S3} = -1.08$, $p = .28$, and $Z_{S5} = 1.47$, $p = .14$.

The estimated reliability for RTs within RT bands was high across all RT bands and conditions (see Table 7). Since reliabilities did not increase with ascending RT bands, the increases in correlations between RT and g cannot be attributed to decreases in error variance with ascending RT bands.

Altogether, the results from a traditional WPA differed from the results of the sequential regression and the MLM approach with unstandardized regression weights. Although there were consistent increases in correlations across RT bands in the S1 and S5 condition (Spearman's rank correlation $r = -.82$ to $-.91$) suggesting a WPR, the size of the differences in correlations across RT bands was not significant. Thus the traditional WPA described the WPR, but the actual test for increases in correlations across RT bands was not significant in all three experimental conditions. The results from the MLM approach with standardized regression weights were in line with the results from the z-Test and indicated that there is no WPR. Considering that the sequential regression approach with standardized regression weights estimated equal parameters as in the MLM approach but overestimates the significance of these parameters, the results of the sequential regression approach with standardized regression weights may be interpreted the same way.

In contrast, both the sequential regression approach and the MLM approach on the level of unstandardized regression weights showed significant increases in unstandardized regression weights from best to worst performance bands in all three conditions that support the WPR. On the one hand, the consistency of increases in unstandardized regression weights analyzed in the sequential regression approach and in the MLM approach was substantially larger ($R^2 \geq .99$) than in the analysis of correlations in the traditional WPA and of standardized regression weights (R^2 = .66 to .83). On the other hand, on the level of unstandardized regression weights the increases were all consistent with the WPR, whereas on the level of standardized regression weights and correlations there was a decrease in the size of correlations with ascending RT bands in the S3 condition. All in all, there are considerable differences between results on the level of unstandardized versus standardized regression weights that need to be discussed.

Table 7. Covariance, unstandardized regression weight, standard of RT, correlation, and reliability of RTs across the nine RT bands in all three experimental conditions.

P	S1					S3					S5				
	Cov	$b_{RT,g}$	SD_{RT}	$r_{RT,g}$	Rel.	Cov	$b_{RT,g}$	SD_{RT}	$r_{RT,g}$	Rel.	Cov	$b_{RT,g}$	SD_{RT}	$r_{RT,g}$	Rel.
1	−12.4	−12.4	51.5	−0.24	.98	−33.3	−33.3	96.1	−0.35	.99	−55.0	−55.0	140.5	−0.39	.99
2	−15.5	−15.5	67.3	−0.23	1.00	−40.3	−40.3	111.7	−0.36	1.00	−64.1	−64.1	164.2	−0.39	1.00
3	−20.5	−20.5	80.1	−0.26	1.00	−44.2	−44.2	121.8	−0.36	1.00	−71.4	−71.4	179.3	−0.40	1.00
4	−24.2	−24.2	91.6	−0.26	1.00	−47.5	−47.5	134.4	−0.35	1.00	−78.9	−78.9	201.2	−0.39	1.00
5	−28.8	−28.8	103.0	−0.28	1.00	−50.5	−50.5	146.8	−0.34	1.00	−88.2	−88.2	221.8	−0.40	1.00
6	−33.5	−33.5	116.4	−0.29	1.00	−53.5	−53.5	163.3	−0.33	1.00	−101.2	−101.2	251.0	−0.40	1.00
7	−37.1	−37.1	135.2	−0.27	1.00	−58.4	−58.4	183.2	−0.32	1.00	−117.7	−117.7	283.8	−0.41	1.00
8	−46.1	−46.1	161.4	−0.29	1.00	−75.6	−75.6	228.5	−0.33	1.00	−144.5	−144.5	336.0	−0.43	1.00
9	−67.9	−67.9	245.3	−0.28	.96	−101.1	−101.1	355.7	−0.28	.98	−212.1	−212.1	454.4	−0.47	.98
R		−1.00	1.00	−.82			−1.00	1.00	.90			−1.00	1.00	−.91	

Note: P = number of RT band, Cov = covariance between RT and g, $b_{RT,g}$ = unstandardized regression weight predicting RT by g, SD_{RT} = inter-individual standard deviation of mean RT within each RT band, $r_{RT,g}$ = correlation between RT and g, Rel. = reliability of RT estimated with odd-even correlations—for this Trials were separated into odd and even trials by trial number and reliability of all RT in an performance band was estimated with the Spearman-Brown formula, R = Spearman's rank correlation between RT band and the variable presented in the column.

3.5. Discussion

The present work conceptualized the WPR as a moderated effect of *g* on performance in a cognitive processing task that depends on the performance band or percentile in which performace is measured. Following this idea, we introduced two approaches to analyze the WPR. Both approaches were tested on the level of unstandardized and standardized estimates in an empirical example. unstandardized regression weights quantifying the relation between *g* and RT across RT bands showed perfect monothonic increases from best to worst performance bands. In correspondence to a larger WPR in tasks with higher *g*-loadings [3,17], the increases tended to be larger in more complex conditions. However, comparing the results with unstandardized regression weights to results with standardized regression weights and to results from a traditional WPA showed that increases in unstandardized regression weights do not necessarily correspond to the WPR from a traditional WPA perspective.

3.5.1. Differences in Analyses with Unstandardized and Standardized Estimates

These differences between results with unstandardized regression weights and results with standardized regression weights or results from traditional WPA are due to increases in inter-individual standard deviation of RT across performance bands. Equally to the covariance, the inter-individual standard deviation of RT (SD_{RT}) increased consistently (Spearman's rank correlation: $r = 1.00$) from best to worst performance bands (see Table 7). Furthermore, for a pair of highly correlated variables (e.g., RT_{BP} and RT_{WP}) with different variances, the covariance between these two variables and a third variable *g* increases proportionally to the increase in variance. As correlations between mean RT across performance bands are medium to very high ($r = .55$ to .99 for the present sample), it may be that larger standard deviations of RT in worst performance percentiles lead to higher covariances with *g*, given that the covariance with *g* is a function of the variance for perfectly correlated variables. If this were the case, the increase in unstandardized regression coefficients that basically represents the covariance may be nothing else but a reflection of the increase in inter-individual standard deviation of RT. It seems that this was exactly the case in the present study, because analyses on the level of standardized regression weights that control for increases in variance across performance bands did not show the WPR.

In contrast to Coyle [3] who stated that differences in variance between best and worst performance do not affect the WPR, the present results show that increases in variance from best to worst performance may play a crucial role for the WPR, especially if best and worst performance is highly correlated. Specifically, the WPR on the level of correlations relies on increases in covariance between performance and *g* from best to worst performance that are proportionally larger than and independent from increases in variance in performance from best to worst performance bands. Despite the fact that results of analyses on the level of standardized regression weights may provide the actual WPR, we think that it is noteworthy that variance as well as covariance with *g* in worst performance band is notably larger than in best performance bands. In a nutshell, there are larger inter-individual differences in worst performance RT than in best performance RT that may drive the increase in covariance with *g* from best to worst performance bands. It would be interesting to see in how far this result is present in other basic cognitive processing task as well.

3.5.2. Differences between Newly Introduced WPA and Traditional WPA

Beyond the difference in analyses with unstandardized and standardized regression weights, there are important differences between traditional WPA and the newly introduced methods to analyze the WPR. Specifically, the traditional WPA *described* the increases in correlations accross RT bands, whereas the newly introduced methods provided an acutal *test* and a *quantification* for the magnitude of the WPR. Although the consistency and significance of increases in correlations across RT bands was sometimes tested with Spearman's rank-correlations or a Fisher's *Z*-Test [10], this is not sufficient for quantifying the actual magnitude and shape of increases across RT bands.

Despite the possibility to compute an effect size for the difference in correlations across RT bands (Cohen's q) as a measure fot the magnitude of the WPR, a test for the significance of these differences lacks statistical power [12]. By modelling the increases in unstandardized and standardized regression weights as a moderation, the newly introduced methods provided an actual test for the significance of the WPR and a quantification for the magnitude of the WPR. Furthermore, the newly introduced methods take the full shape of increases in unstandardized or standardized regression weights across all performance bands into account, whereas the Z-test only evaluates in how far correlations from best to worst performance with g differ significantly. And additionally, the test for consistency in increases across performance bands with Spearman's rank correlations is flawed like the sequential regression approach, because it treats the estimated correlations in performance bands as manifest and does not account for the error in estimation. Altogether, the new methods thus overcome some major problems of traditional WPA.

3.5.3. Discussion of Differences between the Sequential Regression and the MLM Approach

With respect to differences between the two newly introduced methods, the results showed that the coefficients estimated to describe the increases in unstandardized and standardized regression weights are equal for both approaches. However, the standard errors of these coefficients were smaller for the sequential regression approach than for the MLM approach. Although smaller standard errors may seem as an advantage of the sequential regression approach, this approach actually underestimates the standard errors because the unstandardized regression weights across RT bands analyzed in the second step are treated as observed [22]. This leads to an overestimation of the significance, and thus may in turn provide an overly liberal judgement regarding the significance of the WPR. In contrast, the MLM approach treats the unstandardized regression weights across RT bands as estimated and thus provides unbiased standard errors. Taken together, the MLM approach is the more accurate method to analyze the WPR on the level of unstandardized and standardized regression weights. Therefore, we strongly recommend using the MLM approach instead of the sequential regression approach, because the sequential regression approach has serious shortcomings from a statistical perspective.

4. Conclusions

All in all, conceptualizing the WPR as moderation of the effect of g on RT by RT band not only allowed to test the WPR but also provided a quantification of the WPR. We thereby introduced a new way of analyzing the WPR that may overcome the problem of weak power when testing the difference of two correlations, takes the whole shape of increases of unstandardized or standardized estimates across performance bands into account, and additionally quantifies the increases of unstandardized and standardized regression weights across RT bands.

These newly introduced methods for analyzing the WPR suggested that there are perfectly consistent increases in unstandardized regression weights across RT bands in all three experimental conditions. The differences between traditional WPA and results with standardized regression weights to the results with unstandardized regression weights are driven by variations in inter-individual standard deviation of performance across RT bands. And finally, the interaction between RT band and g additionally suggests that general intelligence is related to the shape of the intra-individual performance distribution and not only to the mean performance.

Although these results give promising new insights on the WPR, further evidence with the newly introduced methods with larger sample sizes is needed. Especially because the estimation of the MLM approach is complex and robust estimates of the cross-level interactions can only be obtain in sufficiently large samples. Nonetheless, these results provide preliminary evidence for the feasibility of the newly introduced methods and therefore we hope that researchers will adopt these methods in future studies to gain deeper knowledge of the WPR and its underlying processes.

Nevertheless, by reformulating the WPR in regression termini the two new methods have provided a tool for more sophisticated analyses in future research that aims at an explanation of the WPR. The most accurate method to analyze the WPR would be the MLM approach. Although the sequential regression approach provides equal estimates for the actual increases of unstandardized and standardized regression weights and may be more accessible to use, researchers should bear in mind that the sequential regression approach overestimates the significance of these increases. Therefore, we advise researchers to use the MLM approach and otherwise discuss results from the sequential regression approach reluctantly.

Finally, analyzing the WPR on the level of unstandardized regression weights showed that other context variables, such as the inter-individual standard deviation of performance within each RT band, may affect the results when investigating the WPR on the level of correlations. Therefore future studies should consider analyzing the WPR on the level of unstandardized and standardized regression weights, in order to gain further insight into methodological issues, such as increases in inter-individual standard deviation from best to worst performance, that are related to the WPR.

Beyond that, the here presented methods are not restricted to WPA. In fact, these methodological approaches can be implemented within any setting where a certain result is moderated by a continuous third variable, especially with nested data structures. For instance, a researcher wants to evaluate the relationship between processing speed and general intelligence across different age groups in a longitudinal study. Increases or decreases of this relationship with age could be modeled in a sequential regression approach or multi-level models as well. In that sense, the MLM approach presented for the WPA is only one example where such methods can provide powerful and interesting results.

Future research in search for constructs or processes that mediate the WPR effect, could test whether there are differences in the WPR between trials with and without lapses in attention. Further it would be interesting to know whether the relationship between diffusion model parameters and processing speed affects the outcome of these newly introduced WPA. Results from these studies may take further steps towards a refined understanding of the WPR. Because the WPR is stronger in tasks highly related to g [3,17], this may ultimately present a chance for a better understanding of the processes underlying general intelligence.

Acknowledgments: We thank one anonymous reviewer for pointing out that covariance between RT and g may increase from best to worst performance bands partly due to increases in variance of RT from best to worst performance bands and correlations between performance bands. This is an important point and enriched the discussion of WPA on the level of unstandardized versus standardized regression weights.

Author Contributions: Anna-Lena Schubert and Dirk Hagemann designed the experiment. Gidon T. Frischkorn prepared the data. Gidon T. Frischkorn and Andreas B. Neubauer analyzed the data. Gidon T. Frischkorn wrote the paper. Anna-Lena Schubert, Dirk Hagemann, and Andreas B. Neubauer revised and commented on the paper.

Conflicts of Interest: All authors declare no conflict of interest.

References

1. Sheppard, L.D.; Vernon, P.A. Intelligence and speed of information-processing: A review of 50 years of research. *Personal. Individ. Differ.* **2008**, *44*, 535–551.
2. Schubert, A.L.; Hagemann, D.; Voss, A.; Schankin, A.; Bergmann, K. Decomposing the relationship between mental speed and mental abilities. *Intelligence* **2015**, *51*, 28–46.
3. Coyle, T.R. A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence* **2003**, *31*, 567–587.
4. Larson, G.E.; Alderton, D.L. Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence* **1990**, *14*, 309–325.
5. Kranzler, J.H. A test of Larson and Alterton's (1990) worst performance rule of reaction time variability. *Personal. Individ. Differ.* **1992**, *13*, 255–261.
6. Diascro, M.N.; Brody, N. Serial versus parallel processing in visual search tasks and IQ. *Personal. Individ. Differ.* **1993**, *14*, 243–245.

7. Coyle, T.R. IQ is related to the worst performance rule in a memory task involving children. *Intelligence* **2001**, *29*, 117–129.
8. Coyle, T.R. IQ, the worst performance rule, and Spearman's law: A reanalysis and extension. *Intelligence* **2003**, *31*, 473–489.
9. Fernandez, S.; Fagor, D.; Dirk, J.; de Ribaupierr, A. Generalization of the worst performance rule across the lifespan. *Intelligence* **2014**, *42*, 31–43.
10. Rammsayer, T.H.; Troche, S.J. Validity of the Worst Performance Rule as a Function of Task Complexity and Psychometric *g*: On the Crucial Role of *g* Saturation. *J. Intell.* **2016**, *4*, 5.
11. Salthouse, T.A. Relation of Successive Percentiles of Reaction Time Distributions to Cognitive Variables and Adult Age. *Intelligence* **1998**, *26*, 153–166.
12. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; Lawrence Erlbaum Associates: Hilsdale, NJ, USA, 1988.
13. Jensen, A.R. The importance of intraindividual variation in reaction time. *Personal. Individ. Differ.* **1992**, *13*, 869–881.
14. Jensen, A.R. *Clocking the Mind: Mental Chronometry and Individual Differences*; Elsevier: Oxford, UK, 2006.
15. Rammsayer, T.H.; Troche, S.J. Effects of age and the relationship between response time measures and psychometric intelligence in younger adults. *Personal. Individ. Differ.* **2010**, *48*, 49–53.
16. Doebler, P.; Scheffler, B. The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learn. Individ. Differ.* **2015**, in press.
17. Schubert, A.L.; Frischkorn, G.T.; Hagemann, D. A Systematic Replication of the Worst Performance Rule. Available online: https://osf.io/ekzrv (accessed on 30 June 2015).
18. Ratcliff, R.; Schmiedek, F.; McKoon, G. A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence* **2008**, *36*, 10–17.
19. Ratcliff, R.; Thapar, A.; McKoon, G. Individual Differences, Aging, and IQ in Two-Choice Tasks. *Cogn. Psychol.* **2010**, *60*, 127–157.
20. Unsworth, N.; Redick, T.S.; Lakey, C.E.; Young, D.L. Lapses in sustained attention and their relation to excutive control and fluid abilities: An individual differences investigation. *Intelligence* **2010**, *38*, 111–122.
21. McVay, J.C.; Kane, M.J. Drifting From Slow to "D'oh!": Working Memory Capacity and Mind Wandering Predict Extreme Reaction Times and Executive Control Errors. *J. Exp. Psychol. Learn. Mem. Cogn.* **2012**, *38*, 525–549.
22. Skrondal, A.; Laake, P. Regression among factor scores. *Psychometrika* **2001**, *66*, 563–575.
23. Cohen, J.; Cohen, P.; West, S.; Aiken, L. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed.; Routledge: London, UK, 2002.
24. Wilcox, R.R.; Tian, T. Comparing Dependent Correlations. *J. Gen. Psychol.* **2008**, *135*, 105–112.
25. Diedenhofen, B.; Musch, J. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLoS ONE* **2015**, *10*, e0121945.
26. Bauer, D.J.; Preacher, K.J.; Gil, K.M. Conceptualizing and Testing Random Indirect Effects and Moderated Mediation in Multilevel Models: New Procedures and Recommendations. *Psychol. Methods* **2006**, *11*, 142–163.
27. Preacher, K.J.; Zyphur, M.J.; Zhang, Z. A General Multilevel SEM Framework for Assessing Multilevel Mediation. *Psychol. Methods* **2010**, *15*, 209–233.
28. Wainer, H. The Centercept: An Estimable and Meaningful Regression Parameter. *Psychol. Sci.* **2000**, *11*, 434–436.
29. Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *Am. Sci.* **1969**, *57*, 421–457.
30. Jäger, A.; Süß, H.M.; Beauducel, A. *Berliner Intelligenzstruktur-Test: Form 4*; Hogrefe: Göttingen, Germany, 1997.
31. Gignac, G.E. Raven's is not a pure measure of general intelligence: Implications for *g* factor theory and the brief measurement of *g*. *Intelligence* **2015**, *52*, 71–79.
32. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2015.

33. Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D.; R Core Team. nlme: Linear and Nonlinear Mixed Effects Models; R Package Version 3.1-128. Available online: http://CRAN.R-project.org/package=nlme (accessed on 10 May 2016).
34. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA, 2002.

Processing Speed, Working Memory, and Executive Functions: Independent or inter-related predictors of general intelligence

Gidon T. Frischkorn*, Anna-Lena Schubert & Dirk Hagemann

Heidelberg University, Institute of Psychology

Author Note

We want to thank our research assistants – Ben Riemenschneider, Christoph Löffler, Isabel Gebhardt and Larissa Kunoff – who supported us with the data collection and preparation for this project. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Correspondence concerning this article should be addressed to Gidon Frischkorn, Psychologisches Institut, Ruprecht-Karls-Universität Heidelberg, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, email: gidon.frischkorn@psychologie.uni-heidelberg.de, telephone: +49-6221-54-7746

Abstract

Both working memory capacity (WMC) and processing speed (PS) have been discussed as important covariates of individual differences in intelligence. Recent results indicated that especially latencies of ERP components associated with higher-order processing (P2, N2, and P3) may share up to 80% of variance with individual differences in intelligence. WMC has a similar predictive power and thus these two processes cannot explain individual differences in intelligence independently. The current study explores in how far individual differences in executive functions (EFs) may bridge the gap between WMC and PS as predictors of intelligence. We recruited 101 participants who completed three EF tasks – one for each of the three executive functions shifting, updating, and inhibition – while an EEG was recorded. Additionally, we assessed participants' intelligence, WMC, and PS. Results showed that only variance of behavioral RTs consistent across manipulations in the EF tasks was related to WMC, PS, and intelligence. While P3 latencies were not associated with intelligence, they showed significant correlation with WMC and PS, and N1 latencies showed no correlation with any of the three covariates. The variance specific to the manipulations in EF tasks was small for both behavioral RTs and ERP latencies and showed no consistent correlations with each other or with any of the three covariates. These results suggest that EF tasks capture mostly manipulation-unspecific cognitive processes. Hence, individual differences in the impairment due to additional executive processing demands cannot explain why WMC and PS are related predictors of individual differences in intelligence.

Keywords: Intelligence; Processing Speed; Working Memory; Executive Functions; EEG;

Processing Speed, Working Memory, and Executive Functions: Independent or inter-related predictors of general intelligence

The currently most discussed cognitive processes underling individual differences in general intelligence (*g)* are speed of information processing (Jensen, 2006; Schubert, Hagemann, & Frischkorn, 2017; Sheppard & Vernon, 2008), working memory capacity (Ackerman, Beier, & Boyle, 2005; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Kane, Hambrick, & Conway, 2005), and executive functions (Friedman et al., 2006; Jewsbury, Bowden, & Strauss, 2016; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). These three processing domains were often discussed separately or as independent predictors of individual differences in intelligence, focusing on the question which of the processes shows the largest relationship to individual differences in intelligence. Especially regarding processing speed and working memory capacity as predictors of *g*, some researchers argued for them being independent predictors (Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008), while others showed considerable correlations between these two processes (Ackerman, Beier, & Boyle, 2002; Kyllonen & Christal, 1990; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schmitz & Wilhelm, 2016).

Recent results indicated that the neural speed of information processing of higher order cognitive processes explains up to 80% of variance in intelligence (Schubert, Hagemann, & Frischkorn, 2017), matching the amount of variance in intelligence often explained by working memory capacity (Kyllonen & Christal, 1990; Oberauer, Schulze, Wilhelm, & Süß, 2005). This finding strongly suggests that these two processes cannot be independent predictors of individual differences in intelligence. Hence, it follows that speed of information processing and the capacity of working memory have to be inter-related and might be constrained by similar features of the neuro-cognitive system (Dang, Braeken, Colom, Ferrer, & Liu, 2015). The aim of the present study is to bridge the gap between working memory capacity (WMC) and processing speed (PS) as predictors of *g*. On the basis of current theories of working memory (Barrouillet, Portrat, & Camos, 2011; Cowan, 2017; Oberauer, 2002; Oberauer & Kliegl, 2006) that emphasize the role of attentional processes the present study explores in how far executive functions (Miyake et al., 2000) may explain the inter-relation between WMC and PS as predictors of *g*.

The relationship of processing speed with intelligence

Across a variety of different measures there is a consistent negative relationship between speed of information processing and general intelligence (Jensen, 2006). A review of a broad variety of tasks reported an average correlation of r = -.24 of single task measures of PS and *g* (Sheppard & Vernon, 2008). These correlations tend to increase (r = -.40 to -.50) when reaction times from different tasks are aggregated (Kranzler & Jensen, 1991; Schmitz & Wilhelm, 2016; Schubert, Hagemann, & Frischkorn, 2017), indicating that foremost variance shared between different measures for PS is related to *g*. Moreover, the correlation between PS and *g* increases even further, if measures specifically representing the speed of the decision process, such as the *drift rate* from the drift-diffusion model, are used to represent processing speed (Schmiedek et al., 2007).

Separating encoding processes and motor execution from the actual decision process via cognitive modeling is a promising approach to further investigate the relationship between speed of information processing and intelligence (Frischkorn & Schubert, 2018). Specifically, the drift-diffusion model (DDM; Ratcliff, 1978) has often been successfully used to separate the speed of information accumulation, represented by the *drift rate*, from other processes included in reaction times, such as encoding, motor execution or response caution. Results indicated that drift rates show trait-like properties (Schubert, Frischkorn, Hagemann, & Voss, 2016) and show consistent relationships with intelligence ranging from r = .50 to .90 (Ratcliff, Thapar, & McKoon, 2010; Schmiedek et al., 2007; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015). This indicates that it is precisely the speed of information accumulation that is related to general intelligence and not speed of motor execution or encoding.

However, PS can not only be measured via behavioral indicators such as reaction times, but also with neurophysiological indicators such as latencies of event-related potential components (Verleger, 1997). The event-related potential (ERP) decomposes the stream of neural information processing from stimulus onset until the response into distinct components that can be linked to specific cognitive functions. Specifically, individual differences in the latency of an ERP component may reflect individual differences in the neural speed of information processing, meaning that a higher speed of information processing results in shorter latencies of an ERP components. Moreover, the onset of an ERP component may also serve as an indicator of its functionality. While ERP components occurring early after stimulus onset are mostly related to stimulus encoding, ERP components with a later onset are foremost connected to higher-order processing.

Despite rather weak and inconsistent results on the relationship between latencies of ERP components and intelligence (Schulter & Neubauer, 2005), a number of empirical results showed consistent negative relations between the latency of the P3 component or the mismatch negativity (MMN) and intelligence (Bazana & Stelmack, 2002; McGarry-Roberts, Stelmack, & Campbell, 1992). In detail, more intelligent people displayed shorter latencies of the P3 (Bazana

& Stelmack, 2002; McGarry-Roberts et al., 1992; Troche, Houlihan, Stelmack, & Rammsayer, 2009), which is an ERP component that is associated with stimulus evaluation and categorization, context updating, and context closure. Furthermore, this relationship of latencies of ERP components with intelligence is mediated via behavioral RTs (Schubert et al., 2015) suggesting that neural processing speed may functionally underlie faster information processing on a behavioral level.

The mostly inconsistent relationship of ERP latencies with intelligence can be attributed to different problems, such as (1) small sample sizes (N < 50), (2) questionable selection of electrode sites for the measurement of ERP latencies, and (3) quantification of relationships with single task measures that confound task-related fluctuations with the trait-like neural processing speed of a person. A recent study addressing these issues by measuring ERP latencies for three different tasks at two measurement occasions could demonstrate that the shared variance of later ERP components (P2, N2, & P3) explained up to 65% of variance in general intelligence (Schubert, Hagemann, & Frischkorn, 2017). Moreover, the results of this study indicated that the variance of single task latencies of ERP components included a large proportion of task- and condition-specific variance, which may be irrelevant to the relationship of ERP latencies with general intelligence, as well as substantial unsystematic error variance. This finding may explain the inconsistent results from earlier studies using only single task measures. Taken together, all these results from behavioral and neuro-psychological studies indicate that there is a strong relationship between speed of higher order information processing and intelligence.

The relationship of working memory capacity with intelligence

In addition to speed of information processing, also working memory capacity has been closely linked to individual differences in intelligence (Ackerman et al., 2005; Kane et al., 2005;

Oberauer et al., 2005). Comparable to cognitive ability tasks, measures of working memory capacity (WMC) are highly correlated with each other (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004; Oberauer, Süß, Wilhelm, & Wittman, 2003; Unsworth, Fukuda, Awh, & Vogel, 2014) and resemble a hierarchical structure, with a broad single factor at the highest level, and more domain and process specific factors on the lower level (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Kane et al., 2004; Oberauer et al., 2003; Shah & Miyake, 1996). Correlations between the broad factor of WMC and *g* are very high, ranging from r = .70 to .90 (Conway et al., 2002; Kane et al., 2005; Kyllonen & Christal, 1990; Oberauer et al., 2005). These high correlations have led to a vivid discussion in how far WMC and intelligence may be isomorphic (Ackerman et al., 2005; Kane et al., 2005; Oberauer et al., 2005), ultimately resolved by concluding WMC explains a large proportion of individual differences in intelligence.

In addition to these correlational studies there are results from an experimental study suggesting that overloading working memory while performing a test of fluid intelligence affected performance in the intelligence measure (Rao & Baddeley, 2013). In detail, participants were asked to remember a set of three digits and count backwards while working on a matrix reasoning task. Results showed that especially the time needed to solve an item increased compared to a silent control and an articulatory suppression condition. Altogether these results suggest that working memory is not only strongly related to individual differences in intelligence but may actually causally underlie variations in *g*.

While this strong relationship between working memory and intelligence is undisputed, researchers do not agree which process within working memory is central to the relationship of working memory and intelligence. Some researchers argue that the relationship is best explained by similar demands on short-term memory storage (Colom et al., 2008; Colom, Flores-Mendoza, Quiroga, & Privado, 2005; Shahabi, Abad, & Colom, 2014), and others argue that processes specific to complex span tasks such as attention regulation are the reason for the strong association between working memory and intelligence (Conway et al., 2002; Unsworth et al., 2014). As there is robust evidence that the capacity of working memory is related to attentional processes (Chuderski, Taraday, Nęcka, & Smoleń, 2012; Kane & Engle, 2003; McVay & Kane, 2009, 2012; Meier & Kane, 2013) and current theories of working memory assume that attention plays a major role in maintenance of memory items regardless of concurrent processing (Oberauer, Farrell, Jarrold, & Lewandowsky, 2016; Souza & Vergauwe, 2018), it is plausible that both the capacity of short-term memory storage and additional demands in complex span tasks are strongly related to the same attentional processes within working memory (Barrouillet et al., 2011; Wilhelm, Hildebrandt, & Oberauer, 2013).

Executive Functions: Bridging the gap between processing speed and working memory?

Candidates for the attentional processes underlying both maintenance of items in short-term memory as well as additional demands of complex span tasks are executive functions. Executive functions (EFs) are defined as attentional control mechanisms (Karr et al., 2018; Miyake et al., 2000) that are used to (a) focus attention on relevant information while ignoring irrelevant information (i.e. inhibition), (b) encode new information to memory while removing outdated and no longer relevant information (i.e. updating), or (c) switch between different tasks (i.e. shifting). While it is still under debate in how far these different EFs have to be separated or share common variance (Friedman & Miyake, 2017; Miyake & Friedman, 2012), the majority of results suggests that there is considerable overlap between the three EFs (Karr et al., 2018). Moreover, EFs have recently been subsumed within the hierarchical structure of intelligence (Jewsbury et al., 2016). In detail, updating was integrated within a general memory factor g_m, while shifting and inhibition were integrated in the general speed factor g_s, pointing towards relations of EFs with both memory and processing speed.

With respect to speed of information processing, the results from Schubert et al. (2017) support the proposal that executive functions may underlie the relationship of processing speed with intelligence as well. In detail, the latency of the P3 component showed the strongest association with general intelligence. The P3 component has often been interpreted as an indicator of the efficiency of context-updating (Donchin, 1981; Polich, 2007), and thus shorter latencies of the P3 may reflect a faster inhibition of nonessential processes that in turn ease the transmission of information from attention and working memory regions located frontally in the brain to parietal memory storage processes (Polich, 2007). There is additional support for this hypothesis from behavioral studies showing strongest relations between inhibition and updating with intelligence (Wongupparaj, Kumari, & Morris, 2015).

Beyond that, cognitive as well as neural theories of intelligence are in line with this theoretical perspective. The process-overlap theory (POT; Kovacs & Conway, 2016) assumes that attentional control mechanisms are among the domain-general processes that act as a bottleneck constraining performance in a broad range of cognitive tasks. Moreover, the parieto-frontal integration theory (P-FIT; Jung & Haier, 2007) proposes that individual differences in the efficiency of information transmission from frontal association cortices and parietal brain regions may explain individual differences in *g*. P-FIT has been widely supported by results from structural and functional neuroimaging studies (Burgess, Gray, Conway, & Braver, 2011; Colom, Jung, & Haier, 2007; Colom & Thompson, 2013; Gläscher et al., 2010). Altogether, attention regulation mechanisms such as EFs may provide a theoretical account to bridge the gap between processing speed and working memory as predictors of individual differences in *g*.

The present study

We conducted the present study to investigate in how far individual differences in executive functions (EFs) may underlie the relationship of processing speed (PS) and working memory capacity (WMC) with intelligence (*g*). To that end, we administrated three different EF tasks each tapping one of the executive functions (i.e. shifting, updating, and inhibition). To further differentiate between behavioral and neurophysiological indicators of executive functions we recorded the EEG while participants worked on these three EF tasks. Additionally, we assessed participants' general intelligence, their working memory capacity, and speed of information processing to investigate in how far individual differences in EF tasks explain the relationship between these three constructs.

More specifically, we aimed to address two major points: First, we wanted to investigate in how far performance in the different experimental conditions of EF tasks measure performance specific to the manipulations that are related to the respective EFs, or performance that is unspecific with respect to the experimental manipulations. And second, we were interested how these two components of performance in EF tasks are related to WMC, PS, and *g*. All in all, joining the individual differences constructs of intelligence, WMC, and PS with executive functions may provide insights in how far individual differences in EFs may explain the relationship between WMC and PS as predictors of intelligence.

Methods

Sample

We recruited a community sample of 107 participants via newspaper ads and flyers. The 101 participants[1] who attended both sessions of the experiment were on average 39.1 years (SD_{age} = 14.5, Min_{age} = 18, Max_{age} = 61), and 52.5 % were female. Participants had a heterogeneous educational background (42.6% university degree, 42.6% college degree, 10.9 % high school degree, 3.9% did not report educational background) and were rewarded with 50€ for their participation. While the overall sample size is comparably small for structural equation modeling, we secured that it had sufficient power (1-β > .80) to assess model fit with the RMSEA ($H0_{\text{RMSEA}}$ =.05, $H1_{\text{RMSEA}}$ = .10, df_{Model} = 50, α = .05, N_{Min} = 97).

General Procedure

The study consisted of two sessions that were approximately four months apart. In the first occasion, participants completed three executive functioning tasks – a Switching task, an *N*-Back task, and the Attention Network Test – while an EEG was recorded. For this occasion, participants were seated in a dimly lit, sound-attenuated EEG cabin and tested individually. In the second session, participants' intelligence, working memory capacity, and processing speed were measured with paper-pencil tests and computerized tasks. In addition, participants completed three knowledge tests and two personality questionnaires not reported here. This session was conducted in groups of up to four participants. Both sessions took approximately 3 hours, and the sequence of tasks within the two occasions was the same for all participants.

[1] Only data from the participants who attended both sessions was analyzed and reported in the manuscript.

Measures

Executive Functions. The three executive functions – shifting, updating, and inhibition – were each assessed with one task. To additionally asses in how far all executive functions rely on the same attention process conceptualized as executive attention by Posner and Petersen (1990), a flanker manipulation was implemented in each of the three tasks (Eriksen & Eriksen, 1974).

Switching Task. The Switching task was adapted from Sauseng et al. (2006). An illustration of the trial procedure can be found in Figure 1A (p. 50). In this task, participants saw a digit from 1 to 9 (except the number 5) colored red or green that was presented centrally on a black screen. Participants either had to decide whether the digit was smaller or larger than five or whether it was odd or even, depending on the color of the presented number. Prior to the onset of the target stimulus, a light grey fixation cross was presented centrally on the screen for 400-600ms. Between onset of the target stimulus and the fixation cross, an inter-stimulus interval (ISI) consisting of a blank screen was shown for 400-600ms that was presented until participants responded and stayed on the screen for another 500ms to avoid offset potentials in the EEG due to perceptual changes on the screen. Participants responded via keypresses on the keyboard by pressing a left key “d” if the digit was smaller than five or odd, and a right key “l” if the number was larger than five or even. Before the next trial started, there was an inter-trial interval (ITI) of 1000-1500ms.

The switching task consisted of four different blocks. In the first two blocks, participants had to decide whether the number was less or more than 5 in one block, or odd or even in the other block, irrespective of the stimulus color. These *control blocks* consisting of 48 trials (8 digits x 2 colors x 3 repetitions) did not require any task switching and were used to quantify global switch costs in comparison to the two switching blocks. In the third block, participants were instructed to decide whether the digit was smaller or bigger than 5 for red stimuli and odd or even for green stimuli. This *shifting* block consisted of 96 trials (8 digits x 2 tasks x 2 shifting x 3 repetitions) of which the to-be-conducted task switched in half of the trials. In the last block, additional flanker stimuli that were congruent, neutral, or incongruent to the target stimulus were added to the task and participants were instructed to ignore the flankers while completing the same task as in the *shifting* block. This *shifting flanker* block consisted of 288 trials (8 digits x 2 tasks x 2 shifting x 3 flanker x 3 repetitions). The color of flanker and target stimuli was always the same, therefore congruency of the flankers was only manipulated on the content level (i.e. the numerical information), but not on the task cue level. Incongruent flankers were always a different digit than the target stimulus, but could indicate the same response as the target stimulus (e.g., red 8 as flankers and a red 6 as target are both larger than 5).

All participants completed 16 practice trials per block. The experimental trials within all blocks were pseudo-randomly sorted following some constraints: In all blocks, digits were not allowed to repeat more than three times in a row. Likewise, stimulus color, and thus tasks in the switching blocks, and responses were not allowed to repeat more than three times in a row.

N-Back Task. Participants completed a 2-Back task that was adapted from Scharinger et al. (2015). The trial procedure of the *N*-Back task can be seen in Figure 1B (p. 50). Participants saw a series of light grey letters (H, C, F, or S) shown centrally on the screen one after the other. For each letter, participants decided whether or not it was identical to the letter presented two steps before. Between subsequent letters, there was an ISI consisting of a blank screen that was shown for 1000-1500ms. The letters were always shown for 2500ms irrespective of the time participant needed to respond. This way we ensured that all participants had the same time to encode the new letter and decide whether it matched the letter two steps back. Additionally, by not changing the perceptual input after participants' responses, we avoided offset potentials in the EEG. Participants responded via keypress, pressing a left key "d" if the current letter did not match the letter two steps back, and pressing a right key "l" if the current letter matched the letter two steps back.

The *N*-Back task consisted of two blocks. In the *2-back* block, participants completed 96 trials (4 letters x 2 match x 12 repetitions) of the 2-Back task, preceded by two introductory trials requiring no response as there were no letters two steps prior to presentation. In the *2-back flanker* block, participants completed 384 trials (4 letters x 2 match x 4 flanker x 12 repetitions) of the 2-Back task with additional flanker stimuli. Unlike in the Shifting Task, there were four levels of the flanker manipulation within this block. There either were no flanker stimuli – like in the 2-back block – or flanker stimuli that were either congruent, neutral or incongruent. Moreover, the flanker block was separated into three sub-blocks consisting of 128 trials, in order to give participants short breaks. Like in the 2-back block, the experimental trials of all sub-blocks were preceded by two introductory trials requiring no response.

Participants completed 16 practice trials that were repeated until participants' average accuracy in these 16 practice trials was above chance. Experimental trials were pseudo-randomly sorted with the constraint that responses and thus the match conditions were allowed to repeat a maximum of two times. Additionally, flanker congruency was not allowed to repeat more than two times in the flanker block.

Attention Network Test (ANT). In the Attention Network Test (Fan et al.), participants had to decide whether an arrow pointed left or right (see Figure 1C for the trial procedure). The arrow (i.e., the target stimulus) could appear above or below a fixation cross that was located centrally on the screen. Furthermore, the centrally presented arrow was flanked by two more arrows to the left and right. The flanking arrows were either pointing in the same direction (congruent), in the other direction (incongruent), or were without arrow heads indicating any direction (neutral).

Each trial started with a light grey fixation cross presented centrally on the black screen for 400-1600ms, followed by a short cue stimulus presented for 100ms. There were four different cue options: (1) There was no cue and the fixation cross remained on the screen, (2) there was a central cue at the position of the fixation cross, (3) there was a double cue above and below of the fixation cross, or (4) there was a spatial cue located either above or below the fixation cross validly cueing the position of the following target stimulus. Between the cue stimulus and the target stimulus there was an ISI of 400ms with the fixation cross being presented centrally on the screen. Then the target stimulus, i.e. the central arrow, and flanker arrows appeared above or below the fixation cross on the screen and participant had to decide whether the central arrow pointed right or left. Participants responded via key press, pressing the left key “d” if the arrow pointed left, and the right key “l” if the arrow pointed right. The target stimulus and flanker stimuli remained on-screen for 1700ms, irrespective of the speed of the response. Before the next trial started, there was an ITI again consisting of the light grey fixation cross presented centrally on the screen that lasted 1700ms.

The ANT consisted of three blocks of 96 trials each (2 direction x 2 location x 4 cue x 3 flanker x 2 repetitions) that were pseudo-randomly sorted. Specifically, all of the four experimental factors were allowed to repeat a maximum of three times in subsequent trials. 24 practice trials were conducted prior to the first experimental block. In between blocks, participants took short breaks and read a short reminder of the task instructions.

Intelligence (G*f*). Intelligence was measured with the short-version of the Berlin Intelligence Structure test (BIS; Jäger, Süß, & Beauducel, 1997). The BIS is based on the bimodal Berlin Intelligence Structure model (Jäger, 1982). According to this model, the 15 tasks of the BIS short version are grouped into four operation-related (processing capacity, memory, processing speed, and creativity) and three content-related (verbal, numerical, and figural) components of fluid intelligence. Each task combines one operation-related component with one content-related component of intelligence. The standard scores of the five tasks with verbal, numerical, and figural content were aggregated across operations and used as separate indicators of fluid intelligence.

Working Memory Capacity (WMC). Working memory capacity was measured with four tasks from the working memory test battery by Lewandowsky et al. (2010). Specifically, we used the memory updating task, two complex span tasks, and a spatial short-term memory task. Following the scoring script provided by Lewandowsky et al. (2010), performance was measured by the mean proportion of correctly remembered items for each task separately.

Processing Speed (PS). We measured participants' processing speed with two elementary cognitive tasks (ECTs), the Posner letter matching task and the Sternberg memory scanning task. These tasks are commonly used as an indicator of basic information processing speed.

Posner Letter Matching Task. In the Posner Letter Matching task (Posner & Mitchell, 1967), participants decided whether two letters were physically or semantically identical. The participants completed two different blocks, first a physical identical and second a name identity block. Each of the two blocks consisted of 40 trials preceded by 10 practice trials, in half of which the two letters matched physically or semantically corresponding to the block instructions.

The two letters comprising the target stimulus were selected from a pool of five letters (a, b, f, h, q) that could be capitalized or not. Each trial started with a fixation cross presented centrally on the screen for 1000-1500ms. Immediately after that, the letter pair was shown on the screen. Participants then responded via key press, pressing either a right or left key with their index fingers, indicating whether the two letters were physically or semantically identical. Response mapping of the keys was counterbalanced across participants. After the response, the trial ended and the next trial started after an ITI of 1000-1500ms. We used the mean logarithmized RT of correct responses in the two blocks as a measure of processing speed.

Sternberg Memory Scanning Task. In the Sternberg Memory Scanning task (Sternberg, 1969), participants saw a memory set of digits from zero to nine and had to decide whether a subsequently presented probe digit was contained in the memory set or not. Participants completed two blocks, first one block with memory set size three and second one block with set size five, each consisting of 40 trials, preceded by 10 practice trials. In each block, the probe digit was contained in the memory set in half of the trials. Each trial started with a fixation cross presented centrally on the screen for 1000-1500ms. Then the digits comprising the memory set were presented sequentially on the screen for 1000ms with an ISI of 400-600ms. After the last digit, a question mark was presented for 1800-2200ms, followed by the probe digit. Participants then responded via key press, pressing either a left or right key with their index finger. The response mapping of keys was counterbalanced across participants. After the response, the trial ended and the next trial started after an ISI of 1000-1500ms. Again, we used the mean logarithmized RT of correct responses in the two experimental blocks as an indicator of processing speed.

EEG Recording

In the first session, we recorded the participants' EEG during the three EF tasks (Switching Task, *N*-Back task, and ANT) with 32 equidistant Ag–AgCl electrodes. In addition, we used the aFz as ground electrode, and Cz as a common online reference. The signal of all 32 channels was offline re-referenced to an average reference. We kept all electrode impedances below 5 kΩ, and recorded the EEG signal continuously with a sampling rate of 1000 Hz (band-pass 0.1–100 Hz). The data was filtered offline with a low-pass filter of 12 Hz.

Data Analyses

Behavioral data. To ensure that intra-individual outliers in reaction times measured in EF tasks and ECTs did not distort our results, we discarded trials with RTs shorter than 150ms or longer than 3000ms. Then, we discarded any incorrect trials and trials in which the logarithmized reaction times of correct responses deviated more than 3 SD from the mean logarithmized reaction time of each participant within the different conditions in each task. Finally, we calculated the mean logarithmized reaction time as the dependent variable. As accuracies were very high ($M > .90$) and showed little to no variation in all EF tasks and ECTs (see Table 1), we refrained from analyzing accuracy measures.

For all measures, we conducted additional uni- and multi-variate outlier analyses on the between-person level. First, univariate outliers deviating more than 3 SDs from the mean were deleted, resulting in 0.0 % to a maximum of 2.3 % of subjects being excluded across the different measures. Second, multi-variate outliers on the different measures within each cognitive process were identified. Only in the ANT did the Mahalanobis distance for one participant exceed the critical value of $\chi^2(12) = 39.9$. This subject was excluded case wise for the ANT. Finally, multivariate outliers across the measures of different cognitive processes were again identified via the Mahalanobis distance. As no subject exceeded the critical value of $\chi^2(23) = 49.7$, no data was discarded in this final step.

Electrophysiological data. The event-related potentials (ERPs) were calculated separately for each EF tasks and conditions with the ERPLAB toolbox (Lopez-Calderon & Luck, 2014). All ERPs were time-locked to the stimulus onset with a baseline of 200ms before stimulus onset for the Switching and the N-Back task and 700ms to 500ms for the ANT. The baseline for the ANT was earlier due to the cue stimulus 500ms before stimulus onset that elicited considerable activity in the EEG. After stimulus onset epochs continued for 1000ms, resulting in 1200ms epochs for the Switching and the N-Back task and 1700ms epochs for the ANT. First, channels and epochs with gross artifacts were rejected based on the standard settings implemented in EEGLAB. Second, ocular artifacts and generic discontinuities were corrected via ICA and artifact ICs were identified using the ADJUST algorithm (Delorme & Makeig, 2004; Mognon, Jovicich, Bruzzone, & Buiatti, 2011).

As we were interested in the neural speed of information processing we determined latencies of EPR components instead of evaluating amplitudes that are associated with processing capacity. Furthermore, to differentiate between the neural speed of earlier versus later processes in the neural stream of information processing we analyzed the N1 and P3 latency (see Figure 2 for grand average ERPs). The latency of the N1 was determined at the frontal electrode site over midline and the latency of the P3 was determined at the parietal electrode over midline for all EF tasks. Peak latencies were determined separately for each EF task and the conditions within each task. For participants that did not show a N1 or P3 component in the average ERP, peak latencies were coded as missing in the respective condition of the respective EF task. Finally, univariate-outliers within each condition of the three EF tasks exceeding ±3SDs from the mean latency were discarded. We did not analyze amplitudes of ERP components, as recent research suggest that the latency of ERP components is strongly associated with intelligence (Schubert, Hagemann, & Frischkorn, 2017)

Statistical analyses. As an initial manipulation check, we analyzed whether the experimental manipulations in the EF tasks showed the usual effects on behavioral response times. This ensured that the EFs supposed to be required by specific experimental manipulations were actually demanded within the respective task. In addition, we reported the corresponding results for latency measures of ERP components. For all ANOVAs, we corrected violations of sphericity by adjusting the degrees of freedom with the Greenhous-Geisser correction. For post-hoc comparisons, *p*-values were corrected with the Tukey method.

Following these experimental analyses, we ran structural equations models for behavioral and EEG measures of the EF tasks separately. First, we established separate measurement models for the EF tasks. All of these models were set up as bi-factor models with all indicators across the different blocks and experimental manipulations loading on a general behavioral or neural speed factor, and factors specific to the experimental manipulations or blocks within each EF task. This approach allowed us to separate task-general and manipulation-specific variance in the EF tasks, with the manipulation-specific factors capturing individual differences in executive functions associated with specific experimental manipulations. These bi-factor models will answer the first of the two major points we wanted to investigate within the present study, namely, in how far performance within one condition of an EF task represents general performance, or performance specific to the manipulation that is linked to the respective EF.

The best fitting bi-factor models for the EF tasks were then merged and covariates were entered into the model. In a first step, we analyzed the three covariates – general intelligence, processing speed, and working memory capacity – separately; in a second step, we included all covariates simultaneously to additionally assess the inter-relations between covariates and answer the second of the two major points we wanted to investigate in the present study: How is general and manipulation-specific variance from EF task related to WMC, PS, and intelligence, and do these relationships provide evidence that individual difference in EF might represent the missing link between WMC and PS as a predictor of intelligence?

We assessed model fit of all structural equation models using the comparative fit index (CFI; Bentler, 1990) and the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992). We considered model fit as acceptable with CFI > .90 and RMSEA < .10 (Bentler, 1990; Browne & Cudeck, 1992; Hu & Bentler, 1999). When model evaluation diverged between the two fit criteria, we evaluated model fit with the more favorable fit index, because previous research has shown that goodness-of-fit statistics tend to underestimate absolute model fit in small samples (Kenny, Kaniskan, & McCoach, 2015; Schubert, Hagemann, Voss, & Bergmann, 2017). For comparisons of two models, we required more complex models to show a lower AIC than more parsimonious models with an AIC difference > 10 to retain the more complex model (Burnham & Anderson, 2002). Finally, we assessed statistical significance of model parameters with the two-sided critical ratio test. If parameters did not differ significantly from zero, we fixed them to zero and estimated the SEM again. Thus, only parameters significantly different from zero are reported in the results section.

Results

Manipulation Check: EF tasks

To ensure that experimental manipulations within the EF tasks demanded the respective attentional control mechanisms, we ran within-subject ANOVAs for the three EF tasks. The mean reaction time and proportion of correct responses across the different experimental conditions within the three tasks are displayed in Table 1, descriptive statistics for the latencies of ERP components are displayed in Table 2. Grand Averages of the ERPs are displayed in Figure 2. For brevity, we only report the effect size estimates of the critical manipulations, the full statistical results of the ANOVAs can be found in the supplementary material (osf.io/6trne). For descriptive plots that display the effects of experimental manipulations on behavioral RT in the three EF tasks see Figure 3.

Switching. For reaction times there were substantial global switch costs as indicated by the difference between *control* and both the *shifting* and *shifting flanker* blocks, $\omega_p^2 = .94$. Within shifting blocks, responses times were faster in trials with task repetition than task switches, $\omega_p^2 = .69$, indicating large local switch costs over all conditions (see left part of Figure 3A). While there was a small difference in N1 latency across blocks, $\omega_p^2 = .11$, the direction of the effect contradicted the usual global switch costs (i.e. longer latencies for *shifting* than for *control* blocks). In addition, we obtained no local switch cost on the N1 latency, $\omega_p^2 = .01$. P3 latencies showed no global switch costs, as P3 latency did not vary across blocks, $\omega_p^2 = .02$. However, P3 latency was slightly shorter for switch than for repeat trials, $\omega_p^2 = .04$, contradicting the usual direction for local switch costs. Altogether, the Shifting manipulation was successful for behavioral reaction times.

With respect to the flanker manipulation in the *shifting flanker* block, response times were slowest in trials with incongruent flankers and response times did not differ between trials with neutral and congruent flankers, $\omega_p^2 = .31$ (see right part of Figure 3A). In contrast, neither N1, $\omega_p^2 = .00$, nor P3 latency, $\omega_p^2 = .00$, differed between flanker conditions. Similar to the shifting manipulations, the flankers showed an effect for behavioral reaction that resembles the standard inhibition effect of flanker stimuli (c.f. Eriksen & Eriksen, 1974)

***N*-Back.** Neither behavioral reaction times nor N1 latencies varied between the *2-back* and the *2-back flanker* block, both $\omega_p^2 = .00$. Only P3 latencies were slightly shorter for the *2-back* than for the *2-back flanker* block, $\omega_p^2 = .09$. In contrast, only response times were faster for match than for no-match trials in both blocks, $\omega_p^2 = .87$ (see left part of Figure 3B). This is a common difference between match and no match retrievals from memory. While N1 did not differ between the two match conditions, $\omega_p^2 = .00$, P3 latencies were slightly shorter for no match than for match trials, $\omega_p^2 = .07$, contradicting the usual direction of this effect.

All three dependent variables varied between flanker conditions in the flanker block, with the largest effect for behavioral response times, $\omega_p^2 = .60$ (see right part of Figure 3B), and slightly smaller effects for N1, $\omega_p^2 = .21$, and P3 latency, $\omega_p^2 = .14$. However, this flanker effect was mostly due to differences between the neutral and no flanker condition with congruent or incongruent flankers. The critical inhibition effect – i.e. longer reaction times in incongruent than in congruent trials – was, however, only present for behavioral reaction times in match trials. In sum, there was no consistent inhibition effect of flanker stimuli in the N-Back task on behavioral RTs.

ANT. Overall, participants' response time as well as ERP latencies varied across cue conditions (see Figure 3C). Again this effect was largest for behavioral response time, $\omega_p^2 = .87$, and smaller for N1, $\omega_p^2 = .49$, as well as P3 latency, $\omega_p^2 = .14$. While all three dependent variables showed an alerting effect (i.e. longer reaction times/latencies for no cues than for double cues), an orienting effect (i.e. shorter latencies for spatial than for central cues) was only present for behavioral reaction times. More importantly, behavioral reaction times showed a strong inhibition effect, $\omega_p^2 = .93$, with slowest response times for trials with incongruent flankers compared to congruent or neutral flankers. However, neither N1, $\omega_p^2 = .04$, nor P3 latency, $\omega_p^2 = .05$, showed this inhibition effect. In sum, the manipulation of inhibition again was successful for behavioral reaction times.

SEM Analysis: Bi-factor models for the EF tasks

Switching task. The bi-factor model capturing general and condition-specific variance of behavioral RTs in the Shifting task (see Figure 4A) fitted well to the data, $\chi^2(53) = 41.1$, $p <$ .884, CFI = 1.00, RMSEA = .00, 90% CI = [.00, .03]. The general processing speed factor explained between 43 to 76% of the variance of the manifest indicators, while all condition specific factors together explained between 36 to 51% of the variance in the manifest indicators. In detail, the global shifting factor explained between 6 to 51%, the local shifting factor 9 to 10%, and the flanker factor between 21 to 23% of the variance in the manifest indicators. Additional factors for inhibition or facilitation effects of flanker stimuli had non-significant variances and were thus not included in the final model. Taken together, between 76 to 95% of variance in manifest indicators was explained by both manipulation-specific factors and the general factor.

The bi-factor model for the N1 latency in the Shifting task (see Figure 5A) showed a good fit to the data, $\chi^2(60) = 51.7$, $p < .767$, CFI = 1.00, RMSEA = .00, 90% CI = [.00, .05]. In detail, the general factor of neural information processing speed captured 71 to 88 % of variance of manifest N1 latencies across the different conditions. Furthermore, only the factor capturing individual differences in the flanker effect had a variance significantly different from zero, capturing 8 to 9 % of variance in manifest N1 latencies. The other manipulation-specific factors had variances not significantly different from zero and were not included in the model. In sum, manipulation-specific factors and the general factor explained between 77 to 89% of variance in manifest N1 latencies.

Likewise, the bi-factor model for P3 latencies (see Figure 6A) showed a good fit to the data, $\chi^2(56) = 47.3$, $p < .983$, CFI = 1.00, RMSEA = .00, 90% CI = [.00, .05]. In specific, the overall speed factor explained between 61 to 93 % of variance in manifest P3 latencies, while all condition specific factors together captured between 11 to 25% of variance in manifest P3 latencies. However, there was only a condition-specific factor for global switch cost that captured between 9 to 11%, and a flanker-specific factor that captured between 13 to 14% of variance in manifest P3 latencies. Variances of manipulation-specific factors for local switch costs, facilitation or inhibition had non-significant variances. In this, all these factors together explained between 81 to 93% of variance in manifest P3 latencies.

***N*-Back task.** The bi-factor model for the behavioral RTs in the *N*-Back task (see Figure 4B) fitted well to the data, $\chi^2(56) = 60.2$, $p < .326$, CFI = 1.00, RMSEA = .03, 90% CI = [.00, .07]. The general processing speed factor again explained the largest proportion of variance in manifest RTs in the *N*-Back task with 43 to 81%. All condition-specific factors together explained 18 to 52% of variance. Specifically, the flanker factor explained between 27 to 29%, the no match factor explained between 18 to 21% and the facilitation factor 3% of the variance in manifest indicators. The factor for inhibition effects of flanker stimuli showed a non-significant variance and was thus not included in the model. All these factors together explained between 81 to 95% of variance in manifest response times.

For the N1 latency in the N-Back task (see Figure 5B), the bi-factor model showed a good fit to the data as well, $\chi^2(56) = 52.2$, $p < .619$, CFI = 1.00, RMSEA = .00, 90% CI = [.00, .06]. The general speed factor captured between 44 to 80 % of variance in manifest N1 latencies, and only the flanker-specific factor had a variance significantly different from zero and captured 8 to 9 % of variance in manifest N1 latencies. Apart from that, none of the other manipulation-specific factor had a variance significantly different from zero. Thus, the two factors with significant variances explained between 52 to 92% of variance in manifest N1 latencies.

For the P3 latency, the bi-factor model (see Figure 6B) fit the data acceptably, $\chi^2(55) = 72.0$, $p < .062$, CFI = .98, RMSEA = .06, 90% CI = [.00, .09]. The general neural speed factor explained between 62 to 81% of variance, while all condition-specific factors explained between 8 to 22 % of variance in manifest P3 latencies. In detail, both the factor for no match trials and the factor for the flanker manipulation explained 8 to 10% of variance and the factor for the facilitation effect of flanker 7% of variance in manifest P3 latencies. The factors for inhibition effects of the flanker stimuli did not differ from zero and was not included in the final model. Together these factors explained between 80 to 89% of variance in manifest P3 latencies.

ANT. Consistent with results of the two other EF tasks, the bi-factor model for behavioral RTs in the ANT (see Figure 4C) showed a good fit as well, $\chi^2(83) = 91.8$, $p < .238$, CFI = 1.00, RMSEA = .03, 90% CI = [.00, .07]. Again, consistent with results from the other two EF tasks, the general processing speed factor explained between 86 to 96% of variance in manifest indicators. Contrary to results from the other two EF tasks, the condition-specific factors explained a lower amount of variance in manifest variables with only 2 to 9%.

Specifically, the inhibition factor explained 7% and the no cue and spatial cue factor about 2% of variance in manifest variables. Condition-specific factors for the other cue conditions and for facilitation showed non-significant variances and were not included in the model. Taken together, all factors explained between 93 and 96% of variance in manifest response times.

The bi-factor model for N1 latencies (see Figure 5C) fit the data acceptable, $\chi^2(83) = 114.7$, $p < .012$, CFI = .942, RMSEA = .06, 90% CI = [.03, .09]. In this model, the general factor of N1 latencies across all experimental manipulations captured between 36 to 62% of variance in manifest N1 latencies, while the only experimental factor significantly differing from zero, the factor for spatial cues, captured 17 to 18% of variances in the respective manifest N1 latencies. All other manipulation-specific factors did not have variance significantly different from zero. Altogether these factors explained between 47 to 64% of variance in manifest N1 latencies.

For the P3 latency in the ANT, the bi-factor model (see Figure 6C) showed a good fit to the data as well, $\chi^2(83) = 79.7$, $p < .581$, CFI = 1.00, RMSEA = .00, 90% CI = [.00, .06]. With 50 to 72% of variance the general factor capturing variance consistent across all experimental manipulation explained most of the variance in manifest P3 latencies. In this model, the factor for the no cue condition capturing 13% and the factor for the spatial cue condition capturing 19 to 20% of variance had variances significantly different from zero. The other manipulation-specific factors did not have a variance significantly different from zero. In sum, these factors explained between 62 to 75% of variance in manifest P3 latencies.

SEM Analysis: Relationship of EFs with WMC, PS, and Intelligence

Joint Models for the three EF tasks. A joint model of behavioral reaction times in the three EF tasks indicated only correlations between the three general factors measured in the three EF tasks (r_S = .37 – .76). Additionally estimating correlations between manipulation-specific factors of the three EF tasks did not improve model fit, ΔAIC = -30.8, $\Delta\chi^2(27) = 23.2$, $p = .672$. Moreover, joining the three general processing speed factors into one factor did not impair model fit, ΔAIC = -1.8, $\Delta\chi^2(1) = 0.2$, $p = .643$, and represented a more parsimonious account of the covariance structure. Specifically, the task-general factor explained all variance of the general factor from the shifting task, 56 % of variance of the general ANT factor, and 27 % of variance of the general N-Back factor. As this model showed a good fit to the data, $\chi^2(528) = 647.8$, $p < .001$, CFI = .98, RMSEA = .05, 90% CI = [.0533, .06], it was retained for further analyses with the three covariates.

Similarly, the joint model for N1 latencies indicated correlations between the general N1 factors in all three EF tasks (r_s = .62 - .93). Exploratory analysis revealed that additionally estimating a correlation between the flanker factor from the shifting task with the flanker factor from the *N*-back task ($r = .81$) and with the spatial cue factor from the ANT ($r = .50$) improved the model fit, ΔAIC = 16.4, $\Delta\chi^2(2) = 20.4$, $p < .001$. In general, there was however no consistent pattern of correlations between manipulation-specific factors that could have indicated shared variance between the different manipulations. In addition, merging the general N1 latency factors from the three different EF tasks in one task-general N1 factor did not impair model fit and represented a more parsimonious account for the data, ΔAIC = -1.3, $\Delta\chi^2(2) = 2.7$, $p = .261$. This task-general N1 factor captured all variance of the general N1 factor for the N-Back task, 87% of variance of the general N1 factor for the shifting task, and 47% of variance of the general N1 factor for the ANT. While this model still showed only a mediocre fit to the data, $\chi^2(536) = 953.5$, $p < .001$, CFI = .864, RMSEA = .09, 90% CI = [.08, .10], it still was retained for further analysis as it was the best fitting model.

The joint model for P3 latencies indicated correlations between the general P3 factors (r_s = .48 - 67) as well. Again exploratory analysis revealed that additionally estimating four correlations between manipulation-specific factors improved the model fit, ΔAIC = 19.6, $\Delta\chi^2(4)$ = 27.6, $p < .001$. Specifically, correlations between the global shifting factor and the flanker factor in the *N*-Back task (r = .73), between the flanker factor in the shifting task and the double cue factor from the ANT (r = -.45), between the flanker factor in the N-back task and the no cue factor from the ANT (r = -.59), and between the facilitation factor from the N-Back task and the double cue factor from the ANT (r = .53) were significantly different from zero. Still, there was no consistent pattern of correlations between manipulation-specific factors that would indicate a general factor of the different manipulations across EF tasks. Moreover, joining the three general P3 factors in one overarching factor for the P3 latency in the three EF tasks did not impair model fit and represented a more parsimonious account for the data, ΔAIC = -3.7, $\Delta\chi^2(2) = 0.3$, p = .845. In this model, the task-general P3 factor explained all variance in the general P3 factor for the shifting task, 42% of variance in the general P3 factor for the *N*-Back task, and 47% of variance of the general P3 factor for the ANT. Albeit, this model still had only a mediocre fit, $\chi^2(529) = 908.2$, $p < .001$, CFI = .867, RMSEA = .09, 90% CI = [.08, .09]. It still was retained for further analysis because it represented the best solution of all estimated models.

Altogether, the joint modeling of the three EF tasks indicated that general performance in the three EF tasks was consistently correlated and could be merged into one factor. Although there were some significant correlations between manipulation-specific factors, there were no consistent patterns within these correlations suggesting that individual differences with respect to specific manipulations were divergent rather than unitary.

Bi-variate models of EF tasks and covariates. Detailed results of the bi-variate models between EF tasks and the three covariates can be found in the analysis scripts available at: osf.io/6trne. There were no consistent correlations between the three covariates (i.e. WMC, PS, and intelligence) and manipulation-specific factors in the EF tasks. Interestingly, this was the case both for behavioral RTs and latencies of ERP component measures in the EF tasks. Thus, these correlations were not estimated in the joint model with all three covariates and the three EF tasks. In contrast, the general factors that represented similar individual differences across the three EF tasks showed considerable correlations and were thus included in the model joining all three covariates and the three EF tasks. Detailed results will be reported in the next section.

Joint modeling of all covariates and EFs. A combined model with all three covariates and behavioral RTs in the three EF tasks showed a good fit to the data, $\chi^2(940) = 1168.6$, $p < .001$, CFI = .97, RMSEA = .05, 90% CI = [.04, .06]. The path diagram of this model is shown in the top part (A) of Figure 7. Specifically, the factor merging behavioral performance in the three EF tasks showed a large positive correlation with processing speed in the ECTs ($r = .77$), and slightly lower and negative correlations with both *Gf* ($r = -.55$) and WMC ($r = -.49$). In addition, results indicated a very high correlation between *Gf* and WMC ($r = .95$), and medium correlations of *Gf* and WMC with PS in the ECTs ($r = -.46$ to $-.55$). Due to the strong association between PS in ECTs and the general performance factor of EFs, we simplified the model by estimating one general processing speed factor consisting of EFs and ECTs and one factor tentatively named *higher cognition* summarizing *Gf* and WMC (see bottom part of Figure 7 for the path diagram). These simplifications did not impair model fit, ΔAIC = -4.0, $\Delta\chi^2(7) = 10.0$, $p = .191$, and the model itself fit the data well, $\chi^2(947) = 1178.6$, $p < .001$, CFI = .97, RMSEA =

.05, 90% CI = [.04, .06]. This model indicated a medium correlation between the factor for *higher cognition* and general processing speed (r = -.54).

The SEM combining N1 latencies from the EF tasks with all three covariates showed only a mediocre fit to the data, $\chi^2(947) = 1926.6$, $p < .001$, CFI = .77, RMSEA = .10, 90% CI = [.09, .11]. All correlations between the task general N1 factor and covariates were low (r_s = -.16 to .06) and non-significant (all $p_s > .160$). In additions, assuming no covariance between the task-general N1 latency factor and the three covariates did not impair model fit, ΔAIC = -1.7, $\Delta\chi^2(3) = 4.3$, $p = .230$, further indicating that there was no correlation between the N1 latencies in the three EF tasks and any of the three covariates (see Figure 8A for a path-diagram of this model).

Finally, the joint model of P3 latencies in the EF tasks with all three covariates showed a mediocre fit to the data, $\chi^2(941) = 2366.2$, $p < .001$, CFI = .68, RMSEA = .12, 90% CI = [.11, .13]. Because the size of our sample did not allow for a more complex model to be estimated, we provisionally retained this model. It must be noted that no strong conclusions can be drawn due to unsatisfactory model fit. In detail, the task-general P3 factor showed significant correlations with WMC (r = -.38), and PS from ECTs (r = .45), while the correlation with intelligence (r = -.16) was non-significant (p = .179). Setting the correlation between intelligence and the task-general P3 factor to zero did not impair model fit, ΔAIC = -0.1, $\Delta\chi^2(1) = 1.9$, $p = .173$, but correlations with the other two covariates changed (see Figure 8B for a path diagram).

Discussion

The present study aimed to disentangle the relationship between individual differences in processing speed, working memory capacity and executive functions with general intelligence. Specifically, we were interested in two different points: (1) in how far performance in EF tasks represented general or manipulation-specific aspects. And (2) which of these two different aspects of performance in EF tasks was related to intelligence, working memory capacity (WMC), and processing speed (PS). Overall, performance in specific conditions within EF tasks seemed to largely capture general variance rather than variance specific to an experimental manipulation. Furthermore, manipulation-specific variance in EF tasks did not show any consistent relationships among the different EF tasks and with the three covariates, while general variance in behavioral RTs was related to all three covariates, and general variance in P3 latencies was related to WMC and PS.

Performance in EF tasks: What does it measure?

Before taking on the question in how far performance in EF tasks is related to WMC, PS and intelligence, we addressed the question what is measured by reaction times as performance measures in EF tasks. This is an important point as former studies investigating this relationship have used various indicators for individual differences in EFs. Some studies have used performance from specific conditions in an EF task that should require one specific executive function (e.g., RTs for incongruent conditions in a Stroop task; Wongupparaj et al., 2015) or average performance across conditions (e.g., mean proportion correct in updating tasks; Miyake et al., 2000), while differences between specific experimental conditions in EF tasks were used in other studies (e.g. the difference between RTs in switch versus repeat trials in a shifting task; Friedman et al., 2006). Interestingly, these different measures have often been mixed within studies (Friedman & Miyake, 2017; Miyake et al., 2000; Wongupparaj et al., 2015), and in the most cases difference measures have been used for inhibition and shifting tasks, while average performance across conditions has been used for updating tasks.

From a theoretical perspective, the use of either measure is not entirely unproblematic. Using performance from a single condition or average performance across conditions may confound different variance sources. In this case, such measures may contain variance specific to experimental manipulations that require specific executive functions and more general variance linked to processing speed or memory capacity (Frischkorn & Schubert, 2018). In contrast, difference measures assume that cognitive processes are additive and that experimental conditions vary in all but a single cognitive process (Donders, 1868, 1969). Yet, it is likely that cognitive processes are not additive and that inserting an additional cognitive process interacts with other cognitive processes required in the task (Alexander, Trengove, & van Leeuwen, 2015; Friston et al., 1996; Schubert et al., 2015).

Our results from the bi-factor models of the three EF tasks indicate that behavioral RTs and ERP latencies from one condition within any EF task represented mostly general performance. Specifically, the general factors summarizing the variance consistent across all manipulations did capture the largest proportion of variance in manifest indicators across all EF tasks (on average 68%), while each manipulation-specific factor captured considerably smaller variance proportions (on average 14%). Hence, behavioral or neural measures from a single condition or average performance across all conditions will mostly represent individual differences in general rather than manipulation-specific cognitive processes.

As manipulation-specific factors can be interpreted as latent difference scores, the small amount of variance of manipulation-specific factors might also explain why difference scores in experimental paradigms often show low reliabilities (Hedge, Powell, & Sumner, 2018). In detail, when calculating the difference between two correlated experimental conditions, the small amount of systematic variance in this difference (i.e. the variance of manipulation-specific factors) gets outweighed by unsystematic error variance that gets amplified when calculating the difference between highly correlated variables. One very recently proposed solution to this may be to account for trial-to-trial noise by adopting a hierarchical modeling approach (Rouder & Haaf, 2018). However, it remains to be seen whether this approach solves the reliability issues of difference measures, and provides an increment above the here used method of latent difference scores that are virtually error free.

In conclusion, researchers have to consider that the selection of a specific measure such as difference scores or performance in a single experimental condition may change both the interpretation of the measure and the relationship with covariates. Specifically, results from studies that have not used difference scores as indicators of EFs could also be interpreted as indication that general processing speed (when using RTs in shifting or inhibition tasks) or memory/processing capacity (when using accuracies in updating tasks) are related to general intelligence instead of individual differences specific to executive functions. Since difference measures are not unproblematic either (Friston et al., 1996), developing theoretically founded measures for executive functions is a critical step towards accurately assessing the relationship between EFs and individual differences in other cognitive processes such as intelligence, WMC, and PS.

Executive Functions: Still no bridge across the gap.

While theoretically and empirically executive functions are supposed to underlie working memory capacity and thus to be related to intelligence as well (Kane, Conway, Hambrick, &

Engle, 2007; Kane & Engle, 2003; Unsworth, 2010; Unsworth et al., 2014), the results of the present study indicated that variance specific to experimental manipulations is consistently unrelated to working memory capacity, processing speed, and general intelligence. Moreover, manipulation-specific factors did not show any consistent correlational pattern with each other, indicating that executive functions required by the different experimental manipulations are divergent rather than unitary. This is in line with recent results suggesting that individual differences in executive functions, specifically inhibition, may not be as unitary as suggested (Rey-Mermet, Gade, & Oberauer, 2018; Stahl et al., 2014). In sum, correlations between difference measures from executive functioning tasks both with each other and with external criteria seem to be small and largely inconsistent (Hedge et al., 2018; Rey-Mermet et al., 2018; Stahl et al., 2014), calling into question (1) whether individual differences in executive functions are unitary at all and (2) whether they underlie the relationship between WMC, PS and intelligence.

Instead, only variance consistent across experimental manipulations showed relationships with the three covariates. Specifically, behavioral RTs in EF tasks showed medium-sized negative relationships with working memory capacity and general intelligence ($r = -.49$ to $-.55$), and a large positive relationship with PS from ECTs ($r = .77$). P3 latencies showed medium correlations with processing speed ($r = .40$), a small negative correlation with WMC ($r = -.28$), but no relationship with intelligence. Finally, N1 latencies showed no significant relationships with any of the three covariates. While smaller relationships of ERP latencies are often observed due to their tendency for lower reliability (Cassidy, Robertson, & O’Connell, 2012), the present results do not suggest that this is a problem, because residual error variance captured only a small proportion of variance in manifest ERP latencies. Nevertheless, it is important to note that model fit was unsatisfactory for all structural equation models including ERP latencies, especially P3 latencies, and the covariates (i.e. intelligence, WMC, and PS). Thus, the results for ERP latencies need to be replicated with tasks better suited for the EEG and a larger sample. Until then they should be interpreted with caution.

Although relationships of ERP latencies with intelligence were smaller and not as consistent as suggested by results from a recent study (Schubert, Hagemann, & Frischkorn, 2017), the present results replicated the finding that the speed of higher-order cognitive processes occurring later in the stream of neural processing is related to cognitive processes. In detail, latencies of the N1 component showed no significant relationship with general intelligence, working memory capacity, and processing speed, while latencies of the P3 component were positively correlated with processing speed, and negatively correlated with WMC. These results are in line with the theoretical interpretation by Schubert et al. (2018) and further support the idea that specific neural or cognitive processes underlie the relationship between general intelligence and information processing (Kievit, Davis, Griffiths, Correia, & Henson, 2016; McVay & Kane, 2012; van Ravenzwaaij, Brown, & Wagenmakers, 2011).

Speed of higher-order information processing: Basis of general intelligence?

Although behavioral results indicated that processing speed measured in both EF tasks and ECTs is related to intelligence and WMC, the present results did not replicate the large relationship between latencies of later ERP components (P2, N2, and P3) with general intelligence (Schubert, Hagemann, & Frischkorn, 2017). In fact, the present results did not show any correlation of P3 latency with general intelligence. Nevertheless, there was a small correlation of P3 latency with working memory capacity (r = -.28) and a medium correlation with basic information processing speed (r = .40). Still, the non-significant correlation of P3 latency with general intelligence in the present study and the overall smaller correlations demand an explanation.

One important aspect is that the present study used considerably different tasks. On the one hand the present tasks were more complex resulting in longer reaction times and more errors. And on the other hand, the present tasks integrated multiple cognitive processes to be performed simultaneously. The study by Schubert et al. (2017) used elementary cognitive tasks such as the Posner Letter matching task or the Sternberg task, whereas the present study measured latencies of ERP components in more complex executive functioning tasks. In detail, ECTs are designed to systematically vary the demand on one specific cognitive process (Hick, 1952; Posner & Mitchell, 1967; Sternberg, 1969). Thus, individual differences in latencies of ERPs computed in ECTs may specifically capture individual differences in the speed of this specific process (e.g., memory retrieval) the ECTs are designed to tap.

In contrast, the tasks used in the present study required at least two different cognitive processes while processing the target stimulus. In all three tasks, a decision with respect to the target stimulus had to be made, while additionally irrelevant information had to be ignored (i.e., inhibition in the ANT, and in flanker blocks in the Shifting and N-Back task), the decision task to be conducted had to be determined (i.e., shifting), or the target stimulus had to be encoded in memory while outdated information was being removed (i.e., updating in the N-back task). It is likely that these different processes run at least partly in parallel. As soon as there is a latency jitter between this two processes this will smear out the ERP (Ouyang, Herzmann, Zhou, & Sommer, 2011) and render the two processes inseparable. Thus, individual differences in average ERP latencies may mix up individual differences with respect to the decision process and other processes such as inhibition, updating, and shifting.

Results from both behavioral and neural measures suggested that individual differences specific to executive functions were not related to each other and other cognitive processes. However, the variance stemming from these EF processes might still be reflected in latencies of ERP components in addition to variance due to individual differences of the decision process. This uncorrelated additional variance may have masked the strong relationship of purely decision-specific variance in the latency of later ERP components with general intelligence.

Processing speed, working memory capacity and general intelligence: The missing link

Neural speed of information processing showed small and inconsistent relationships with general intelligence, working memory capacity, and processing speed. Nevertheless, the present results replicate the negative relationship (r = -.54) of behavioral processing speed with both intelligence and WMC and the strong relationship between the latter processes (Ackerman et al., 2005; Kyllonen & Christal, 1990; Schmiedek et al., 2007; Schmitz & Wilhelm, 2016; Schubert et al., 2015). Moreover, results still showed that individuals with higher working memory capacity and faster behavioral processing speed have shorter latencies of the P3 component. Taken together, we observed a considerable overlap between individual differences in speed of information processing, working memory and general intelligence that requires a theoretical explanation.

Even though researchers have often argued for a causal relationship between basic cognitive processes such as processing speed or working memory and general intelligence, there may also be a confounding variable that affects all these different cognitive processes. For instance, a recent study using a psychopharmacological manipulation of processing speed with nicotine indicated that the speed of neural information processing might not causally underlie individual differences in general intelligence (Schubert, Hagemann, Frischkorn, & Herpertz, 2018). In detail, nicotine administration did increase neural as well as behavioral processing speed, while not showing any effect on performance in a matrix reasoning task. Hence, processing speed may not causally underlie individual differences in intelligence despite being correlated with intelligence.

Conclusion

Altogether, the present results further emphasize the important role of processing speed and working memory for individual differences in general intelligence. In contrast, executive functions did not underlie either individual differences in processing speed or working memory capacity and thus did not explain why these two domains are related to general intelligence. Nevertheless, it is reasonable to assume that individual differences in both processing speed and working memory capacity arise due to similar limitations in the cognitive system (Meiran & Shahar, 2018; Wilhelm & Oberauer, 2006). A promising approach to further investigate this idea might lie in joining theoretically grounded measures for behavioral indicators of processing speed and working memory capacity (e.g. cognitive models; Frischkorn & Schubert, 2018), with biological indicators of neural processing related to these two processes (c.f. Schubert, Nunez, Hagemann, & Vandekerckhove, 2018). More comprehensive insights on the basic cognitive processes underlying individual differences in general intelligence may thus be gained by associating structural and function architectural features of the brain related to intelligence (Hilger, Ekman, Fiebach, & Basten, 2017; Menon & Uddin, 2010) with cognitive process domains such as working memory capacity and processing speed. Ultimately, this may provide the integration of working memory capacity and processing speed as related predictors of intelligence that could not be reached by executive functions.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. D. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*(4), 567–589. https://doi.org/10/bjtjnz

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10/drqmhx

Alexander, D. M., Trengove, C., & van Leeuwen, C. (2015). Donders is dead: cortical traveling waves and the limits of mental chronometry in cognitive neuroscience. *Cognitive Processing*, *16*(4), 365–375. https://doi.org/10/f7zgts

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175–192. https://doi.org/10/bq7rsv

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*(1), 71–92. https://doi.org/10/d6hwsp

Bazana, P. G., & Stelmack, R. M. (2002). Intelligence and information processing during an auditory discrimination task with backward masking: an event-related potential analysis. *Journal of Personality and Social Psychology*, *83*(4), 998–1008.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10/dbj

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10/dbn

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*(4), 674–692. https://doi.org/10/cvxsd7

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). New York: Springer-Verlag. Retrieved from //www.springer.com/de/book/9780387953649

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*, *49*(5), 659–664. https://doi.org/10/gd3vr9

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. https://doi.org/10/f3zwhw

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. https://doi.org/10/dn58z8

Colom, R., Flores-Mendoza, C., Quiroga, M. Á., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences*, *39*(5), 1005–1014. https://doi.org/10/fvckcg

Colom, R., Jung, R. E., & Haier, R. J. (2007). General intelligence and memory span: Evidence for a common neuroanatomic framework. *Cognitive Neuropsychology*, *24*(8), 867–878. https://doi.org/10/bn8rqz

Colom, R., & Thompson, P. M. (2013). Understanding Human Intelligence by Imaging the Brain. In *The Wiley-Blackwell Handbook of Individual Differences* (pp. 330–352). Wiley-Blackwell. https://doi.org/10.1002/9781444343120.ch12

Conway, A. R. ., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. . (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10/frs9t8

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170. https://doi.org/10/gbvchk

Dang, C.-P., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2015). Do processing speed and short-term storage exhaust the relation between working memory capacity and intelligence? *Personality and Individual Differences*, *74*, 241–247. https://doi.org/10/f6xc38

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10/bqr2f2

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10/ff2cmm

Donders, F. C. (1868). Over de snelheid van psychische processen. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool (1868–1869)*, *2*, 92–120. Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2316352

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10/cjj95t

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331. https://doi.org/10/ddp3gg

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10/d23822

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10/f9sqwm

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10/bmb68s

Frischkorn, G. T., & 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/gd3vqn

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage*, *4*(2), 97–104. https://doi.org/10/drwjs5

Gläscher, J., Rudrauf, D., Colom, R., Paul, L. K., Tranel, D., Damasio, H., & Adolphs, R. (2010). Distributed neural system for general intelligence revealed by lesion mapping. *Proceedings of the National Academy of Sciences*, *107*(10), 4705–4709. https://doi.org/10/b9h5df

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10/gddfm4

Hick, W. E. (1952). On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10/fc97vw

Hilger, K., Ekman, M., Fiebach, C. J., & Basten, U. (2017). Efficient hubs in the intelligent brain: Nodal efficiency of hub regions in the salience network is associated with general intelligence. *Intelligence*, *60*, 10–25. https://doi.org/10/f9jg3q

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. https://doi.org/10/dbt

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzlestungen: 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: Form 4*. Göttingen: Hogrefe.

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

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell—Horn—Carroll model. *Journal of Experimental Psychology: General*, *145*(2), 220–245. https://doi.org/10/f79zsq

Jung, R. E., & Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences*, *30*(02), 135. https://doi.org/10/d8ddm9

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in Working Memory Capcacity as Variation in Executive Attention and Control. In A. R. A. Conway & M. Jarrold (Eds.), *Variation in Working Memory* (pp. 21–46). Oxford University Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10/fwhz92

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related onstructs: Comment on Ackerman, Beier and Boyle (2005). *Psychological Bulletin*, *131*, 66–71. https://doi.org/10/fgtzfp

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10/cntzrs

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*. https://doi.org/10/gd3vsx

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10/f7kdvc

Kievit, R. A., Davis, S. W., Griffiths, J., Correia, M. M., & Henson, R. N. (2016). A watershed model of individual differences in fluid intelligence. *Neuropsychologia*, *91*, 186–198. https://doi.org/10/f892mx

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. https://doi.org/10/gd3vr6

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric g: Unitary process or a number of independent processes?. *Intelligence*, *15*(4), 397–422. https://doi.org/10/bz2p63

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10/bxmdv4

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*(2), 571–585. https://doi.org/10/c7646z

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10/gdz7rc

McGarry-Roberts, P. A., Stelmack, R. M., & Campbell, K. B. (1992). Intelligence, reaction time, and event-related potentials. *Intelligence*, *16*(3), 289–313. https://doi.org/10/fcrtwf

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 196–204. https://doi.org/10/d2bmhv

McVay, J. C., & Kane, M. J. (2012). Drifting From Slow to “D’oh!”: Working Memory Capacity and Mind Wandering Predict Extreme Reaction Times and Executive Control Errors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *38*, 525–549. https://doi.org/10/c945qk

Meier, M. E., & Kane, M. J. (2013). Working memory capacity and Stroop interference: Global versus local indices of executive control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 748–759. https://doi.org/10/gd59t9

Meiran, N., & Shahar, N. (2018). Working memory involvement in reaction time and its contribution to fluid intelligence: An examination of individual differences in reaction-time distributions. *Intelligence*, *69*, 176–185. https://doi.org/10/gdxqtm

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*(5), 655–667. https://doi.org/10/fmz3ch

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10/fzdfvs

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10/bkksp2

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*(4), 621–640. https://doi.org/10/cp4w8m

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features: Automatic spatio-temporal EEG artifact detection. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10/cwhtpn

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10/bpqg2q

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799. https://doi.org/10/f8t2hx

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10/bm923b

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working Memory and Intelligence--Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/10/cfjxgw

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193. https://doi.org/10/fs4vfj

Ouyang, G., Herzmann, G., Zhou, C., & Sommer, W. (2011). Residue iteration decomposition (RIDE): A new method to separate ERP components on the basis of latency variability in single trials. *Psychophysiology*, *48*(12), 1631–1647. https://doi.org/10/bqvh3b

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10/hmf

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, *74*(5), 392–409. https://doi.org/10/bm5nqh

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42. https://doi.org/10/b486gb

Rao, K. V., & Baddeley, A. (2013). Raven's matrices and working memory: A dual-task approach. *The Quarterly Journal of Experimental Psychology*, *66*(10), 1881–1887. https://doi.org/10/gd59vd

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10/fjwm2f

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/fndcr6

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501–526. https://doi.org/10/gcx8pf

Rouder, J., & Haaf, J. M. (2018). A Psychometrics of Individual Differences in Experimental Tasks. https://doi.org/10/gfdbw2

Sauseng, P., Klimesch, W., Freunberger, R., Pecherstorfer, T., Hanslmayr, S., & Doppelmayr, M. (2006). Relevance of EEG alpha and theta oscillations during task switching. *Experimental Brain Research*, *170*(3), 295–301. https://doi.org/10/dm39wk

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating: When flanker meets the n-back. *Psychophysiology*, *52*(10), 1293–1304. https://doi.org/10/f7qn24

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/dwr688

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). https://doi.org/10/gd3vrp

Schubert, A.-L., Frischkorn, G. T., Hagemann, D., & Voss, A. (2016). Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence*, *4*(3), 7. https://doi.org/10/gd3vs3

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/gch83n

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. https://doi.org/10/gffjb9

Schubert, A.-L., Hagemann, D., Voss, A., & Bergmann, K. (2017). Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology*, *77*, 29–45. https://doi.org/10/f943hk

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. https://doi.org/10/f7mtv3

Schubert, A.-L., Nunez, M. D., Hagemann, D., & Vandekerckhove, J. (2018). Individual differences in cortical processing speed predict cognitive abilities: a model-based cognitive neuroscience account. *BioRxiv*, 374827. https://doi.org/10/gfhzdz

Schulter, G., & Neubauer, A. C. (2005). Zentralnervensystem und Persönlichkeit [Central nervous system and personality]. In J. Henning & P. Netter (Eds.), *Biopsychologische Grundlagen der Persönlichkeit* (pp. 35–190). Munich, Germany: Elsevier.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4–27. https://doi.org/10/cgvnhr

Shahabi, S. R., Abad, F. J., & Colom, R. (2014). Short-term storage is a stable predictor of fluid intelligence whereas working memory capacity and executive function are not: A comprehensive study with Iranian schoolchildren. *Intelligence*, *44*, 134–141. https://doi.org/10/f576d3

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/fn3gm2

Souza, A. S., & Vergauwe, E. (2018). Unravelling the intersections between consolidation, refreshing, and removal. *Annals of the New York Academy of Sciences*, *1424*(1), 5–7. https://doi.org/10/gdxvb7

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*(2), 850–886. https://doi.org/10/f62q5t

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–457. Retrieved from https://www.jstor.org/stable/27828738

Troche, S. J., Houlihan, M. E., Stelmack, R. M., & Rammsayer, T. H. (2009). Mental ability, P300, and mismatch negativity: Analysis of frequency and duration discrimination. *Intelligence*, *37*(4), 365–373. https://doi.org/10/dm4nz7

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*(1), 16–28. https://doi.org/10/bmxh4j

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10/f54shj

van Ravenzwaaij, D., Brown, S., & Wagenmakers, E.-J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition*, *119*(3), 381–393. https://doi.org/10/dwf8d6

Verleger, R. (1997). On the utility of the P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131–156. https://doi.org/10/fvv22s

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*, 433. https://doi.org/10/gd3vs7

Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*(1), 18–50. https://doi.org/10/dgsf8g

Wongupparaj, P., Kumari, V., & Morris, R. G. (2015). The relation between a multicomponent working memory and intelligence: The roles of central executive and short-term storage functions. *Intelligence*, *53*, 166–180. https://doi.org/10/gd3vsr

Tables

Table 1

Descriptive Statistics for the Executive Functioning tasks

Task	Block	Task Shifting	Flanker	M_{RT} (SD_{RT})	M_{Pc} (SD_{Pc})
Shifting	Control LM			583.23 (94.12)	0.93 (0.01)
	Control OE			679.28 (130.57)	0.80 (0.02)
	Shifting (SH)	Switch		1'383.33 (306.83)	0.95 (0.09)
		Repeat		1'181.51 (245.18)	0.97 (0.08)
	SH Flanker	Switch	congruent	1'176.39 (249.79)	0.98 (0.03)
			neutral	1'181.13 (246.60)	0.98 (0.03)
			incongruent	1'215.83 (251.07)	0.98 (0.03)
		Repeat	congruent	1'072.24 (234.69)	0.99 (0.02)
			neutral	1'080.16 (242.77)	0.99 (0.02)
			incongruent	1'132.30 (248.04)	0.99 (0.03)
	Block	**Match**	**Flanker**		
N-Back	2-Back	False		1'108.53 (186.25)	0.91 (0.10)
		True		911.32 (190.45)	0.95 (0.06)
	2-Back Flanker	False	no	1'059.88 (189.14)	0.95 (0.08)
			congruent	1'144.80 (202.73)	0.94 (0.09)
			neutral	1'106.65 (190.48)	0.93 (0.10)
			incongruent	1'133.33 (202.27)	0.94 (0.08)
		True	no	843.02 (169.66)	0.97 (0.04)
			congruent	912.05 (163.71)	0.97 (0.05)
			neutral	859.21 (163.17)	0.96 (0.05)
			incongruent	934.95 (165.75)	0.97 (0.05)
		Cue	**Flanker**		
ANT		no	congruent	703.76 (93.69)	1.00 (0.00)
			neutral	684.50 (85.34)	1.00 (0.00)
			incongruent	825.63 (103.92)	0.98 (0.05)
		central	congruent	669.78 (97.38)	1.00 (0.00)
			neutral	660.10 (90.57)	1.00 (0.00)
			incongruent	809.02 (102.68)	0.98 (0.04)
		double	congruent	653.44 (92.16)	1.00 (0.00)
			neutral	652.69 (90.99)	1.00 (0.00)
			incongruent	790.15 (105.19)	0.98 (0.04)
		spatial	congruent	593.61 (93.37)	1.00 (0.00)
			neutral	593.61 (88.29)	1.00 (0.00)
			incongruent	706.48 (113.98)	0.99 (0.03)

Note. RT = Reaction Time in ms; Pc = Proportion correct responses; LM = Less More; OE = Odd-Even

Table 2
Descriptive statistics for the ERP latencies in the Executive Functioning tasks

Task	**Block**	**Task Shifting**	**Flanker**	**N1 Lat**		**P3 Lat**	
				M	(SD)	*M*	(SD)
Shifting	Control LM			117.41	(24.27)	465.39	(89.93)
	Control OE			118.01	(23.88)	467.83	(91.32)
	Shifting (SH)	Switch		120.43	(22.39)	473.87	(92.62)
		Repeat		120.59	(21.54)	466.81	(96.82)
	SH Flanker	Switch	congruent	114.84	(22.81)	492.48	(88.66)
			neutral	114.16	(24.46)	497.88	(105.80)
			incongruent	116.44	(23.67)	489.48	(82.99)
		Repeat	congruent	112.72	(22.91)	487.65	(87.63)
			neutral	114.52	(23.57)	480.70	(96.00)
			incongruent	115.01	(22.13)	481.13	(85.33)
	Block	**Match**	**Flanker**				
N-Back	2-Back	False		126.10	(25.49)	505.05	(97.14)
		True		130.29	(25.88)	478.47	(101.20)
	2-Back Flanker	False	no	121.49	(28.32)	510.05	(97.96)
			congruent	122.86	(25.26)	512.43	(101.48)
			neutral	122.86	(28.14)	495.34	(91.48)
			incongruent	131.90	(27.12)	505.58	(109.58)
		True	no	121.50	(24.79)	508.64	(84.91)
			congruent	121.32	(26.31)	501.34	(85.81)
			neutral	122.84	(24.23)	464.29	(76.57)
			incongruent	131.78	(24.34)	469.60	(99.87)
		Cue	**Flanker**				
ANT		no	congruent	144.05	(30.91)	431.58	(52.40)
			neutral	148.33	(28.00)	440.05	(56.18)
			incongruent	150.61	(31.07)	448.16	(51.45)
		central	congruent	140.80	(28.54)	425.77	(62.13)
			neutral	142.43	(29.97)	419.55	(55.21)
			incongruent	149.54	(31.73)	436.70	(61.62)
		double	congruent	135.93	(24.54)	416.18	(60.74)
			neutral	137.48	(20.83)	408.56	(51.77)
			incongruent	136.59	(24.30)	425.86	(49.24)
		spatial	congruent	127.67	(17.61)	422.12	(69.08)
			neutral	131.81	(17.70)	418.43	(70.76)
			incongruent	130.47	(22.34)	431.10	(61.16)

Note. Lat = latency, LM = Less More, OE = Odd-Even;

Table 3
Descriptive Statistics for the elementary cognitive tasks

Task	Block	Match	M_{RT} (SD_{RT})		M_{Pc} (SD_{Pc})	
Posner	PI	False	673.90	(118.19)	0.99	(0.03)
		True	623.83	(105.12)	0.99	(0.02)
	NI	False	753.08	(141.04)	0.99	(0.02)
		True	693.93	(120.58)	0.98	(0.03)
Sternberg	S3	False	885.29	(193.78)	0.99	(0.02)
		True	882.70	(202.77)	0.98	(0.03)
	S5	False	1051.46	(291.60)	0.98	(0.04)
		True	1017.01	(250.26)	0.97	(0.05)

Note. RT = Reaction time in ms; Pc = Proprotion correct; PI = Physical Identity; NI = Name Identity; S3 = Set Size 3; S5 = Set Size 5.

Figures

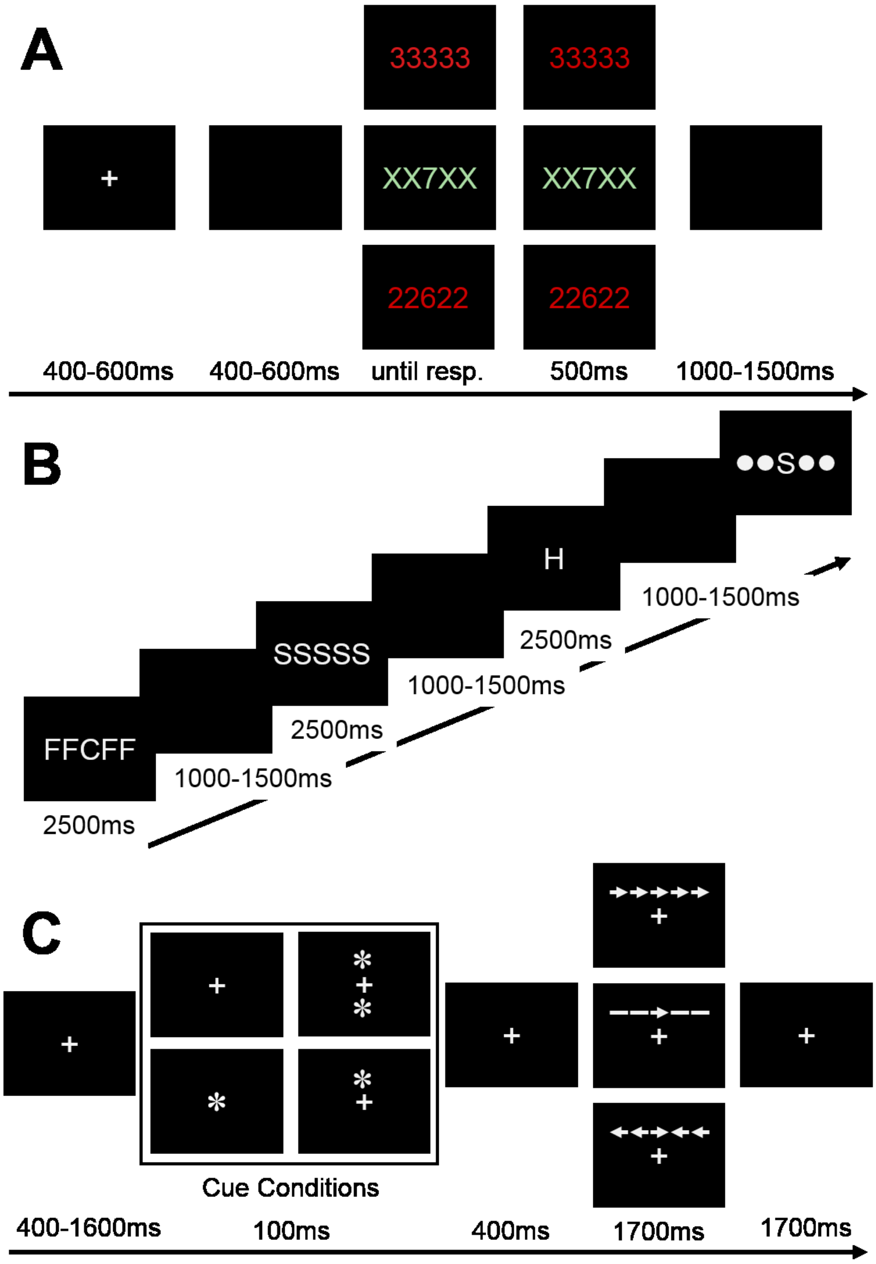

Figure 1. Trial Procedure of the three EF tasks for (A) Task Shifting with red (dark) as cue for less/more decision and green (light) for odd/even decision (Switching Task), (B) Updating (N-Back task), and (C) Inhibition (Attention Network Test, ANT). Presentation times are given below the different screens in the trial procedure. In the Shifting task and the N-Back task, the flanker stimuli as shown above were only presented in the flanker blocks. The other blocks in these two tasks did not include flanker stimuli and only showed the central target stimulus.

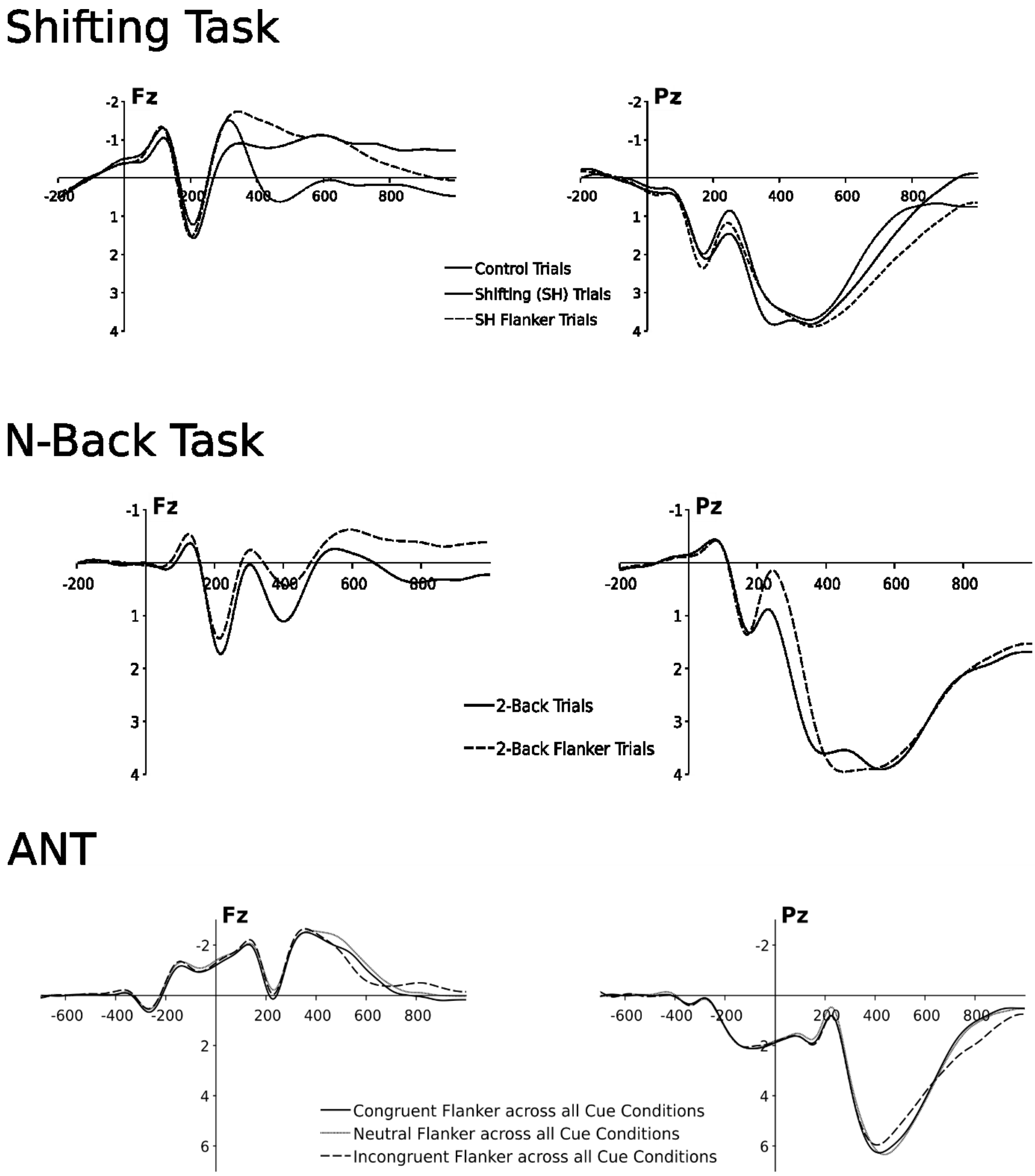

Figure 2. Grand Averages of the ERPs in the three executive function tasks across the experimental blocks. More detailed differences between specific experimental functions were omitted for readability. Time is displayed on the x-axis in milliseconds and the potential on the y-axis in µV. N1 latency was determined at Fz, and P3 latency at Pz.

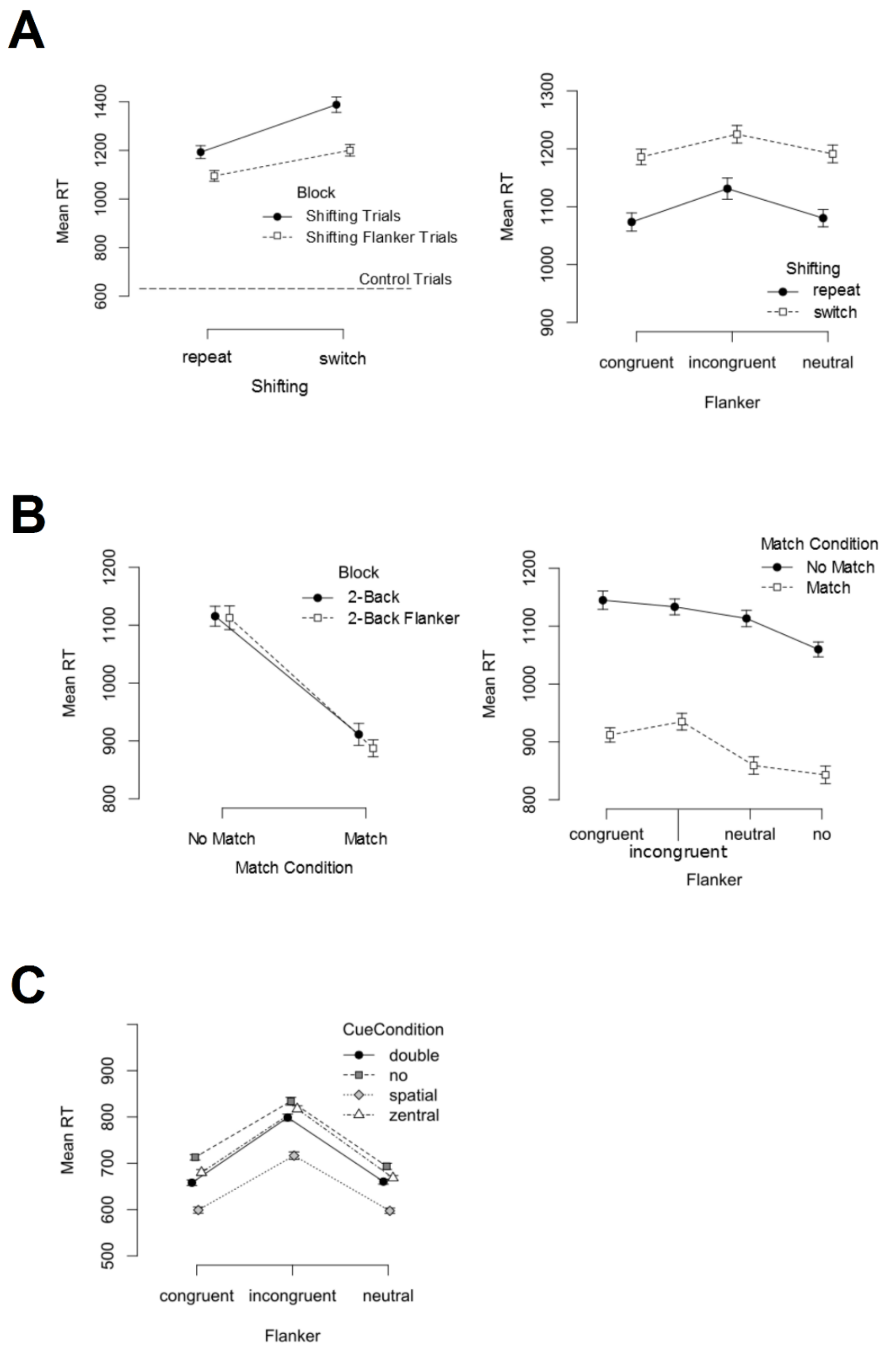

Figure 3. Descriptive Plots displaying the effects of experimental manipulations on behavioral RTs in the three EF tasks. The top panel (A) displays the effects in the Shifting task, the mid panel (B) the effects in the N-Back task, and the bottom panel (C) the effects in the ANT.

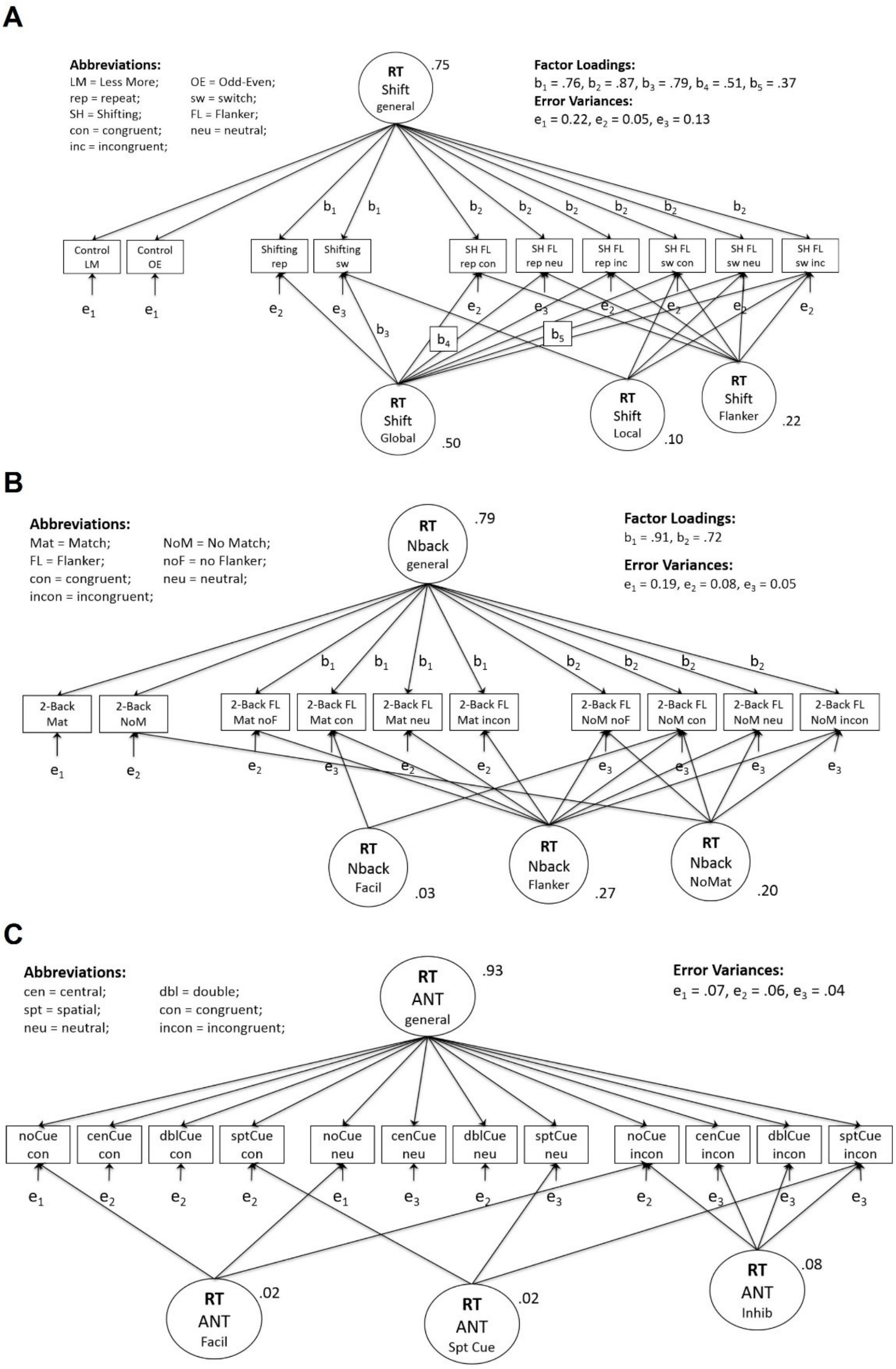

Figure 4. Path-diagrams of the Bi-factor models for behavioral RT in the three executive function tasks. The top part (A) shows the model for the shifting task (Shift), the middle part (B) shows the model for the N-Back task, and the lower part (C) the model for the Attention Network Test (ANT). All loadings that are not explicitly stated were fixed to one and unstandardized parameters are reported.

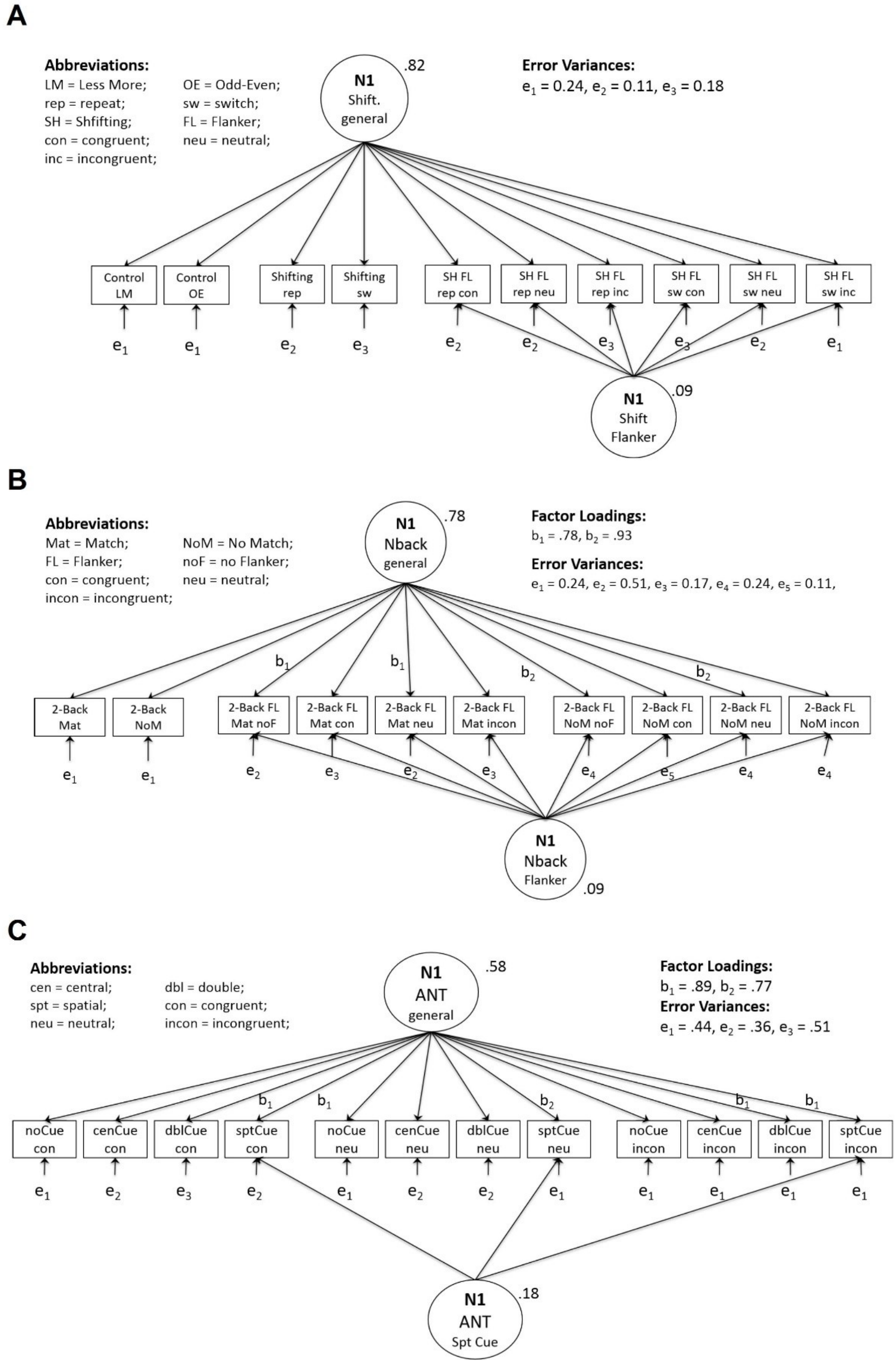

Figure 5. Path-diagrams of the Bi-factor models for N1 latency in the three executive function tasks. The top part (A) shows the model for the shifting task, the middle part (B) shows the model for the N-Back task, and the lower part (C) the model for the ANT. All loadings that are not explicitly stated were fixed to one and unstandardized parameters are reported.

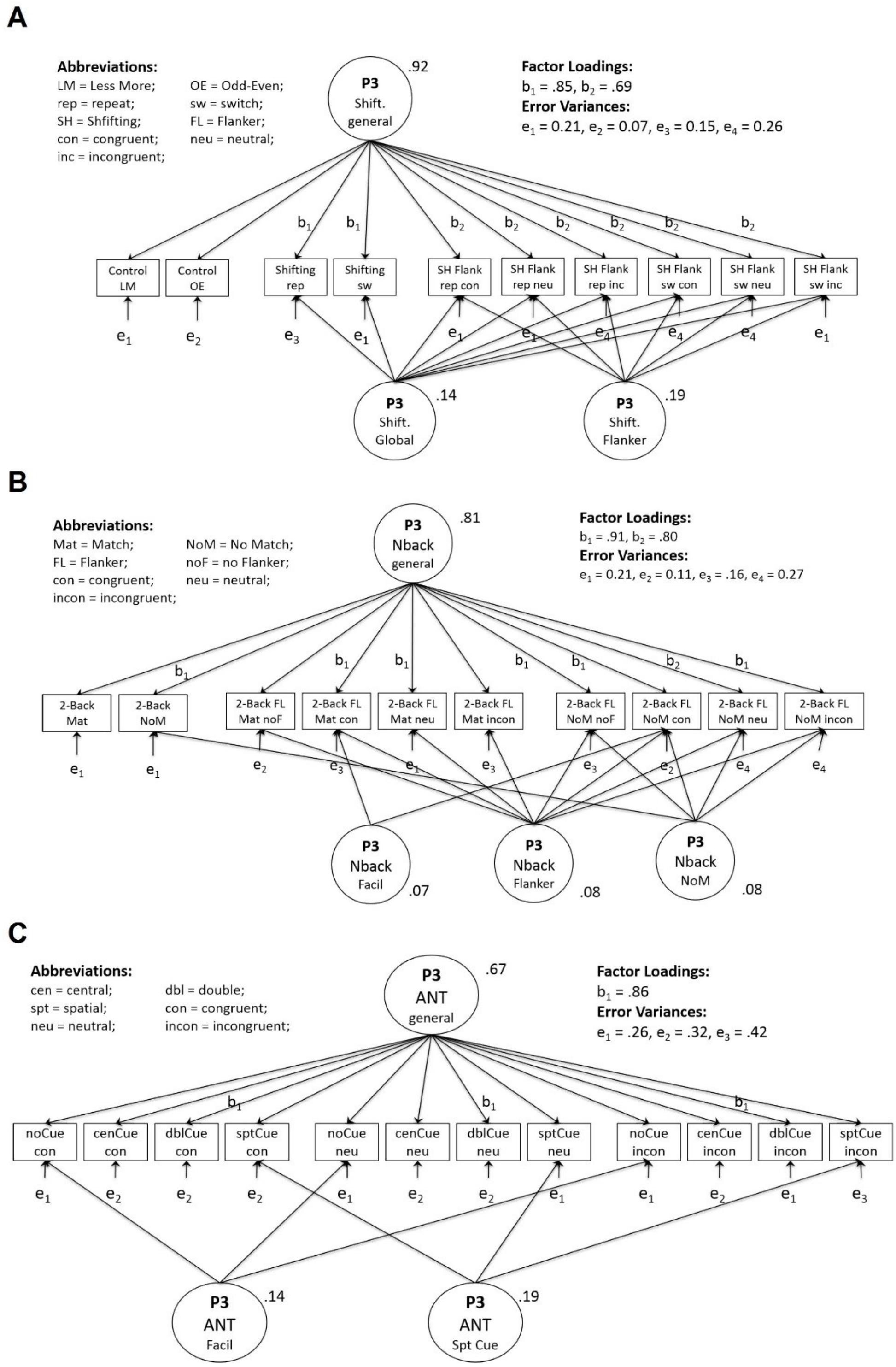

Figure 6. Path-diagrams of the Bi-factor models for P3 latency in the three executive function tasks. The top part (A) shows the model for the shifting task, the middle part (B) shows the model for the N-Back task, and the lower part (C) the model for the ANT. All loadings that are not explicitly stated were fixed to one and unstandardized parameters are reported.

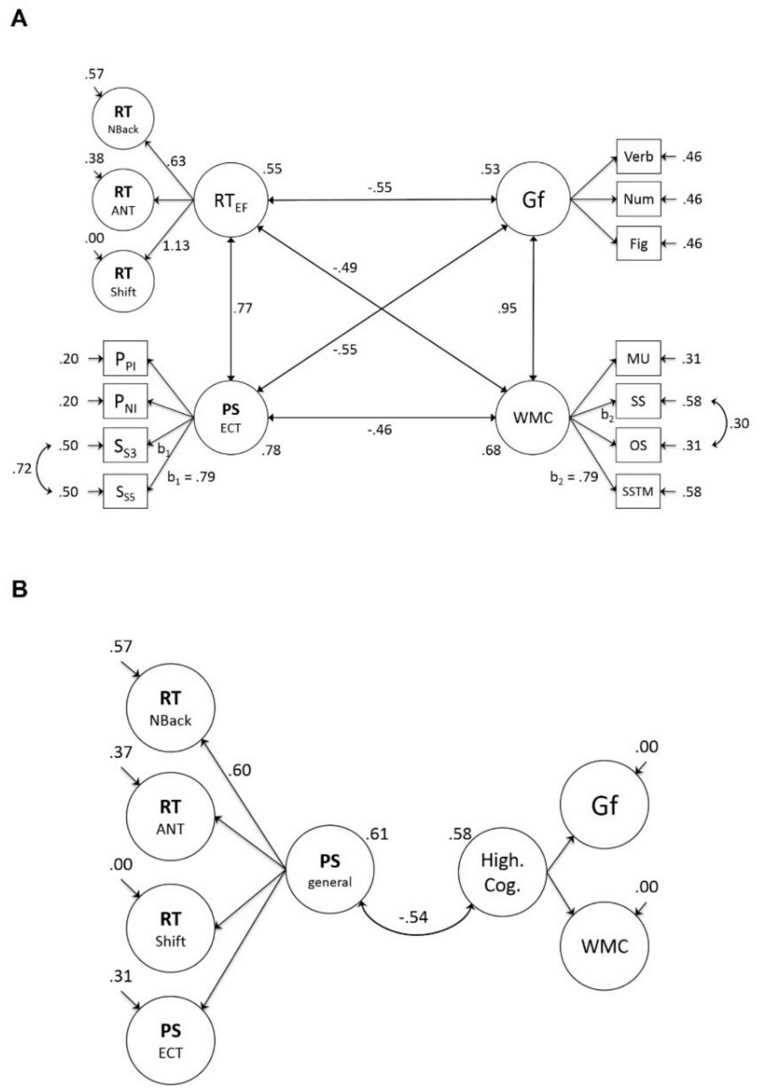

Figure 7. Path-diagrams for SEM of behavioral reaction times (RTs) in three EF tasks with the three covariates processing speed (PS), intelligence (Gf), and working memory capacity (WMC). Gf indicators are: verbal (Verb), numerical (Num), and figural (Fig) score from the BIS. Indicators for WMC are: proportion correct in a memory updating (MU), a sentence span (SS), an operation span (OS), and a spatial short-term memory (SSTM) task. Processing Speed indicators are: name identity (NI), and physical identity (PI) RTs from the Posner task (P), and set size 3 (S3) and set size 5 (S5) RTs from the Sternberg task (S). The top part (A) shows a correlational model, whereas the bottom part (B) shows a simplified model joining PS and EF performance into one general processing speed factor and Gf and WMC into a factor for higher cognitive abilities. Manipulation-specific factors of EF tasks are not depicted as all relationships of these factors with other factors were fixed to zero. All loadings that are not explicitly stated were fixed to one. Parameters are unstandardized except for correlations and differ all significantly from zero.

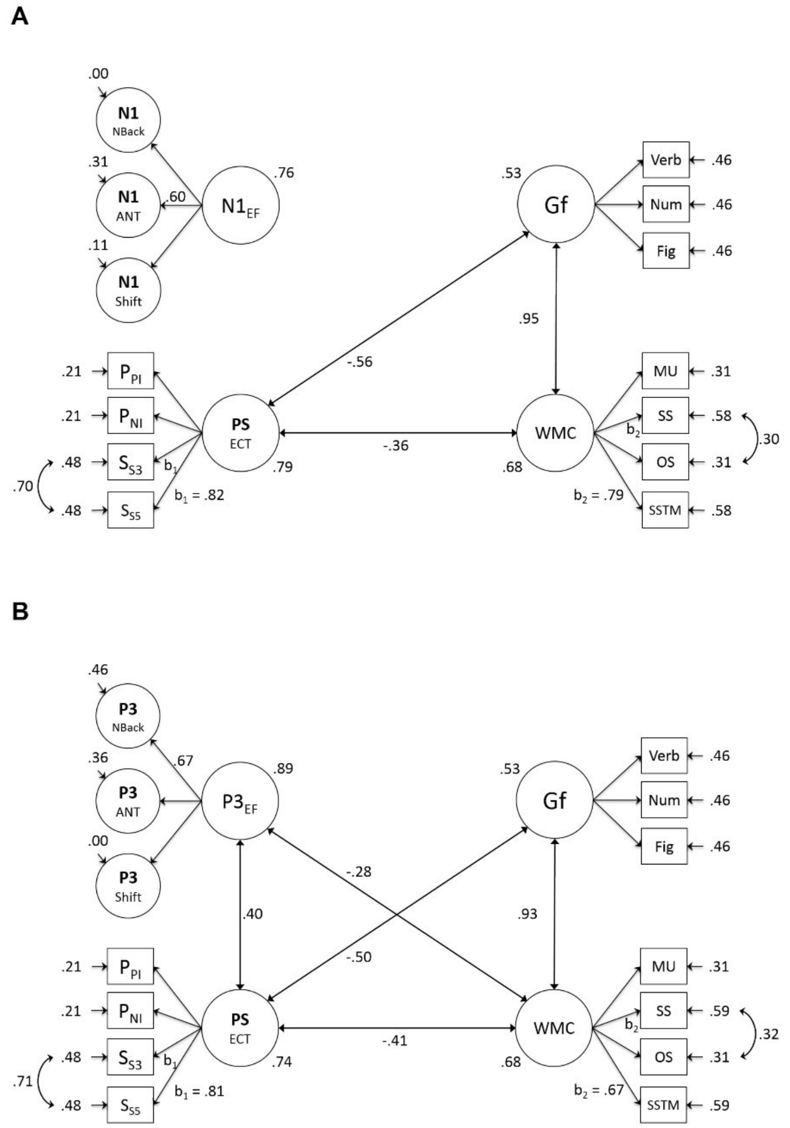

Figure 8. Path-diagrams for SEM of ERP latencies in three EF tasks with the three covariates processing speed (PS), intelligence (Gf), and working memory capacity (WMC). Gf indicators are: verbal (Verb), numerical (Num), and figural (Fig) score from the BIS. Indicators for WMC are: proportion correct in a memory updating (MU), a sentence span (SS), an operation span (OS), and a spatial short-term memory (SSTM) task. Processing Speed indicators are: name identity (NI), and physical identity (PI) RTs from the Posner task (P), and set size 3 (S3) and set size 5 (S5) RTs from the Sternberg task (S). The top part (A) shows the model for N1 latencies, and the bottom part (B) shows the model for P3 latencies. Manipulation-specific factors of EF tasks are not depicted as all relationships of these factors with other factors were fixed to zero. All loadings that are not explicitly stated were fixed to one. Parameters are unstandardized except for correlations and differ all significantly from zero.