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

The relationship between mental speed and mental abilities

presented by
Anna-Lena Schubert

year of submission
2016

Dean: Prof. Dr. Birgit Spinath
Advisor: Prof. Dr. Dirk Hagemann

Acknowledgements

This dissertation would not have been possible without the support of several admirable individuals who contributed ideas, conversations, and guidance.

First and foremost I want to thank my supervisor Prof. Dr. Dirk Hagemann, who has always offered constant and reliable help and has astonished me with his extraordinary commitment many times. His attitude of joviality, collegiality, and professionalism allowed me to grow both as a researcher and as an individual.

A special mention is warranted for Prof. Dr. Joachim Funke who kindly agreed to review my thesis.

Much of the work presented here is the fruit of numerous discussions and collaborations at the institute. In particular, I want to thank Prof. Dr. Andreas Voss for his valuable, clear-sighted and encouraging guidance. Moreover, I want to thank my roommates, colleagues, and friends, Dr. Katharina Bergmann and Gidon Frischkorn, for our continuous battle of wits, for their shared caffeine addiction, and for making our office a place that feels like home. I am further indebted to my colleagues Dr. Andrea Schankin, Benjamin Tauber, Andreas Neubauer, Veronika Lerche, Ulf Mertens, Carolin Baumann, Max Ihmels, Dr. Rosalux Falquez, and Dr. André Mata for their continuous support.

My gratitude also goes to Marianne Beschorner for her friendly support in administrative as well as personal matters. Moreover, I want to thank our team of research assistants – Ben Riemenschneider, Bjarne Schmalbach, Christoph Löffler, Eva Becker, Jana Villioth, Lynn Gärtner, Malin Hildebrandt, and Quynh Trinh Nguyen – for their kind and reliable assistance and their dedication to my dissertation project.

Finally, I want to express my sincerest gratitude to those not primarily involved in this dissertation. I want to thank my parents Annette and Klaus Schubert for their constant and unconditional support. Furthermore, I want to thank a few outstanding individuals whose friendship I have been fortunate to enjoy – Martin L., Sabine F., Laura M., Lea F., Kai D., Wolfram L., Nicolas F., Sven T, Erik Z., and Christoph S. Without them, these past three years would have been so much duller.

Table of contents

List of scientific publications of the publication-based dissertation

I. Manuscript
Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A., & Bergmann, K. (2015). Decomposing the Relationship between Mental Speed and Mental Abilities. *Intelligence, 51*, 28-46. doi: 10.1016/j.intell.2015.05.002

II. Manuscript
Schubert, A.-L., Hagemann, D., Bergmann, K., & Voss, A. (submitted). Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology*.

III. Manuscript
Schubert, A.-L., Hagemann, D, & Frischkorn, G. T. (submitted). General intelligence is little more than the speed of higher-order processing. *Nature*.

1. Introduction

General intelligence (*g*) - the common variance shared by different measures of cognitive ability - is a captivating psychological construct with a high predictive validity for educational attainment, job performance (Schmidt & Hunter, 2004), development of expertise (Wai, 2014), general health (Der, Batty & Deary, 2009), longevity (Deary, 2008), and well-being (Pesta, McDaniel, & Bertsch, 2010). Therefore, it is hardly surprising that many popular definitions of intelligence emphasize its role in achieving positive life outcomes through overcoming problems or changing the environment by reasoning. In their 1995 report, the Board of Scientific Affairs of the APA defined intelligence as "[the] ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77). Similarly, Robert Sternberg described intelligence as "a mental activity directed toward purposive adaptation to, selection, and shaping of real-world environments relevant to one's life" (Sternberg, 1985), and David Wechsler understood intelligence 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). Because general intelligence is such a powerful predictor of life outcomes, identifying which cognitive, neurophysiological, and genetic factors give rise to individual differences in intelligence – especially in *g* – is of great relevance to different areas of applied research.

The emergence of *g* in any cognitive test battery raises one of the greatest theoretical challenges in the identification of processes underlying individual differences in intelligence. Heterogeneous measures of cognitive abilities typically share 40-50 percent of their variance. Moreover, *g* factors from different cognitive test batteries are nearly perfectly correlated (Johnson, Nijenhuis, & Bouchard, 2008). Both the emergence of *g* and its great invariance across test batteries suggest that there may be some domain-wide process or property underlying individual differences in cognitive abilities (Spearman, 1923). Mental speed is one often proposed candidate property of information processing affecting different cognitive abilities that may underlie individual differences in general intelligence (Jensen, 2006).

2. Mental speed

In the 19th century, Frans C. Donders was the first researcher to conduct experimental research on the speed of information processing by systematically varying the complexity of response time tasks and analyzing the subsequent change in response times (RTs). He was particularly interested in inserting additional processing demands into simple response time tasks and measuring the subsequent increase in response times. This would allow him to identify the time required for the inserted mental process by subtracting the response times in the original paradigm from the response times in the more complex paradigm (Donders, 1868/1969). This *subtraction method* became very popular, because it would allow the measurement of the speed of specific cognitive processes if the assumption that additional processing demands can be inserted into a simple response time paradigm without affecting any other processes were true. Hence, even today many response time tasks are based on this assumption and allow the calculation of difference scores to identify the speed of specific processes.

Ever since, response times have become increasingly popular in psychology and the chronological measurement of cognitive processes has fueled much of the early physiological research (Cattell, 1886; Wundt, 1908-1911). The first studies on individual differences in RTs were conducted by Francis Galton, who collected behavioral and physiological data from more than 10,000 participants and (among many other enquiries) analyzed whether group differences in demographic variables predicted individual differences in RTs (Galton, 1908). Galton assumed that individual differences in mental abilities could be predicted by response times to external stimuli. However, the low reliability of response time measurements and the lack of more sophisticated statistical methods at the turn of the century prevented him and subsequent researchers such as Clark Wissler from finding any meaningful associations between RTs and other variables (Galton, 1908; Spearman, 1904; Wissler, 1901). Nevertheless, two later studies in the first half of the 20th century already reported small correlations between RTs and intelligence tests (Peak & Boring, 1926; Roth, 1964).

More recent research on mental speed has overcome these problems by using standardized response time devices, computerized chronometric measurements, and higher trial numbers to increase the reliability of measurements. Typical response time tasks are very simple and have only marginal cognitive requirements, so that individual differences in strategy use have little or no impact on response times. In a recent review of 172 studies published between 1955 and 2005, Sheppard and Vernon (2008) reported an average correlation of r = -.24 between different measures of mental speed and mental abilities. Although the moderate size of this correlation does not warrant any claim that mental speed may be *the sole* cognitive basis of general intelligence, it indicates that more intelligent individuals tend to have a consistently higher speed of information processing.

Despite the large number of studies on the relationship between mental abilities and mental speed, only few attempts have been made to systematically define mental speed. Most studies use the operational definition (if any) that any measurement of response times measures mental speed. In a recent call for papers for an upcoming special issue on "Mental Speed and Response Times in Cognitive Tests" in the Journal of Intelligence, Oliver Wilhelm described mental speed "[…] as a construct label" that "[…] is also known under the labels 'processing speed', 'elementary cognitive speed', 'clerical speed', etc.". He added that mental speed "[…] was also used as a tool to express mental work, for example time required per correct response or number of correct responses per time unit" (Wilhelm, 2015). This very broad definition of mental speed subsumes many different chronometric measures from elementary cognitive speed (response times in extremely easy response time paradigms) over clerical speed (the speed in which someone can perform repetitive tasks typically encountered in clerical occupations, e.g. typing speed or the speed at which papers can be filed) to speeded testing (the number of correct responses per time unit in a test of any complexity).

A narrower definition was proposed by Arthur Jensen who defined mental speed "[as] the actual time taken to process information of different types and degrees of complexity" (Jensen, 2006, p. ix). For the purpose of this doctoral thesis, I will extend this definition by adding that mental speed is "*the actual time taken to process information of different types and degrees of complexity" (Jensen, 2006, p. ix) goal-directedly*. I decided to extend Jensen's definition, because in most studies analyzing the relationship between mental abilities and mental speed, mental speed is measured as the speed of goal-oriented information processing culminating in some kind of decision. Hence, this definition emphasizes that mental speed constitutes not only of the speed of information encoding and retrieval from short- and long-term memory systems, but also of the speed of goal-directed information processing and decision making.

The small number of existing definitions of mental speed is only one symptom of a serious theoretical issue: Because the one thing all studies on mental speed have in common is that response times are used as variables in some multivariate context, there has never been any conceptual discussion about what the construct actually encompasses. If mental speed was simply used as a synonym for response times in the context of individual differences, the scope of the construct would be incredibly limited. Even worse, mental speed could never be located in a nomological network including other psychological constructs, because it would lack correspondence rules between the theoretical construct and its measurement as the two would be identical (Cronbach & Meehl, 1955). As elaborated by McCorquodale and Meehl

(1948), a theoretical construct has to contain a surplus meaning going beyond its measurement and involving processes that are not directly observable.

In order to overcome this conceptual problem, establish a theoretical conceptualization of mental speed, and allow its localization in a nomological network, I suggest three steps. In a first step, the measurement of mental speed has to be expanded beyond the measurement of response times using state-of-the-art psychological and physiological methods, allowing statements and inferences about mental speed without referring to response times as its only empirical operationalization. In a second step, an internal nomological network of mental speed has to be established, in which the relationships between different operationalizations of mental speed and between mental speed measured in different paradigms can be located. In a third step, mental speed has to be located in a larger nomological network including other psychological constructs such as intelligence and executive functions, and theoretical and empirical linkages between the constructs have to be specified.

2.1 The measurement of mental speed

In individual differences research, mental speed is typically measured as response times in so-called elementary cognitive tasks (ECTs). These ECTs are tasks with very low cognitive demands that maximize the empirical control of task complexity and minimize unwanted sources of variance in individual differences. Many of them consist of several conditions with increasing complexity that follow the logic of Donders' subtraction method and thus allow calculating difference scores between conditions to identify the speed of specific cognitive processes.

One very popular example of such ECTs is the Hick paradigm, which is a single and choice response time task with stimuli and corresponding buttons arranged in a semi-circle. Because the information processing demands in this paradigm increase linearly with the logarithm of choice alternatives (Hick, 1952), individual response times can be regressed on them and the resulting individual slope parameters can be used as estimates for the individual "rate of gain of information" (Roth, 1964). Similarly, increases in response times due to increases in short-term or long-term memory demands in ECTs can be used to estimate the speed of short- or long-term memory access by calculating difference or slope parameters across conditions (Hunt, 1983; Sternberg, 1969).

If difference and slope parameters in ECTs allowed measuring the speed of specific cognitive processes, correlations between these parameters and mental abilities would allow disentangling the unspecific response time-intelligence association into different associations between the speed of specific cognitive processes and mental abilities. Moreover, these correlations should be free from the contaminating variance of motor responses, as response demands are equal across conditions and thus reflected in the intercept, not in the slope of the regression. Thus, if the slope was a linear function of the time taken to process *i* units of information, and if more intelligent individuals had a higher mental speed, their resulting slopes should be shallower than the slopes of less intelligent individuals as they would be able to process more units of information per time, as illustrated in Figure 1.

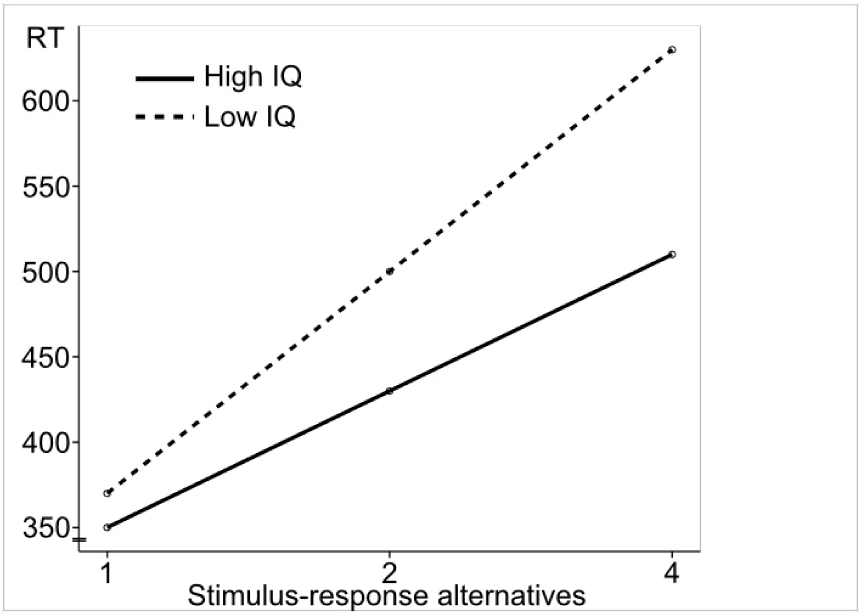

Figure 1. Schematic regression of response times on the logarithm of stimulus-response alternatives for a low and a high IQ group.

There has been no empirical evidence supporting this theoretical assumption, which warrants some skepticism of the idea that difference and slope parameters reflect specific cognitive processes. Although task complexities of ECTs are low, increasing task demands across conditions may not only affect the amount of information that has to be processed in some cognitive system such as short- or long-term memory, but may also affect task demands on other cognitive processes such as attention, encoding, working memory, or response preparation. If this were the case, it may be argued that difference and slope parameters are not indicators for the time required for a specific cognitive process, but for a conglomerate of processes differing between conditions. Therefore, more sophisticated methods that allow identifying the information processing components that differ between ECT conditions are needed to evaluate the validity of this assumption.

2.2 Using diffusion model and event-related potentials to evaluate the validity of elementary cognitive tasks (Manuscript 1)

In order to identify which information processing components differ between ECT conditions, the stream of information processing during each of these conditions has to be decomposed into different processes or process-related parameters. Both the mathematical modeling of response times and the simultaneous measurement of the electroencephalogram (EEG) during response time tasks allow this decomposition into process-related parameters and a comparison of process-specific demands across conditions.

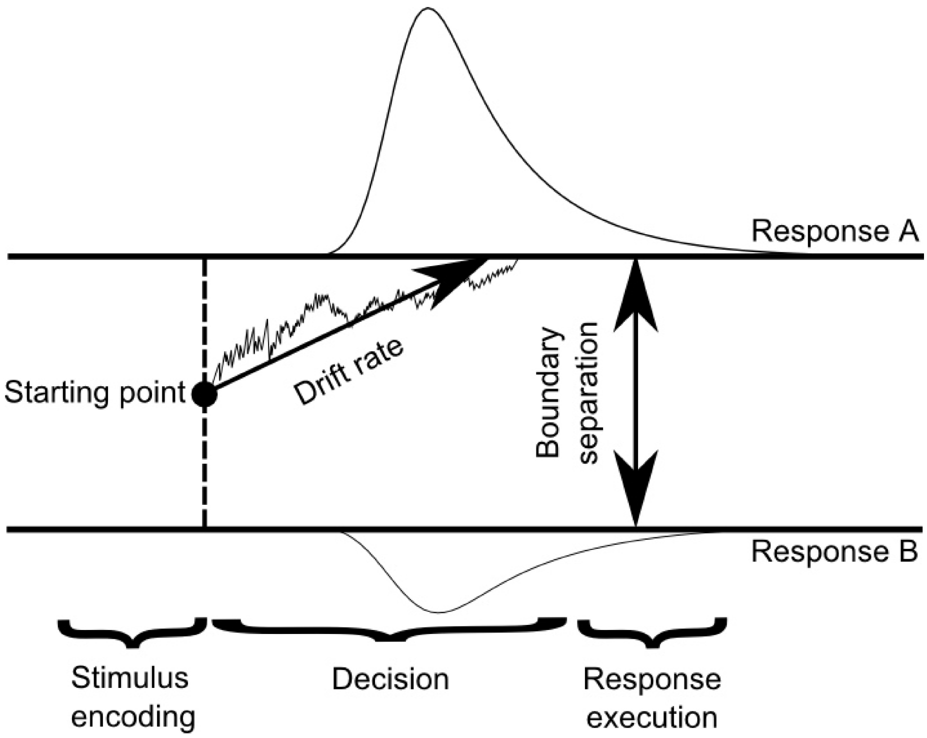

Figure 2. Information accumulation begins at the starting point and continues with a mean drift rate *v* (affected by random noise) until one of two thresholds is hit. Boundary separation depicts the response caution. Outside of the information accumulation process, non-decision time t_0 quantifies the time for stimulus encoding and response execution.

One very popular mathematical model of binary response time tasks is the diffusion model (Ratcliff, 1978), which estimates different cognitive parameters by fitting response time-distributions predicted from these parameters to empirical response time-distributions. The basic model is depicted in Figure 2. As is described in more detail in manuscript 1, p. 3, and manuscript 2, p. 5, the diffusion model is a process model of decision making with the underlying assumption that information is accumulated continuously until the diffusion process, which is a stochastic random-walk process, terminates after reaching one of two thresholds. This information accumulation process is driven by a constant systematic component, the *drift*, and the strength of this systematic component is called the *drift rate* (v), which is one of four parameters estimated in the basic diffusion model. The other parameter suited to identify which cognitive processes differ between ECT conditions is the non-decision time (t_0), which is the time taken for cognitive processes unrelated to decision making, such as encoding, memory retrieval, and motor reaction time. If ECT conditions differed in only a single cognitive process, condition differences should be reflected in changes in *only one* of the parameters. If both drift rate and non-decision time changed across conditions, demands would increase for different cognitive processes simultaneously, and difference and slope parameters would not validly isolate the speed of specific cognitive processes.

The same logic applies to the decomposition of the stream of electrophysiological information processing by means of the event-related potentials (ERP), which allows the identification of functionally distinct electrophysiological components (e.g., the N100 or P300) between stimulus onset and response. Each ERP component used here is defined by its polarity in comparison to the pre-stimulus baseline activity, its latency (here after stimulus onset), and its topography. Moreover, a large body of experimental research sheds light on which experimental manipulations elicit and affect specific ERP components, allowing a functional interpretation of ERP components. Again, if the assumption that ECT conditions differ in only a single cognitive process were true, condition differences should be reflected in changes in *only one* ERP component.

To test the validity of this assumption, 40 participants between 18 and 75 years from different educational and occupational backgrounds performed three ECTs – the Hick task, the Sternberg memory scanning task, and the Posner letter matching task (for a detailed description of these tasks see Manuscript 1, p. 2 and pp. 4-6) – while their EEG was recorded. In each of these tasks it is assumed that conditions differ only in the addition of a specific cognitive process. In the Sternberg memory scanning task, for example, participants were presented a memory set consisting of one, three, or five numbers and had to decide whether a subsequently presented probe stimulus was part of the memory set. Because the only demands supposed to increase across conditions are the demands on short-term memory search and retrieval, the slope of the regression of response times on memory set size is supposed to reflect the speed of short-term memory access (Sternberg, 1969).

Response times increased with increasing information processing-demands across task conditions, $\omega^2 = .64 - .71$, which is consistent with previous research (Hick, 1952; Sternberg, 1969; Posner & Mitchell, 1967). As predicted by the subtraction assumption of ECTs, in the Hick task condition differences were only reflected by changes in the non-decision component (probably reflecting increasing encoding or motor response demands), $\omega^2 = .60$, but not by changes in the drift rate, $\omega^2 = .03$. In the Posner letter matching task condition differences were only reflected by changes in the drift rate (reflecting the increasing difficulty of information accumulation), $\omega^2 = .48$, but not in the non-decision component, $\omega^2 = .04$. In the Sternberg memory scanning task, however, we observed both decreasing drift rate, $\omega^2 = .91$, and increasing non-decision component parameters, $\omega^2 = .48$, with increasing memory load, which should not occur under the subtraction assumption. This finding suggests that both encoding as well as information accumulation demands increase as a function of task demands in the Sternberg memory scanning task.

Moreover, the ERP analysis provided evidence against this assumption for all three ECTs, because increasing task demands were reflected in amplitude changes in several ERP components in each of the three tasks. In the Hick task, we found that P200 amplitudes were greater in the 1 bit than in the 2 bit condition, $\omega^2 = .41$, and that N200 amplitudes were greater in the 2 bit than in the 1 bit condition, $\omega^2 = .42$. These effects were greatest at central and central parietal electrode sites. We observed no main effect of condition on P300 amplitudes, $\omega^2 = .00$, but a two-way interaction between condition and caudality, $\omega^2 = .25$, indicated that P300 amplitudes were greater in the 1 bit than in the 2 bit condition at central electrode sites, $\omega^2 = .09$, and tended to be smaller at frontal electrode sites, $\omega^2 = .07$. These three ERP components reflect different specific cognitive processes. While not much is known about the cognitive correlates of the posterior P200 (Luck, 2005), the posterior N200 component has been associated attention (Folstein & Van Petten, 2008), and the P300 component has been associated with context-updating (Donchin, 1981; Polich, 2007) and context-closure (Verleger, 1988). Hence, condition differences in the Hick paradigm differ at least both in attentional and in updating demands.

In the Sternberg memory scanning task, we observed that N300 amplitudes increased, $\omega^2 = .40$, and that P300 amplitudes decreased, $\omega^2 = .36$, with increasing memory set size. No effect of condition was found for earlier ERP components such as the N150 and P200, all ω^2s $< .06$. The N300 component has been associated with spatial, structural, and categorical incongruences of visual stimuli (Demiral, Malcolm, & Henderson, 2012; Hamm, Johnson, & Kirk, 2002).

In the Posner letter matching task, N140 and N300 amplitudes were greater in the name identity than in the physical identity condition at frontal electrode sites, $\omega^2 = .05 – .13$. In addition, P210 amplitudes were greater in the physical identity than in the name identity condition at frontal electrode sites, $\omega^2 = .16$, and P300 amplitudes were greater in the physical identity than in the name identity condition at central electrode sites, $\omega^2 = .05$. The N150 has been discussed in the context of early lexical encoding (Spironelli & Angrilli, 2009), while the anterior P200 has been associated with short-term memory processing (Dunn, Dunn, Languis, & Andrews, 1998) and the detection of simple target stimuli (Luck & Hillyard, 1994). Thus, the physical and name identity conditions differ in the amplitude of at least four ERP components associated with distinct cognitive processes.

Parts of our electrophysiological results are consistent with previous research in which condition differences in several ERP components (P200, P390) were reported for a 2- and 4-choice response time task (Falkenstein, Hohnsbein, and Hoormann, 1994), and with studies in which increasing the memory set size in the Sternberg memory scanning paradigm also led to decreasing P300 amplitudes (Brookhuis, Mulder, Mulder, & Gloerich, 1983; Ford, Roth, Mohs, Hopkins, & Kopell, 1979; Gomer, Spicuzza, & O'Donnell, 1976; Houlihan, Stelmack, & Campbell, 1998; Pelosi et al.,1992). However, our analyses went beyond previous studies because we parameterized the whole stream of information processing with several ERP components, whereas most previous studies focused only on one or two components (often the P300).

Taken together, these results indicate that ECT conditions differ in several neuro-cognitive parameters and that slope and difference parameters are no valid estimates for the speed of a specific cognitive process. On the one hand this is worrying, because measuring the speed of specific cognitive processes is necessary to assess the relationship between different kinds of processing speed. Moreover, variance in behavioral response times consists of individual differences in mental speed and of individual differences in motor speed, which could be easily separated by means of slope and intercept parameters if the subtraction method was valid. On the other hand, more sophisticated methods such as mathematical modeling and event-related potentials allowing the decomposition of cognitive processes involved in response time tasks are readily available. These methods provide estimates for the speed of these specific processes so that the relationship between them can be assessed. Finally, using the latency of ERP components as a measure of mental speed provide an alternative operationalization of mental speed that expands its measurement beyond response times.

Both diffusion models and the EEG provide exciting new opportunities to identify the speed of cognitive processes. The EEG has seen decades of use in research on individual differences and has been successfully applied in the context of dispositional mood and personality (e.g., Fink, Grabner, Neuper, & Neubauer, 2005; Hagemann et al. 1999), dispositional approach/avoidance behavior (e.g., Hewig, Hagemann, Seifert, Naumann, & Bartussek, 2006), mental abilities (e.g. Grabner, Fink, Stipacek, Neuper, & Neubauer, 2004; McGarry-Roberts, Stelmack, & Campbell, 1992), and creativity (Fink, Graif, & Neubauer, 2009). In comparison, the diffusion model has only recently been applied in individual differences research.

3. Individual differences in diffusion model parameters

The diffusion model has only recently seen a huge rise in popularity, because in the last years several software solutions have been published that allow fitting the diffusion model and estimating model parameters without extensive programming knowledge (Voss & Voss, 2007; Vandekerckhove & Tuerlinckx, 2007, 2008; Wagenmakers, van der Maas, & Grasman, 2007). Therefore, there have not yet been many applications of the diffusion model in research on individual differences, although it has already been applied to gain a better understanding of individual differences in attention (Nunez, Srinivasan, & Vandekerckhove, 2015), in impulsivity (e.g., Stahl et al., 2014), in mental abilities (e.g., Ratcliff, Thapar & McKoon, 2010, 2011; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007), in numeracy (Ratcliff, Thompson, and McKoon, 2015), and in word recognition (Yap, Balota, Sibley, & Ratcliff, 2012). These first findings in these studies are very promising and encourage using the diffusion model in individual differences research.

3.1 The relationship between drift rate and intelligence (Manuscript 1)

The relationship between drift rate and general intelligence is of great theoretical interest, because the drift rate – i.e., the strength and direction of the systematic influence on the diffusion process – is often seen as the ability parameter of the diffusion model, whereas the other model parameters describe decision preferences and decision cautiousness (Vandekerckhove, 2015). Hence, it allows estimating the speed of information accumulation without the contaminating variance of sensory processing or motor response speed, both of which are captured in the non-decision time parameter.

Schmiedek et al. (2007) reported a correlation of $r = .79$ between a latent reasoning ability factor and a latent drift rate factor from eight response time tasks (including verbal, numerical, and spatial tasks) in a student sample. Ratcliff et al. (2010) analyzed the relationship between the vocabulary and matrix reasoning subtests of the Wechsler intelligence test and a latent drift rate factor from a numerosity discrimination, recognition memory, and lexical decision response time task for three age groups (college age, 60-74 years old, 75-90 years old). Correlations ranged from $r = .60$ to .90 for the vocabulary subtest, and from $r = .36$ to .85 for the matrix reasoning subtest. In a further analysis of response time data from an item and associative recognition task in this sample, Ratcliff et al. (2011) reported manifest correlations ranging from $r = .18$ to .67 between the matrix reasoning subtest and drift rates, and correlations ranging from $r = .28$ to .68 between the verbal subtest and drift rates. We also observed a correlation of $r = .50$ between a drift rate factor from three response time tasks (Hick task, Sternberg memory scanning task, Posner letter matching task) and intelligence (Manuscript 1). Taken together, these first results are very promising and encourage using the diffusion model to study the relationship between mental speed and mental abilities.

3.2 The role of model fit in experimental vs. multivariate research questions

Nevertheless, the use of diffusion models in individual differences research is still at its early stages. The majority of studies in which the diffusion model is applied are concerned with questions of model comparison in experimental paradigms: Just as we did in Manuscript 1, diffusion model parameters can be compared across experimental conditions in order to identify which cognitive processes differ between conditions, which is extremely useful to infer the cognitive processes responsible for changes in response time and accuracies resulting from an experimental manipulation. Not only can the estimated parameters be compared across conditions and tested for significance, but the model fit of a diffusion model with all parameters fixed across conditions can be compared to the model fit of a model with some parameters allowed to vary between conditions. If the model fit of the second model (e.g., with two separate drift rates estimated for the two conditions) is better than the model fit of the first one (e.g., with a common drift rate estimated for the two conditions), it can be inferred that the experimental manipulation in this paradigm affects a specific part of the diffusion process (e.g., the ease of information accumulation).

Model fit indices are used to decide which of several alternative models describes the empirical data best. They typically weigh the discrepancy between the empirical and predicted data against the complexity of the model, rewarding the parsimony of a model by adding a penalty for the number of model parameters. One popular fit index is the Akaike Information Criterion (AIC; Akaike, 1973), which is based on information theory and asymptotically estimates how much information is lost by one model in comparison to another. Another popular fit index is the Bayes Information Criterion (BIC; Schwarz, 1978), which is despite its name not derived from information theory, but from Bayesian statistics, and rewards parsimonious models more strongly than the AIC. Both the AIC und the BIC, however, do not provide any information about the absolute model fit, i.e. they only help to decide which of two or more models described the empirical data *best*, but not if any of these models described the data *well*.

As long as one is interested in questions of model comparison an evaluation of absolute model fit is not necessarily needed, because the question which model accounts for the empirical data best is the very question one is interested in. When model parameters are instead to be used in further analyses (e.g., in correlational analyses in individual differences research), it is important to know that individual model parameter values are meaningful in the sense that the model provides a valid description of the empirical data. Otherwise, correlations between model parameters and other ability measures may be underestimated if there is a high degree of noise in the model parameter values, which is likely if some participants' model parameters have no predictive validity for their observed response time distributions. Therefore, these cases have to be identified prior to further analyses by some kind of model fit evaluation criterion with cut-off values for acceptable absolute model fit.

3.3 Statistical tests of model fit

The most straight-forward evaluation of model fit is a statistical test of model significance. Model fit is acceptable if the null hypothesis that the empirical and the predicted response time distributions do not diverge cannot be rejected. The p-value associated with the test statistic indicates how likely it is to obtain the test statistic (or a more extreme value of this test statistic) if the H_0 is true. p-values smaller than a certain threshold (e.g., $p < .05$) indicate that the model does not explain the data well, whereas p-values larger than .05 are supposed to indicate that model fit is acceptable.

However, there a two major limitations to this approach. First, the test power of the test statistic increases with increasing trial numbers. As all models are only parsimonious approximations to complex cognitive processes, model predictions will always differ from the observed data. Therefore, the test statistic is always going to be significant when trial numbers are large even though the model predictions deviate only negligibly from the observed data (Cohen, 1994).

Second, a *p*-value larger than .05 does not necessarily indicate that the model accounts for the data *well*. Accepting a model implies accepting the null hypothesis with an unknown error probability, as the probability of the observed data given that the H_1 is true is unknown. If the statistical power of the test was known, this probability could be computed. However, an estimation of the statistical power is usually not feasible as there are no conventions for the sizes of diffusion model parameters independent of experimental tasks (such as standardized values for small and large drift rates). Thus, a model might be accepted just because the statistical power is too small to detect a significant deviation from the observed data. Moreover, accepting models with *p*-values larger than .05 may suggest that any model above this threshold accounts for the data equally well, which is obviously not the case.

3.4 Using the root mean square error of approximation to evaluate the goodness of fit of diffusion models (Manuscript 2)

Because existing fit indices such as the AIC and BIC only quantify relative model fit, and because statistical tests such as the χ^2 test are not suited to evaluate goodness of fit due to their dependency on trial numbers and test power, an absolute index of model fit – a goodness-of-fit (GOF) index – is needed. Such as GOF index should quantify the degree of deviation from perfect model fit and that be largely independent of trial numbers.

The root mean square error of approximation (RMSEA; Steiger & Lind, 1980) is one of the most popular GOF indices used in structural equation modeling (SEM; Jackson, Gillaspy, and Purc-Stephenson, 2009). It is based on the noncentrality parameter of the χ^2 distribution. As described in more detail in Manuscript 2, the noncentrality parameter of the χ^2 distribution can also be calculated in the diffusion model framework when the model is fitted to empirical data using the χ^2 statistic as a minimization criterion. The RMSEA is relatively independent of sample size, rewards parsimonious models, has a minimum value of 0 against which the deviation of any specific model from perfect fit can be compared, and it allows calculating confidence intervals around the point estimate of the RMSEA and conducting power analyses (see Browne & Cudeck, 1993; MacCallum, Browne, & Sugawara, 1996). Because these properties would also be very desirable for a GOF index used in the diffusion model framework, we evaluated how well the RMSEA performs as a decision criterion in the evaluation of goodness-of-fit in comparison to the χ^2 criterion (Manuscript 2).

In two simulation studies we showed that the RMSEA is superior to the χ^2 criterion at evaluating goodness of fit in the diffusion model. For this purpose, we simulated data from the diffusion model and manipulated different factors in these simulations, such as the number of model parameters in the generating model, the number of trials, the degree of noise or outlier contamination added to the data, and the number of estimated model parameters.

Subsequently, we assessed how many correct models were rejected by the two decision criteria and how rejection rates were influenced by these different factors. Rejection rates based on the χ^2 criterion increased with increasing trial numbers irrespective of model fit when there was some degree of noise (Study 1) or contamination (Study 2) in the response time data, whereas rejection rates based on the RMSEA were largely invariant with regard to trial numbers. This result is consistent with previous simulation studies that assessed RMSEA performance in the SEM framework, where the RMSEA has been shown to be relatively unaffected by sample size (e.g., Chen, Curran, Bollen, Kirby, & Paxton, 2008).

Moreover, RMSEA values for well-fitting models were comparable to values typically observed for well-fitting models in the SEM framework. Only when the number of trials was very small (i.e., 100 trials in total or 50 trials per condition) there was a larger spread in the distribution of RMSEA values than expected for a well-fitting model. Hence, the cut-off value at which only 5% of the models were incorrectly rejected was much larger than the cut-off value of .05 typically considered indicating good fit in structural equation modeling (Browne & Cudeck, 1993; MacCallum et al., 1996). This tendency to underestimate the goodness of fit when the number of trials was small is consistent with previous simulation studies in structural equation modeling that have shown that the RMSEA rejects too many models when both the degrees of freedom and the sample size are small (Chen et al., 2008; Curran, Bollen, Chen, Paxton, & Kirby, 2003; Kenny, Kaniskan, & McCoach, 2014).

Although the RMSEA is supposed to rewards parsimonious models, it did not succeed in rewarding them sufficiently when evaluating absolute model fit in Study 1. Models in which variability parameters of diffusion model parameters were estimated (e.g., the variability of the drift rate or of the non-decision components) always provided a better fit for the data as indexed by the RMSEA than models without variability parameters – even when the generating model did not include these parameters. This preference for models with inter-trial variabilities may lead to a spurious acceptance of models, as some of our simulations suggested that models with inter-trial variabilities can even account for data heavily contaminated with random noise. However, this may not reflect a fault of the RMSEA, but may rather reflect the greater flexibility of models with estimated inter-trial variabilities in comparison with models in which these variabilities are fixed to zero. Future studies will have to explore whether the estimation of the other model parameters is more biased or more precise when measurement error is captured by inter-trial variability parameters.

Because the RMSEA may underestimate the goodness of fit in diffusion models with small trial numbers, we only considered simulations with trial numbers ≥ 500 when recommending cut-off values for the identification of badly-fitting models. In addition, we suggested different cut-off values for model with and without estimated inter-trial variabilities. Cut-off values were derived by identifying the RMSEA value at which only 5% of the correct models were incorrectly rejected. Based on our simulation results, we suggested that RMSEA values of ≤ 0.08 indicate *acceptable* model fit when inter-trial variabilities are estimated, and that RMSEA values of ≤ 0.08 indicate *acceptable* model fit when inter-trial variabilities are fixed to zero. Overall, we suggested that RMSEA values ≤ 0.05 indicate *good* model fit regardless of whether inter-trial variabilities are estimated.

Taken together, these simulation results support the idea that the RMSEA can be used to evaluate the goodness of fit in the diffusion model framework unless trial numbers are very small. It is superior to the χ^2 statistic in its empirical independency from trial numbers and could supplement other means of evaluating model fit such as graphical tests as a more objective measure. The cut-off values for specific instantiations of the model we suggested in Manuscript 2 can be immediately used in subsequent applications of the diffusion model in individual differences research. Thus, individuals' response time data that cannot be accounted for by the diffusion model can be removed from further multivariate analyses in order to ensure that all model parameters in these analyses are valid estimates of cognitive processes. Thus, establishing the RMSEA as a way to evaluate goodness of fit in the diffusion model is an important step towards extending the measurement of mental speed in individual differences research beyond the measurement of mean response times.

4. The nomological network of mental speed: Factor structure and stability

After introducing different measurement methods of mental speed, I will now outline an internal nomological network of mental speed, in which the relationships between different operationalizations of mental speed and between mental speed measured in different paradigms can be located. Moreover, I will discuss the stability of mental speed in the light of its potential role in explaining individual differences in general intelligence.

4.1 The factor structure of mental speed

Summarizing previous research on the factor structure of mental speed is cumbersome due to a great heterogeneity in participant samples, experimental paradigms, number of variables derived from these paradigms (response times, decision times, motor times, difference parameters, etc.), and analysis plans. Carroll (1993) commented about his reanalysis of 39 data sets of response time tasks that “[…] the available evidence is not sufficient to permit drawing any solid conclusions about the structure of reaction time variables” (p. 6). Over two decades of research later, a meta-analysis or even a systematic review of the factor structure of response times is no more feasible than in Carroll’s time. Nevertheless, a positive manifold seems to emerge in the majority of studies reporting correlations between different response time measures, sometimes in addition to group factors that have not yet been identified in a systematic review of the literature because of the great heterogeneity across different studies (Jensen, 2006). Moreover, whenever decision times (i.e., response times reflecting the speed of information processing) and movement times (i.e., psychomotor response times) are experimentally separated (e.g., with a home button-setup), they tend to load on two orthogonal factors (Carroll, 1993), suggesting that unrelated properties of the cognitive system are responsible for individual differences in the respective response times.

Table 1 summarizes eleven studies in which correlations between different measures of mental speed (e.g., decision times and movement times) and/or between response times in different tasks were reported. I conducted principal component analyses (PCAs) based on the reported correlation matrices if PCAs were not already reported in the original studies. On *N*-weighted average, the first principal component accounted for 52.1 % (SD = 15.9 %) of the total variances, suggesting that a *general mental speed* gives rise to inter-individual differences in a variety of response time tasks. This large *general mental speed* factor is very reminiscent of *g*, which typically explains about 40-50 % of the variance in mental abilities tests (Mackintosh, 2011). It is unclear whether *general mental speed* shares other characteristics of *g* such as its high temporal stability (Larsen, Hartmann, & Nyborg, 2008) and its functional invariance (Johnson et al., 2008).

Table 1 also illustrates the great heterogeneity in participant samples and response time paradigms used in previous research. The amount of variance explained by the first principal component tends to be higher when only response, decision, and/or inspection times are entered into the analysis (Hale & Jensen, 1994; Kyllonen, 1985; Levine, Preddy, & Thorndike,1987; McGarry-Roberts et al., 1992; Miller & Vernon, 1996; Neubauer, Spinath, Riemann, Borkenau, & Angleitner, 2000), and tends to be lower when movement times are included in the correlation (Kranzler & Jensen, 1991; O’Connor & Burns, 2003; Roberts & Stankov, 1999), which is consistent with Carroll’s analysis who reported that response time and movement times loaded on two orthogonal factors (Carroll, 1993). When movement times were removed from the correlation matrices of the three studies, the percent of variance explained by the first principal component increased on average by 10.9 % from 22.9 % - 41.5 % to 35.6 - 43.7 %.

There are a couple of other interesting trends evident in these eleven studies. Both Levine et al. (1987) and Neubauer and Bucik (1996) manipulated the content material of their chronometric tasks (verbal, numerical, and figural/spatial) and did not find support for a hierarchical structure of mental speed with content-related first-order factors. Comparing the factor structure of the two studies involving children (Levine et al., 1987; Miller & Vernon, 1996) to the factor studies of the other studies may provide preliminary evidence for a factorial invariance of mental speed across the age span at least into middle adulthood, although more research is clearly needed.

Overall, the factor structures of three of the four the student samples (Kranzler & Jensen, 1991; O'Connor & Burns, 2003; Roberts & Stankov, 1999) seem more complex than the factor structures of the more heterogeneous samples, suggesting that the factor structure of mental speed may change depending on the sample's cognitive abilities (but see Hale & Jansen, 1994).

There are no systematic investigations of the factor structure of other measures of mental speed, but in some studies correlations between diffusion model parameters or ERP latencies were reported. A structural equation model with a latent drift rate, non-decision time, and boundary separation factor that allowed for correlations between these latent factors provided an acceptable fit for data from a battery of choice response time tasks (Schmiedek et al., 2007). Nunez et al. (2015) found that including a common drift rate factor for each participant improved the predictive validity of a hierarchical Bayesian diffusion model in different conditions of a perceptual decision making task. All in all, research on the correlation of drift rates across different response time paradigms suggests that a common drift rate factor may underlie individual differences in drift rate across different paradigms.

There is even less research on the factor structure of ERP latencies in typical mental speed paradigms. McGarry-Roberts et al. (1992) reported correlations between P300 latencies in different response time tasks that ranged from $r = -.04$ to $r = .63$, and that were best described by a two-factorial solution (eigenvalues: 2.46, 1.32, 0.85, 0.71, 0.42, 0.24) in my reanalysis of their data. However, these two factors cannot easily be interpreted due to significant cross-loading from two of the six tasks, and the small sample size of only 30

Table 1

Summary of eleven studies in which correlations between different measures of mental speed (e.g., decision times and movement times) and/or between response times in different tasks were reported

Study	Sample	Chronometric tasks	Measures	Variance explained by first principal component
Burns & Nettelbeck (2003)	*n* = 90 general population (54% males, M_{age} = 26.6, SD_{age} = 6.7)	Three IT-like paradigms (alphanumeric task, classical IT paradigm on a computer screen, classical IT paradigm on a LED screen), OMO	Alphanumeric SOA, IT (monitor), IT (LED), OMO-DT, OMO-MT	41.6 % 1) [2.08, 0.90, 0.78, 0.65, 0.59] (43.9 % when MTs are removed) [1.76, 0.88, 0.77, 0.59]
Hale & Jansen (1994)	*n* = 40 undergraduates	Line-length discrimination, CRT, letter classification, visual search, mental rotation, abstract matching, mental paper-folding	RT	64.6 % [4.52, 0.87, 0.47, 0.36, …]
Kranzler & Jensen (1991)	*n* = 101 students (49% males, M_{age} = 20.3, SD_{age} = 1.8)	Hick, OMO, visual search, memory search, Posner letter matching, IT	DT, DTSD, MT, MTSD 3)	22.9 % 1) [6.64, 5.28, 2,40, 1.85, …] (35.6 % when MTs are removed) [5.34, 1.54, 1.33, 1.15, …]
Kyllonen (1985)	*n* = 178 Airforce trainees	SRTs (left hand, right hand), CRTs (L vs. D, even vs. odd, positive vs. negative, vowel vs. consonant), categorization (words, letters), sequential matching (words, letters), simultaneous matching (words, letters)	RT	56.5 % 2) [3.40, 0.96, 0.64, 0.44, …]
Levine, Preddy, & Thorndike (1987)	*n* = 300 children (1/3 4th, 7th, 10th grade each)	CRTs (verbal-perceptual, verbal-semantic, quantitative-perceptual, quantitative-symbolic, spatial-single figure, spatial-complex figures)	DT	62.3 % 1) [3.74, 0.63, 0.59, 0.44, …]
McGarry-Roberts, Stelmack, & Campbell (1992)	*n* = 30 women (age 18 to 25)	SRT, CRT, Sternberg memory scanning, Posner letter matching (physical similarity, semantic similarity, category matching)	DT	75.3 % 1) [4.52, 0.90, 0.26, 0.14, …]
Miller & Vernon (1996)	*n* = 109 children (55% males, M_{age} = 5.5, SD_{age} = 0.9)	CRTs (shape, color, size, number, arrow direction), shape string test, color string test, matching test	RT	64.8 % [5.22, 0.74, 0.68, 0.52, …] 1)

Study	Sample	Chronometric tasks	Measures	Variance explained by first principal component
Neubauer & Bucik (1996)	*n* = 120 general population (53% males, M_{age} = 28.33, SD_{age} = 7.94)	Coding, Sternberg memory scanning, Posner letter matching (all tasks were presented as time-restricted pen-and-paper versions with verbal, numerical, and figural content)	Number of correctly solved items	46.7 % [11.2, 2.17, 1.39, 1.24, …]
Neubauer, Spinath, Riemann, Angleitner, & Borkenau (2000)	*n* = 600 general population (22% males, M_{age} = 34.3, SD_{age} = 13.0)	Sternberg memory scanning (one, three, five digits), Posner letter matching (physical identity, name identity)	RT	58.4% [no correlation matrix or eigenvalues reported]
O'Connor & Burns (2003)	*n* = 102 students and well-educated general population (33% males, M_{age} = 22.0, SD_{age} = 5.9)	Perceptual speed tasks (digit symbol, mental rotation, cross out), IT, CRT (Jensen box), OMO, Triplet Numbers Test, Swaps Test	RT, DT, MT	32.6 % [1)] [5.87, 3.45, 2.10, 1.33, …] (40.1 % when MTs are removed) [5.62, 2.40, 1.34, 0.97]
Roberts & Stankov (1999)	*n* = 179 students and general population (39% males, M_{age} = 12.58, SD_{age} = 6.18)	Fitt's movement, joystick reaction, SRT, tachistoscopic CRT, complex CRT (sequentially lit lights in Jensen box), binary reaction task (arbitrary decision rule associated with lights in Jensen box), single card-sorting, multitask card-sorting (simultaneous word-classification), single word-classification, multitask word-classification (simultaneous card-sorting)	DT, MT	31.5% [1)] [5.67, 2.54, 1.75, 1.37, …] (43.7 % when MTs are removed) [3.93, 1.60, 0.91, 0.81, …]

Note. IT = inspection time; OMO = odd-man-out; SRT = simple reaction time; CRT = choice reaction time; SOA = stimulus onset asynchrony; RT = reaction time; DT = decision time (i.e., when experimentally separated from the motor response via a home button-design); MT = movement time. For more details about the chronometric tasks please refer to the original studies. Squared brackets in the last column show the first four eigenvalues of the principal component analysis.

[1)] Data were reanalyzed based on the correlation matrices reported in the original studies.

[2)] Data were reanalyzed based on the first-order factor correlation matrix reported in Carroll (1993), because correlation matrices of the RT variables were not available in the original study.

[3)] Additionally reported slope and intercept measures were removed from this analysis, because a) they provide no valid estimates of information processing speed (see Manuscript 1), and b) including them causes the correlation matrix to not be positive definite due to the linear dependency between variables.

women warrants further research before firm conclusions can be drawn. Response times and ERP latencies were virtually unrelated in their study with the exception of a correlation of $r = .39$ in the choice response time paradigm. On an intraindividual level, shorter single trial P300 latencies have been repeatedly been shown to predict faster response times (Holm, Ranta-aho, Sallinen, Karjalainen, & Müller, 2006; Kutas, McCarthy, & Donchin, 1997). Clearly more research on the association between the latencies of different ERP components in response time tasks is needed.

4.2 The stability of mental speed

Because *g* is highly stable over time (Carroll, 1993; Larsen et al., 2008), any candidate cognitive process underlying individual differences in *g* should show a similar temporal relative stability. Manifest measures of response times and ERP components may, however, be strongly affected by situational factors such as motivation, familiarity with the testing environment, or fatigue. Hence, the amount of variance with a high temporal stability that is associated with general intelligence may be relatively small in comparison to the amount of variance that is affected by situational factors and that may or may not be related to general intelligence. Consequently, correlations based on single measurements of mental speed may underestimate the relationship between mental speed and mental abilities.

Previous research suggests a good temporal stability of response times and diffusion model parameters over the period of one to two weeks (Clayson & Larson, 2013; Jensen, 2006; Lerche & Voss, submitted; Yap, Balota, Sibley, & Ratcliff, 2012) unless difference or slope parameters are estimated (Roznowski & Smith, 1993). Studies on age effects on mental speed have focused on Brinley plots, in which the response times of older adults are plotted against the response times of younger adults, typically resulting in a slope larger than one and a negative intercept (e.g., Brinley, 1965; Fisk & Fisher, 1994; Myerson & Hale, 1993). Recent research has shown that this finding is not indicative of a general slowing of information processing in older adults, but that it can be explained by a discrepancy in the relative variance in response times across conditions for the two age groups (Ratcliff, Spieler, McKoon, 2000). Although these results shed some light on the processes underlying age-related cognitive slowing, they provide no information about the *relative* stability of mental speed, which has not yet been assessed in systematic longitudinal studies.

There are also only few studies on the temporal stability of ERP latencies. Cassidy, Robertson, & O'Connell (2012) found that four-week test-retest correlations varied substantially across ERP components with higher ($r > .62$) correlations reported for the P1, N1, and N170 peak latency, and lower ($r < .44$) correlations reported for the P3a and P3b peak latencies and two components of error processing (ERN, Pe). Brunner et al. (2013) reported test-retest reliabilities of $ICC = .86$ across 6 – 18 months for the P3 peak latency in a Go/NoGo paradigm, Williams et al. (2005) reported a four-week test-retest correlation of $r = .93$ for the P150 peak latency and of $r = .56$ for the P300 peak latency in a working memory paradigm, and Fabiani, Gratton, Karis, and Donchin (1987) reported an average one-/two-week test-retest correlation of $r = .56$ for the P300 peak latency in different oddball paradigms. The great heterogeneity of these results suggests a) that the temporal stability of ERP peak latencies likely depends on several factors, e.g. the specific component, the specific paradigm, the participant sample, or the test-retest interval, and b) that longitudinal studies on ERP latencies should always report test-retest correlations. Nevertheless, these results also suggest that ERP latencies do not generally have a high temporal stability.

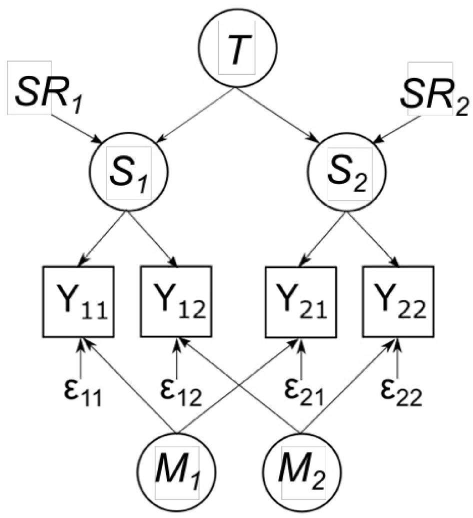

Figure 3. Basic LST model with two parallel measures and two measurement occasions. The model consists of two method factors (M_1, M_2), two state residuals (SR_1, SR_2), one common trait (T), and four measurement error variables (ε_{11}, ε_{12}, ε_{13}, ε_{14}).

However, test-retest correlations may not be ideal estimates of the relative temporal stability of variables in some cases. Test-retest correlations are insofar suboptimal estimates of temporal stability as their size depends not only on the temporal stability of a variable, but also on its reliability (i.e., the proportion of variance not affected by unsystematic measurement error). This proportion of variance may be rather small (resulting in a low reliability), but still highly stable over time (which should be reflected in a high temporal stability). In order to get a better estimate of the temporal stability of a variable, its stability has to be disentangled from its reliability, which can be achieved by means of the latent-state-trait (LST) theory (Steyer, Fering, & Schmitt, 1992; Steyer, Mayer, Geiser, & Cole, 2015).

LST theory is an expansion of classical test theory that takes into account that any measurement is always affected by situational factors. In short, LST theory proposes that the variance of a variable Y_{ij} can be decomposed into the variance of the trait T, the variance of a state residual SR_i, the variance of a method residual M_j, and the variance of an unsystematic error residual ε_{ij}. As illustrated in Figure 3, the stability of a measurement can then be evaluated by comparing the proportion of occasion-specific variances $\sigma^2(SR_1)$ and $\sigma^2(SR_2)$ with the shared trait variance $\sigma^2(T)$ across the two measurement occasions in relation to the total variance $\sigma^2(Y_{ij})$. Thus, LST theory allows estimating how large the amount of variance with a high trans-situational consistency that may be associated with general intelligence is in comparison to the amount of variance that is affected by situational factors and that may or may not be related to general intelligence.

4.3 The psychometric properties and factor structure of mental speed in the latent-state-trait framework (Manuscript 3)

In order to analyze the factor structure and stability of mental speed – measured as response times and ERP latencies –, we recruited 134 participants between 18 and 60 years from different educational and occupational backgrounds and measured response times and ERP latencies in three response time tasks at two measurement occasions that were approximately eight months apart. About four months after the first laboratory session, we measured general intelligence and personality traits in a diagnostic assessment. The response time and ERP latency data from 122 participants who completed at least the first laboratory and the diagnostic assessment were entered separately into structural equation models that were an extension of LST theory (for details, see the method section of Manuscript 3).

For the response time data, a LST model with a common trait T, a state residual SR_i for each of the two measurement occasions, and a hierarchical method factor M_j for each of the three experimental paradigms provided a good fit, $\chi^2(133) = 265.81$, $p <.001$, CFI = .95, RMSEA = .09. The fact that response times measured in different paradigms loaded onto a broad general mental speed factor that explained the greatest amount of variance in manifest RT measurements is consistent with previous studies summarized in Table 1 supporting the idea of a broad general processing speed factor. It should be noted that although the response time paradigms in this study were simple ECTs, the method factors M_j were significant and except for the Posner letter matching paradigm far from negligible in their size, indicating that there are additional modality- or at least task-specific processing speed factors beyond general mental speed, which is consistent with previous studies reporting multi-factorial models of mental speed (e.g., O'Connor & Burns, 2003; Roberts & Stankov, 1999). Moreover, the factor structure might even be more complex when a broader range of response time tasks is included in the analysis.

The variance of the state residual SR_1 was not significant, whereas the variance of SR_2 was significant, but very small in comparison to the common trait variance, indicating that the influence of situational factors on response times is generally negligible. This result is consistent with previous studies reporting good stabilities of response time data (Clayson & Larson, 2013; Jensen, 2006; Lerche & Voss, submitted; Yap et al., 2012).

For the ERP latency data, we compared a general mental speed LST model with a common trait T to a specific mental speed LST model with two common traits, $T_{ealier\ latencies}$ and $T_{later\ latencies}$, that loaded onto earlier (P100, N100) and later (P200, N200, P300) ERP components in the stream of information processing. The specific mental speed model provided a better fit than the general mental speed model, Δ AIC = 177.9. This result is consistent with previous research showing that P300 latencies are correlated across different response time paradigms (McGarry-Roberts et al., 1992). However, our results extend this finding to other ERP components that covary across tasks. We found two separate common traits that loaded onto earlier and later latencies, respectively, which is reminiscent of the scientifically outdated distinction between exogenous and endogenous ERP components. This result suggests that individuals who are faster in higher-order information processing are not necessarily also faster at the onset of information processing.

All state residuals except for the state residual of earlier latencies and of the N200 latency at the second laboratory session were not significant, suggesting that situational influences on ERP latencies were rather small. This is consistent with previous research reporting moderate (Fabiani et al., 1987; Cassidy et al., 2012) and even high (Brunner et al., 2013; Williams et al., 2009) test-retest correlations for the latency of several ERP components, but does not replicate the great heterogeneity in test-retest correlations reported previously (Cassidy et al., 2012).

Despite a comparable temporal stability, ERP latencies had lower consistencies and reliabilities than response times. This suggests that different ERP components measured in the same task and the same ERP component measured across different tasks only share a small amount of variance. Hence, reliably assessing the neural speed of higher-order information processing requires multiple measurements of ERP latencies in different tasks.

Taken together, among these two measures of mental speed, ERP latencies seem to have a more complex factor structure in ECTs than response times. Situational influences are not likely to have a great influence on either measure, which leads to the conclusion that both measures of mental speed have a high temporal stability. The reliability of ERP latencies, however, is substantially lower than the reliability of response times, indicating that the relationship between ERP latencies and general intelligence may be underestimated unless corrected for unreliability or estimated as a latent correlation.

5. The relationship between mental speed and general intelligence

In a review of 172 studies on the relationship between response times and mental abilities, Sheppard and Vernon (2008) reported an average correlation of $r = -.24$ between the two measures. Previous research has largely supported the notion that correlations between response times in the Hick paradigm and other single and choice response time tasks increase linearly as a function of choice alternatives (Jensen, 2006), increasing from $r = -.22$ to -.44 (Sheppard & Vernon, 2008). Similarly, the association between response times in letter and category matching tasks increases as a function of the *g*-loading of tasks (Jensen, 2006). In memory scanning paradigms, however, no relationship between the size of the memory set and the resulting correlation between mental abilities and response times is apparent with correlations ranging from $r = -.25$ to -.45 (Sheppard & Vernon, 2008). In general, correlations between composite measures of mental speed and mental abilities tend to be higher than the correlations of single response time measures. Canonical correlations between different test batteries of response time tasks and general intelligence ranged from $C = .55$ to .72 (Kranzler & Jensen, 1991; Miller & Vernon, 1996; Saccuzzo, Larson, & Rimland, 1986). This suggests that it is the general mental speed factor and not specific facets of mental speed that is associated with general intelligence.

The association between ERP latencies and mental abilities is notably weaker and more inconsistent. Schulter and Neubauer (2005) reviewed ten studies with a total sample size of 1,183 participants reporting negative, and twelve studies with a total sample size of 1,219 participants reporting non-significant correlations between ERP latencies and mental abilities. The most consistent results are found for the latency of the P300 component, which has been interpreted in terms of stimulus evaluation and categorization (Callaway, 1983; McCarthy & Donchin, 1981), context-updating (Donchin, 1981; Polich, 2007), and context-closure (Verleger, 1988). Moreover, somewhat consistent results have also been reported for the latency of the mismatch negativity (MMN), which is evoked by sudden deviations from regular auditory stimulations (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). More intelligent individuals tend to have shorter P300 and MMN latencies than less intelligent individuals (Bazana & Stelmack, 2002; Beauchamp & Stelmack, 2006; McGarry-Roberts et al., 1992; Russo, De Pascalis, Varriale, & Barratt, 2008; Troche, Indermühle, Leuthold, & Rammsayer, 2015).

Considering the results of the LST analysis in Manuscript 3, it is not surprising that the most consistent results are found for the P300, which was the only ERP latency with reliabilities consistently exceeding .60 in all paradigms and at both laboratory sessions. Thus, correlations between other ERP latencies and mental abilities may only be found when ERP latencies are aggregated over different measurements due to their low reliabilities. Moreover, even the weak correlations between P300 latencies and mental abilities are likely underestimated due to the mediocre reliability of P300 latencies and will probably increase when P300 latencies are aggregated or modeled hierarchically across different measurements.

5.1 Explaining the relationship between mental speed and mental abilities

Several theories of the relationship between mental speed and mental abilities have been suggested that can be organized into two broad categories. The first set of theories proposes that chronometric and psychometric measures are correlated because more intelligent individuals have an advantage in some brain-wide property. One candidate property is myelination, as a denser myelin layer may facilitate the speed of impulse propagation along the axon through saltatory conduction (Fields et al., 2014; Yakovlev & Lecours, 1967; Wake, Lee, & Fields, 2011). White matter microstructure has been associated with inhibition (Liston et al., 2006) and working memory capacity (Vertergaard et al., 2011). In addition, processing speed has been shown to mediate the relationship between white matter microstructure and reasoning in children (Ferrer et al., 2013), and between white matter tract integrity and general intelligence in adults (Penke et al., 2012). Moreover, previous research supports the notion that white matter tract integrity is a brain-wide property (Penke et al., 2010). Thus, individual differences in myelination and white matter tract integrity may influence the speed of neural propagation and information processing, and may thus give rise to individual differences in intelligence (Miller, 1994; Penke et al., 2012).

Another candidate brain-wide property proposed by Jensen (2006) is the speed of neural oscillations. Neural oscillation theory proposes that the level of excitatory potential oscillates synchronously among a large number of neurons and that an external stimulus is processed faster and more efficiently when it occurs at the peak than at the trough of the sine wave. Faster oscillations thus lead to more excitatory phases per time interval, which results in less time elapsed until a stimulus occurring at the trough of the oscillatory wave can be processed, leading to faster information processing and subsequently faster and less variable response times. Jensen proposes that the rate of neural oscillations underlies individual differences in general intelligence, because faster information processing (resulting from a higher rate of oscillations) allows the processing of more information without having to temporarily store this information in and then retrieve short-term memory. Thus, information loss is less likely, as less information has to be stored in short-term memory, where it may get lost as a consequence of rapid decay.

Evaluating evidence in favor of the neural oscillation theory is difficult, as the theory is very unspecific in its description of which neurons are supposed to oscillate in synchrony and how this oscillation might be quantified. Jensen (2006) discusses different neural oscillations such as the alpha and beta rhythm, but justifies the theory in terms of single cell activity. Thus, a clear-cut empirical test of the theory is very difficult. A recent meta-analysis of 24 studies supported the notion that intelligent individuals have faster and less variable response times, although SDs of response times were not more strongly related to intelligence than mean response times as predicted by neural oscillation theory (Doebler & Scheffler, 2015). On a neurophysiological level, individual alpha frequency (i.e., the individual peak frequency in the alpha band ranging from 8 to 13 Hz) exhibits trait-like characteristics (Grandy et al., 2013b) and is associated with working memory capacity (Clark et al., 2004) and intelligence (Anokin & Vogel, 1996; Grandy et al., 2013a). Due to the low specifity of the neural oscillation hypothesis, however, it is unclear whether individual alpha frequency is a valid operationalization of the speed of neural oscillations, or whether only single cell recordings would provide an adequate measurement.

The second set of theories proposes that chronometric and psychometric measures are correlated because more intelligent individuals show advantages in some specific process or property of the information-processing system. Likely candidates are executive functions such as shifting, inhibition, updating, selective attention, and attentional control. Previous research found that higher intelligence was moderately related to updating (Benedek, Jauk, Sommer, Arendasy, & Neubauer, 2014; Wongupparaj, Kumari, & Morros, 2015) and inhibition (Wongupparaj et al., 2015), but not to shifting (Benedek et al., 2014; Wongupparaj et al., 2015). Moreover, attentional control (measured as task-unrelated thoughts) has been shown to be associated with response times (McVay & Kane, 2012), working memory capacity (McVay & Kane, 2012; Mrazek et al., 2012) and intelligence (Mrazek et al., 2012). In addition, both response time variability and general intelligence have been related to the activity of the default-mode network (Basten, Stelzel, & Fiebach, 2013; Kelly, Uddin, Biswak, Castellanos, & Milham, 2008; Weissman, Roberts, Visscher, & Woldorff, 2006), which underlies task-unrelated thoughts such as mind wandering, autobiographical planning and daydreaming and which has to be de-activated during stimulus processing to allow a reallocation of resources to task-relevant brain networks (McKiernan, Kaufman, Kucera-Thompson, & Binder, 2003). Taken together, each of these theoretical frameworks suggests that executive functions or attentional control act as a common cause influencing both mental speed and mental abilities.

Another candidate specific process often discussed in the context of the relationship between mental speed and mental abilities is working memory. As illustrated in the context of the neural oscillation hypothesis above, a faster speed of information processing implies that more items can be processed in working memory before they have to be stored in short-term memory. More importantly, the number of items that can be held in short-term memory simultaneously is limited and single items may be forgotten once their trace activation falls under a critical threshold (Barrouillet, Bernardin, & Camos, 2004; Burgess & Hitch, 2006; Page & Norris, 1998). The time-based resource sharing model proposes that attention can be either deployed to task-relevant processing or to the refreshing of memory traces (Barrouillet et al., 2004). A higher mental speed during task-relevant processing results in less time spent on task-relevant processing and thus more time available for memory refreshing, resulting in less time-based forgetting. In addition, higher mental speed may also allow a faster and more efficient refreshing of memory traces during the time the attentional system deploys to refreshing. Previous research has shown that an interaction between individual differences in mental speed and memory span loaded more strongly on a factor of biological intelligence than either of the two measures ($r_{mentalspeed} = .85$, $r_{memoryspan} = .49$, $r_{interaction} = .92$), although the difference in factors loadings between mental speed and its interaction with memory span was not great (Kline, Draycott, & McAndrew, 1994). Moreover, there is evidence that developmental changes in working memory capacity are mediated by developmental changes in mental speed, and that developmental changes in fluid intelligence are in turn mediated by changes in working memory capacity and mental speed (Fry & Hale, 1996). It should be noted, however, that the theory of trace decay in short-term memory is not undisputed (Jonides et al., 2008; Oberauer & Lewandowsky, 2014).

All in all, these theories are very prominent examples of two sets of distinct theories that propose that either some brain-wide property or some specific process underlies the relationship between mental abilities and mental speed. A direct empirical test of these two sets of theories can be achieved by means of event-related potentials.

5.2 Do more intelligent individuals have advantages in the capacity of some brain-wide property or in the speed of specific processes? (Manuscript 3)

In order to test the plausibility of these two sets of hypotheses, we added a hierarchical model of general intelligence to the LST model of ERP latencies in Manuscript 3. The resulting structural equation model thus consisted a) of a LST model with two broad common traits for earlier and later latencies, respectively, and specific common traits for the P200 and P300 latency (other specific traits were not significant), and b) of a hierarchical model of general intelligence with *g* loading on the Berlin intelligence structure test (BIS; Jäger, Süß, & Beauducel, 1997) and Raven's Advanced Progressive Matrices (APM; Raven, Court, & Raven, 1994).

General intelligence was positively correlated with the common trait for earlier latencies, $r = .33$, $p <.001$, and negatively correlated with the common trait for later latencies, $r = -.89$, $p <. 001$. These results support the view that more intelligent individuals show a higher mental speed at specific, but not at all stages of information processing. Hence, they contradict theories proposing that mental speed and mental abilities are related because more intelligent individuals have advantages in some brain-wide property. Instead, more intelligent individuals showed faster neurophysiological mental speed only in ERP latencies that are generally associated with higher-order processing and showed slower mental speed in ERP latencies occurring early in the stream of information processing.

To be more specific, it is the fact that P300 latencies showed the greatest association with general intelligence that supports both specific theories on the relationship between mental speed and mental abilities outlined above. Faster P300 latencies are thought to reflect a more efficient inhibition of task-irrelevant processes that facilitates the updating of temporal-parietal memory storage processes by attentional and working memory processes (Polich, 2007). Thus, the context-updating interpretation of the P300 allows explaining the relationship between mental speed and mental abilities both in terms of underlying individual differences in executive functions and in terms of a more efficient time-based resource sharing or memory trace refreshing in working memory.

ERP latency traits were not correlated with specific intelligence tests beyond their association with *g* and were not related to specific subscales of the BIS such as processing speed. This result is consistent with previous research showing the mental speed measured as response times shows the greatest associations with general intelligence and not with specific first- or second-order factors of intelligence (Jensen, 2006). However, the same result was not obtained for response times in Manuscript 3, because both general intelligence and the processing speed component of the BIS were related to the common response speed trait, but response speed was still more strongly related to general intelligence, $r = .43$, $p < .001$, than to the processing speed component of the BIS, $r = .24$, $p = .005$.

5.3 A tentative cognitive model of individual differences in general intelligence

Based on these results, I propose a tentative cognitive model of mental speed, attentional control, and working memory that may explain individual differences in general intelligence (see Figure 4). The model is an extension of the time-based resource-sharing model (Barrouillet et al., 2004), which regards attention as a bottleneck in working memory processes. According to the time-based resource-sharing model, attention can be either deployed to task-relevant processing or to the refreshing of memory traces, and hence both processing stages can only occur sequentially, never in parallel. The black rectangles in Figure 4 represent time spent on task-relevant processing, which includes all stages of information processing from encoding over information accumulation and manipulation to decision making, but also meta-cognitive processes such as strategy planning, goal setting, and goal monitoring. Note that the inclusion of a random-walk process of information accumulation in the black squares in Figure 4 only serves to illustrate that information accumulation is an important part of task-relevant processing, but does not imply that this is the only process occurring during task-relevant processing. Moreover, task-relevant processing is a very generic term that can be applied to any kind of task requiring multiple processing steps (represented as multiple black boxes) from complex response time tasks over working memory task to intelligence test items. The white rectangles in Figure 4 represent the time that is available to refresh decaying memory traces between processing steps. Lines in the white rectangles represent distinct memory traces that can be refreshed during the available time. Because attention is assumed to be a domain-general process that can only be deployed to one cognitive process at a time, the working memory system oscillates between stages of information processing and memory refreshing.

I propose that individual differences in two components of the extended model – attentional control and mental speed – are related to individual differences in general intelligence. First, more intelligent individuals are thought to possess greater attentional control, which subsequently reduces task-unrelated thoughts and minimizes the effects of distraction and mind wandering on information processing. Specifically, individual differences in attentional control should affect the consistency of information processing, affecting a) the time spent on task-relevant processing and b) mental speed.

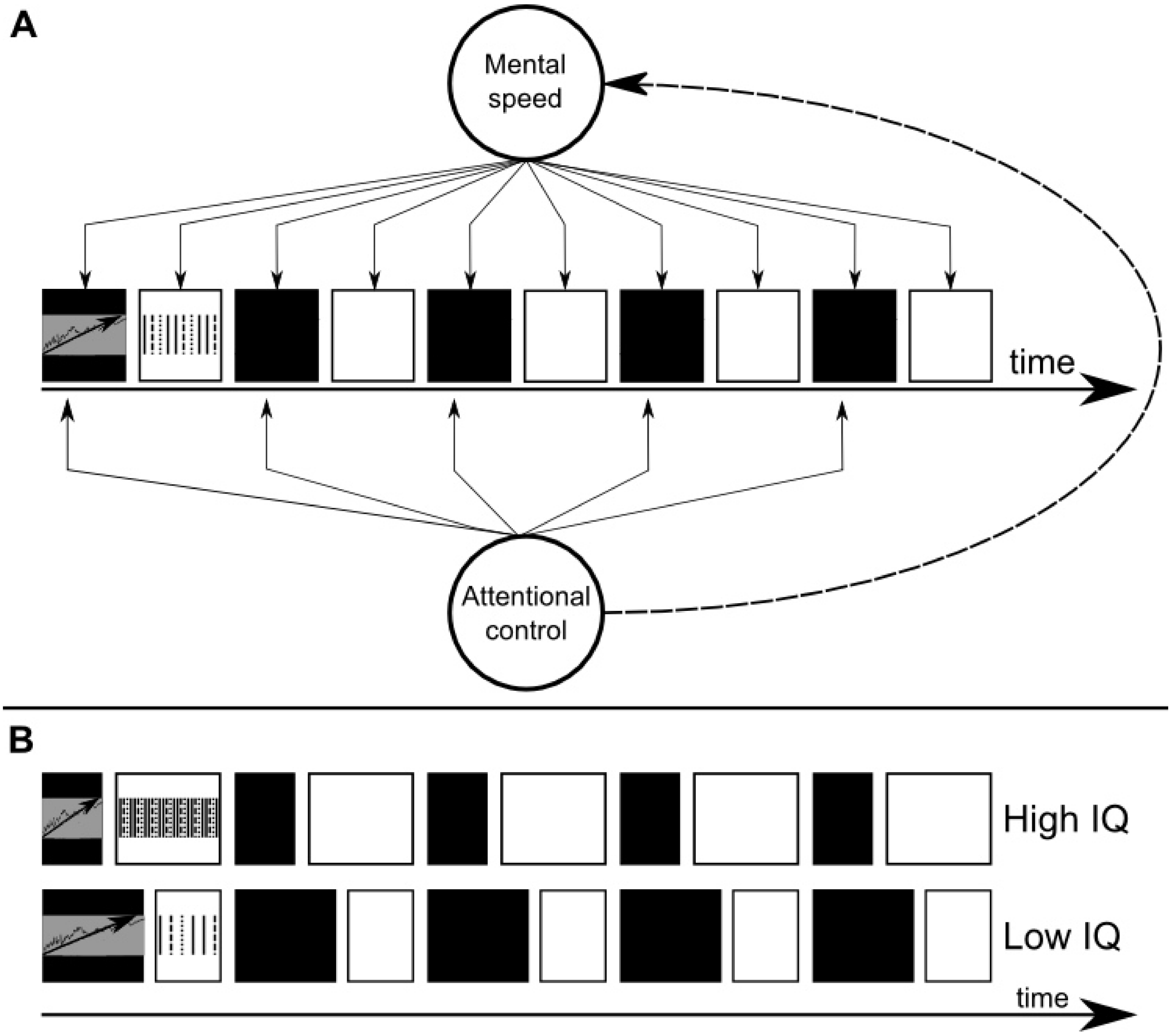

Figure 4. A) A tentative cognitive model of individual differences in general intelligence that is based on the time-based resource-sharing model (Barrouillet et al., 2004). Black rectangles represent the time attention is deployed to task-relevant processing, whereas white rectangles represent the time attention is deployed to the refreshing of memory traces. Attentional control influences the time of task-relevant processing and mental speed, which in turn influences the time of task-relevant processing and the number of memory traces refreshed during a given time interval. B) Model predictions are shown separately for more (high IQ) and less (low IQ) intelligent individuals. The model predicts that more intelligent individuals can spend more time refreshing memory traces (reflected by broader white rectangles), arriving later at the point of working memory breakdown.

Second, a greater mental speed is likely to affect the time that attention has to be devoted to task-relevant processing by facilitating evidence accumulation and decision making. Here, integrating a diffusion model account of decision making leads to the prediction of a higher drift rate for more intelligent individuals, which results in faster and more accurate task-relevant processing. Faster task-relevant processing (represented as the narrower black rectangles for more intelligent individuals in Figure 4B) allows more time spent on the refreshing of memory traces (reflected by broader white rectangles in Figure 4B) before the onset of the next processing step for more intelligent individuals. Moreover, a greater mental speed is also going to increase the number of memory traces that can be refreshed during the time attention is deployed to the refreshing of short-term memory (represented by a higher density of lines in the white rectangles). The model thus predicts that more intelligent individuals can spend more time refreshing memory traces more efficiently, arriving later at the point of working memory breakdown.

Note that the model assumes that any time between task-relevant processing steps is spent on memory refreshing, which does therefore imply that more intelligent individuals do not spend overall less time on the task. This assumption is consistent with previous results showing that the correlation between item response times and reasoning ability is close to zero (Kyllonen, 1985) or even positive (Klein Entnik, Fox, & van der Linden, 2009; Klein Entnik, Kuhn, Hornke, & Fox, 2009) in reasoning tests and disappears with increasing item difficulty in intelligence tests (Goldhammer, Naumann, & Greiff, 2015).

Obviously, this tentative model has to be subjected to empirical tests before any statements about its validity can be made. Such empirical tests might either consist of multivariate analyses of mental speed, attentional control, and working memory parameters estimated based on the mathematical implementation of the time-based resource-sharing model (Oberauer & Lewandowsky, 2011), or of an evaluation of the predictive validity of a computational implementation of the extended model. Alternatively, targeted experimental or pharmacological manipulations might be employed to assess the influence of attentional control and mental speed on model parameters.

6. Summary and Conclusion

The aim of the present work was to overcome problems associated with the theoretical conceptualization and measurement of mental speed in individual differences research. This aim was pursued by first expanding the measurement of mental speed beyond response times, second establishing a nomological network of mental speed by investigating its factor structure across different paradigms and different operationalizations, and third locating it in a larger nomological network including other cognitive abilities.

For this purpose, I showed that the measurement of mental speed can be expanded with the help of diffusion models and event-related potentials. In particular, I showed that difference and slope parameters in elementary cognitive tasks are not suited to identify the speed of specific cognitive processes and that this goal can be better achieved by means of diffusion modeling and ERP analysis. Because difficulties in the evaluation absolute of model fit are one of the greatest hindrances to the proliferation of diffusion models in individual differences research, I demonstrated how the RMSEA may be used to evaluate absolute model fit and to subsequently identify individuals whose data cannot be properly accounted for by the diffusion model.

In a second step, I investigated the factor structure and the psychometric properties of mental speed on a behavioral and neurophysiological level. On the behavioral level, there was evidence for a strong general mental speed factor that accounted for 53 - 71 percent of the variance in single response time measurements. On the neurophysiological, I found evidence for two uncorrelated mental speed factors – one general factor for latencies earlier and one general factor for latencies later in the stream of information processing. Moreover, structural equation modeling also revealed specific trait factors for P200 and P300 latencies. Situational influences such as fatigue or motivation on measurements were negligible on both measurement levels.

In a third step, I started to locate mental speed in a larger nomological network by exploring its relationship with general intelligence. Structural equation modeling revealed that on a behavioral level, general intelligence was associated with generally faster response times, whereas on a neurophysiological level, general intelligence was differentially related to earlier and later ERP latencies and was most strongly associated with ERP latencies reflecting the speed of higher-order processing. Taken together, ERP latencies explained about 85 percent of the variance in general intelligence, whereas response times explained only 19 percent. This result suggests that ERP latencies may provide a purer measurement of mental speed as *"the actual time taken to process information of different types and degrees of complexity" (Jensen, 2006, p. ix) goal-directedly* and that response times may be contaminated by additional processes such as motor planning and execution that are largely unrelated to general intelligence. Finally, I proposed a tentative cognitive model explaining the relationship between mental speed and mental abilities based on individual differences in attentional control and working memory processes.

All in all, these findings show that mental speed is an intriguing concept that is crucial to explaining individual differences in mental abilities. What is remarkable about these results is that neurophysiological mental speed explained 90 percent of the variance in general intelligence. This suggests that ERP latencies could be used for the individual assessment of intelligence in participants not willing or capable to complete intelligence tests, if they could be determined with a higher reliability. Moreover, these results further strengthen the position that any process model of general intelligence has to account for the central role of mental speed in mental abilities.

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Anokhin, A., & Vogel, F. (1996). EEG alpha rhythm frequency and intelligence in normal adults. *Intelligence*, *23*, 1–14. doi:10.1016/S0160-2896(96)80002-X

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

Basten, U., Stelzel, C., & Fiebach, C. J. (2013). Intelligence is differentially related to neural effort in the task-positive and the task-negative brain network. *Intelligence*, *41*, 517–528. doi:10.1016/j.intell.2013.07.006

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. doi:10.1037/0022-3514.83.4.998

Beauchamp, C. M., & Stelmack, R. M. (2006). The chronometry of mental ability: An event-related potential analysis of an auditory oddball discrimination task. *Intelligence*, *34*, 571–586. doi:10.1016/j.intell.2006.03.007

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence*, *46*, 73–83. doi:10.1016/j.intell.2014.05.007

Brinley, J. F. (1965). Rigidity and the control of cognitive sets in relation to speed and accuracy of performance in the elderly. *Dissertation Abstracts*, *26*, 1158–1159.

Brookhuis, K. A., Mulder, G., Mulder, L., & Gloerich, A. (1983). The P3 complex as an index of information processing: The effects of response probability. *Biological Psychology*, *17*, 277–296. doi:10.1016/0301-0511(83)90004-2

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Sage focus editions: Vol. 154. Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications.

Brunner, J. F., Hansen, T. I., Olsen, A., Skandsen, T., Håberg, A., & Kropotov, J. (2013). Long-term test-retest reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo wave in a visual Go/NoGo task. *International Journal of Psychophysiology*, *89*, 106–114. doi:10.1016/j.ijpsycho.2013.06.005

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*, 627–652. doi:10.1016/j.jml.2006.08.005

Burns, N. R., & Nettelbeck, T. (2003). Inspection time in the structure of cognitive abilities. *Intelligence*, *31*, 237–255. doi:10.1016/S0160-2896(02)00120-4

Callaway, E. (1983). The pharmacology of human information processing. *Psychophysiology*, *20*, 359–370. doi:10.1111/j.1469-8986.1983.tb00915.x

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

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*, 659–664. doi:10.1111/j.1469-8986.2011.01349.x

Cattell, J. M. (1886). The time taken up by mental operations. *Mind*, (42), 220–242. doi:10.1093/mind/os-XI.42.220

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*, 462–494. doi:10.1177/0049124108314720

Clark, C. R., Veltmeyer, M. D., Hamilton, R. J., Simms, E., Paul, R., Hermens, D., & Gordon, E. (2004). Spontaneous alpha peak frequency predicts working memory performance across the age span. *International Journal of Psychophysiology*, *53*, 1–9. doi:10.1016/j.ijpsycho.2003.12.011

Clayson, P. E., & Larson, M. J. (2013). Psychometric properties of conflict monitoring and conflict adaptation indices: Response time and conflict N2 event-related potentials. *Psychophysiology*, *50*, 1209–1219. doi:10.1111/psyp.12138

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49*, 997-1003.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. doi:10.1037/h0040957

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, *32*, 208–252. doi:10.1177/0049124103256130

Deary, I. (2008). Why do intelligent people live longer? *Nature*, *456*, 175–176. doi:10.1038/456175a

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia, 50*, 1271-1285. doi:10.1016/j.neuropsychologia.2012.02.011

Der, G., Batty, G. D., & Deary, I. J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, *37*, 573–580. doi:10.1016/j.intell.2008.12.002

Doebler, P., & Scheffler, B. (2015). The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learning and Individual Differences.* doi:10.1016/j.lindif.2015.02.009

Donchin, E. (1981). Surprise! … Surprise? *Psychophysiology*, *18*, 493–513. doi:10.1111/j.1469-8986.1981.tb01815.x

Donders, F. C. (1868/1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. doi:10.1016/0001-6918(69)90065-1

Dunn, B. R., Dunn, D. A., Languis, M., & Andrews, D. (1998). The relation of ERP components to complex memory processing. *Brain and Cognition, 36*, 355-376. doi:10.1006/brcg.1998.0998

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. In P. K. Ackles, J. R. Jennings, & M. G. H. Coles (Eds.), *Advances in Psychophysiology, Vol. 2*. Greenwich, CT: JAI Press, Inc., pp. 1-78.

Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*, 148–160. doi:10.1016/0168-5597(94)90055-8

Ferrer, E., Whitaker, K. J., Steele, J. S., Green, C. T., Wendelken, C., & Bunge, S. A. (2013). White matter maturation supports the development of reasoning ability through its influence on processing speed. *Developmental science*, *16*, 941–951. doi:10.1111/desc.12088

Fields, R. D., Araque, A., Johansen-Berg, H., Lim, S.-S., Lynch, G., Nave, K.-A., . . . Wake, H. (2014). Glial biology in learning and cognition. *The Neuroscientist: a review journal bringing neurobiology, neurology and psychiatry*, *20*, 426–431. doi:10.1177/1073858413504465

Fink, A., Grabner, R. H., Neuper, C., & Neubauer, A. C. (2005). Extraversion and cortical activation during memory performance. *International Journal of Psychophysiology*, *56*, 129–141. doi:10.1016/j.ijpsycho.2004.11.002

Fink, A., Graif, B., & Neubauer, A. C. (2009). Brain correlates underlying creative thinking: EEG alpha activity in professional vs. novice dancers. *NeuroImage*, *46*, 854–862. doi:10.1016/j.neuroimage.2009.02.036

Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journal of Gerontology*, *49*, 81-89. doi:10.1093/geronj/49.2.P81

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*, 152-170. doi:10.1111/j.1469-8986.2007.00602.x

Ford, J., Roth, W., Mohs, R., Hopkins, W., & Kopell, B. (1979). Event-related potentials from young and old adults during a memory retrieval task. *Electroencephalography and Clinical Neurophysiology, 47*, 450-459.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*, 237–241. doi:10.1111/j.1467-9280.1996.tb00366.x

Galton, F. (1908). *Memories of my life*. London: Methuen & Co.

Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's matrices. *Journal of Intelligence, 3*, 21-40.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory-scanning tasks. *Physiological Psychology*, *4*, 61–65.

Grabner, R. H., Fink, A., Stipacek, A., Neuper, C., & Neubauer, A. C. (2004). Intelligence and working memory systems: Evidence of neural efficiency in alpha band ERD. *Cognitive Brain Research*, *20*, 212–225. doi:10.1016/j.cogbrainres.2004.02.010

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013a). Individual alpha peak frequency is related to latent factors of general cognitive abilities. *NeuroImage*, *79*, 10–18. doi:10.1016/j.neuroimage.2013.04.059

Grandy, T. H., Werkle-Bergner, M., Chicherio, C., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2013b). Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults. *Psychophysiology*, *50*, 570–582. doi:10.1111/psyp.12043

Hagemann, D., Naumann, E., Lürken, A., Becker, G., Maier, S., & Bartussek, D. (1999). EEG asymmetry, dispositional mood and personality. *Personality and Individual Differences*, *27*, 541–568. doi:10.1016/S0191-8869(98)00263-3

Hale, S., & Jansen, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science*, *5*, 384–389. doi:10.1111/j.1467-9280.1994.tb00290.x

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*, 1339-1350. doi:10.1016/S1388-2457(02)00161-X

Hewig, J., Hagemann, D., Seifert, J., Naumann, E., & Bartussek, D. (2006). The relation of cortical activity and BIS/BAS on the trait level. *Biological Psychology*, *71*, 42–53. doi:10.1016/j.biopsycho.2005.01.006

Hick, W. E. (1952). On the rate of gain of information. *The Quarterly Journal of Experimental Psychology*, *4*, 11–26. doi:10.1080/17470215208416600

Holm, A., Ranta-aho, P. O., Sallinen, M., Karjalainen, P. A., & Müller, K. (2006). Relationship of P300 single-trial responses with reaction time and preceding stimulus sequence. *International Journal of Psychophysiology*, *61*, 244–252. doi:10.1016/j.ijpsycho.2005.10.015

Houlihan, M., Stelmack, R., & Campbell, K. (1998). Intelligence and the effects of perceptual processing demands, task difficulty and processing speed on P300, reaction time and movement time. *Intelligence*, *26*, 9–25. doi:10.1016/S0160-2896(99)80049-X

Hunt, E. (1983). On the nature of intelligence. *Science*, *219*, 141–146. doi:10.1126/science.6849125

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23. doi:10.1037/a0014694

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 [etc.]: Elsevier.

Johnson, W., Nijenhuis, J. t., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, *36*, 81–95. doi:10.1016/j.intell.2007.06.001

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224. doi:10.1146/annurev.psych.59.103006.093615

Kelly, A. M. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage*, *39*, 527–537. doi:10.1016/j.neuroimage.2007.08.008

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*, 486–507. doi:10.1177/0049124114543236

Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika, 74*, 21-48. doi:10.1007/s11336-008-9075-y

Klein Entink, R. H., Kuhn, J., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, *14*, 54-75. doi:10.1037/a0014877

Kline, P., Draycott, S. G., & McAndrew, V. M. (1994). Reconstructing intelligence: A factor analytic study of the BIP. *Personality and Individual Differences*, *16*, 529–536. doi:10.1016/0191-8869(94)90180-5

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence*, *15*, 397–422. doi:10.1016/0160-2896(91)90003-V

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795. doi:10.1126/science.887923

Kyllonen, P. C. (1985). *Dimensions of information processing speed*. Brooks Air Force Base, TX: Air Force Systems Command, AFHRL-TP-84-56.

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

Lerche, V., & Voss, A. (submitted). The Ratcliff diffusion model as diagnostic tool: Retest reliability of the parameters.

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

Liston, C., Watts, R., Tottenham, N., Davidson, M. C., Niogi, S., Ulug, A. M., & Casey, B. J. (2006). Frontostriatal Microstructure Modulates Efficient Recruitment of Cognitive Control. *Cerebral Cortex*, *16*, 553–560. doi:10.1093/cercor/bhj003

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

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*, 291-308. doi:10.1111/j.1469-8986.1994.tb02218.x

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. doi:10.1037/1082-989X.1.2.130

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*(2), 95–107. doi:10.1037/h0056029

Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford, New York: Oxford University Press.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science*, *211*, 77–80. doi:10.1126/science.7444452

McGarry-Roberts, P. A., Stelmack, R. M., & Campbell, K. B. (1992). Intelligence, reaction time, and event-related potentials. *Intelligence*, *16*, 289–313. doi:10.1016/0160-2896(92)90011-F

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., & Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Journal of cognitive neuroscience*, *15*, 394–408. doi:10.1162/089892903321593117

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:10.1037/a0025896

Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, *17*, 803–832. doi:10.1016/0191-8869(94)90049-3

Miller, L. T., & Vernon, P. A. (1996). Intelligence, reaction time, and working memory in 4- to 6-year-old children. *Intelligence*, *22*, 155–190. doi:10.1016/S0160-2896(96)90014-8

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *141*, 788–798. doi:10.1037/a0027968

Myerson, J., & Hale, S. (1993). General slowing and age invariance in cognitive processing: The other side of the coin. In J. Cerella, J. M. Rybash, W. Hoyer, M. L. Commons, J., Cerella, J. M., Rybash, J. M., . . . Commons, M. L. (Eds.), *Adult information processing: Limits on loss* (pp. 115–141). San Diego, CA, US: Academic Press.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, *24*, 283–288. doi:10.1016/S0166-2236(00)01790-2

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J.,. . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*, 77–101. doi:10.1037/0003-066X.51.2.77

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

Neubauer, A. C., Spinath, F. M., Riemann, R., Borkenau, P., & Angleitner, A. (2000). Genetic and environmental influences on two measures of speed of information processing and their relation to psychometric intelligence: Evidence from the German Observational Study of Adult Twins. *Intelligence*, *28*, 267–289. doi:10.1016/S0160-2896(00)00036-2

Nunez, M. D., Srinivasan, R., & Vandekerckhove, J. (2015). Individual differences in attention influence perceptual decision making. *Frontiers in psychology*, *8*, 18. doi:10.3389/fpsyg.2015.00018

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*, 10–45. doi:10.3758/s13423-010-0020-6

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*, 15–30. doi:10.1016/j.jml.2014.02.003

O'Connor, T. A., & Burns, N. R. (2003). Inspection time and general speed of processing. *Personality and Individual Differences*, *35*, 713–724. doi:10.1016/S0191-8869(02)00264-7

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781. doi:10.1037/0033-295X.105.4.761-781

Peak, H., & Boring, E. G. (1926). The factor of speed in intelligence. *Journal of Experimental Psychology*, *9*, 71–94. doi:10.1037/h0071020

Pelosi, L., Holly, M., Slade, T., Hayward, M., Barrett, G., & Blumhardt, L. D. (1992). Event-related potential (ERP) correlates of performance of intelligence tests. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *84*, 515–520. doi:10.1016/0168-5597(92)90040-I

Penke, L., Maniega, S. M., Bastin, M. E., Valdés Hernández, M. C., Murray, C., Royle, N. A.,. . . Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular psychiatry*, *17*, 1026–1030. doi:10.1038/mp.2012.66

Penke, L., Muñoz Maniega, S., Murray, C., Gow, A. J., Hernández, M. C. V., Clayden, J. D.,. . . Deary, I. J. (2010). A general factor of brain white matter integrity predicts information processing speed in healthy older people. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *30*, 7569–7574. doi:10.1523/JNEUROSCI.1553-10.2010

Pesta, B. J., McDaniel, M. A., & Bertsch, S. (2010). Toward an index of well-being for the fifty U.S. states. *Intelligence*, *38*, 160–168. doi:10.1016/j.intell.2009.09.006

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, *118*, 2128–2148. doi:10.1016/j.clinph.2007.04.019

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, *74*, 392–409. doi:10.1037/h0024913

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

Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin & Review*, *7*, 1–25. doi:10.3758/BF03210723

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

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

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

Raven, J. C., Court, J. H., & Raven, J. (1994). *Manual for Raven's progressive matrices and mill hill vocabulary scales. Advanced progressive matrices*. Oxford: Oxford University Press.

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences*, *11*, 1–120. doi:10.1016/S1041-6080(00)80007-2

Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *11*, 616–622.

Roznowski, M., & Smith, M. L. (1993). A note on some psychometric properties of Sternberg task performance: Modifications to content. *Intelligence*, *17*, 389–398. doi:10.1016/0160-2896(93)90006-Q

Russo, P. M., de Pascalis, V., Varriale, V., & Barratt, E. S. (2008). Impulsivity, intelligence and p300 wave: An empirical study. *International Journal of Psychophysiology*, *69*, 112–118. doi:10.1016/j.ijpsycho.2008.03.008

Saccuzzo, D. P., Larson, G. E., & Rimland, B. (1986). *Speed of information processing and individual differences in intelligence.* Research Report NPRDC TR 86-23. San Diego, CA: Navy Personnel Research and Development Center.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of personality and social psychology*, *86*, 162–173. doi:10.1037/0022-3514.86.1.162

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

Schulter, G., & Neubauer, A. (2005). Zentralnervensystem und Persönlichkeit. In J. Henning & P. Netter (Eds.), *Biopsychologische Grundlagen der Persönlichkeit* (pp. 35-190). München: Elsevier.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*. 6, 461–464.

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

Spearman, C. (1904). 'General intelligence,' objectively determined and measured. *The American Journal of Psychology*, *15*, 201–293. doi:10.2307/1412107

Spearman, C. (1923). *The nature of 'intelligence' and the principles of cognition*. Oxford, England: Macmillan.

Spironelli, C., & Angrilli, A. (2009). Developmental aspects of automatic word processing: Language lateralization of early ERP components in children, young adults and middle-aged subjects. *Biological Psychology, 80*, 35-45. doi:10.1016/j.biopsycho.2008.01.012

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*, 850–886. doi:10.1037/a0033981

Steiger, J. H., & Lind, J. (1980). *Statistically-based tests for the number of common factors.* Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge, New York: Cambridge University Press.

Sternberg, S. (1969). Memory-scanning: mental processes revealed by reaction-time experiments. *American Scientist*. 57, 421–457.

Steyer, R., Ferring, D., & Schmitt, M. J. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment*, *8*, 79–98.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits—Revised. *Annual Review of Clinical Psychology*, *11*, 71–98. doi:10.1146/annurev-clinpsy-032813-153719

Troche, S. J., Indermühle, R., Leuthold, H., & Rammsayer, T. H. (2015). Intelligence and the psychological refractory period: A lateralized readiness potential study. *Intelligence*, *53*, 138–144. doi:10.1016/j.intell.2015.10.003

Vandekerckhove, J. (2015, July). *Cognitive latent variable models.* Talk presented at the 48th Annual Meeting of the Society for Mathematical Psychology, Newport Beach, USA.

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–1026. doi:10.3758/BF03193087

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*, 61–72. doi:10.3758/BRM.40.1.61

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of P3. *Behavioral and Brain Sciences*, *11*, 343–356. doi:10.1017/S0140525X00058015

Vestergaard, M., Madsen, K. S., Baaré, W. F. C., Skimminge, A., Ejersbo, L. R., Ramsøy, T. Z., . . . Jernigan, T. L. (2011). White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children. *Journal of cognitive neuroscience*, *23*, 2135–2146. doi:10.1162/jocn.2010.21592

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*, 767–775. doi:10.3758/BF03192967

Wagenmakers, E.-J., van der Maas, Han L. J., & Grasman, Raoul P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22. doi:10.3758/BF03194023

Wai, J. (2014). Experts are born, then made: Combining prospective and retrospective longitudinal data shows that cognitive ability matters. *Intelligence*, *45*, 74–80. doi:10.1016/j.intell.2013.08.009

Wake, H., Lee, P. R., & Fields, R. D. (2011). Control of local protein synthesis and initial events in myelination by action potentials. *Science*, *333*, 1647–1651. doi:10.1126/science.1206998

Wechsler, D. (1944). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins Co.

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature neuroscience*, *9*, 971–978. doi:10.1038/nn1727

Wilhelm, O. (2015). *Special Issue "Mental Speed and Response Times in Cognitive Tests"*. Retrieved from http://www.mdpi.com/journal/jintelligence/special_issues/mentalspeed

Williams, L. M., Simms, E., Clark, C. R., Paul, R. H., Rowe, D., & Gordon, E. (2005). The test-retest reliability of a standardized neurocognitive and neurophysiological test battery: 'Neuromarker'. *International Journal of Neuroscience*, *115*, 1605–1630. doi:10.1080/00207450590958475

Wissler, C. (1901). The correlation of mental and physical tests. *The Psychological Review: Monograph Supplements*, *3*, i-62. doi:10.1037/h0092995

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

Wundt, W. M. (1908-1911*). Grundzüge der physiologischen Psychologie. Band 1-3* (6th ed.). Engelman: Leipzig.

Yakovlev, P. I., & Lecours A.-R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowski (Ed.). *Regional development of the brain in early life* (pp. 3-70). Oxford, UK: Blackwell Scientific.

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

List of tables

List of figures

List of abbreviations

AIC	Akaike Information Criterion
APA	American Psychological Association
APM	Raven's Advanced Progressive Matrices
BIC	Bayes Information Criterion
BIS	Berlin intelligence structure test
C	Canonical correlation coefficient
CFI	Comparative Fit Index
CRT	Choice reaction time task
DT	Decision time
DTSD	Standard deviation of the decision time
ε_{ij}	Error residual of a manifest variable measured at situation *i* with method *j*
ECT	Elementary cognitive task
EEG	Electroencephalogram
ERP	Event-related potential
g	General intelligence
GOF	Goodness-of-fit
H_0	Null hypothesis
H_1	Alternative hypothesis
ICC	Intraclass correlation coefficient
IQ	Intelligence quotient
IT	Inspection time
LED	Light-emitting diode
LST	Latent state-trait
M_j	Variance of the method redisual *j*
MT	Movement time
MTSD	Standard deviation of the movement time
OMO	Odd-man-out task

PCA	Principal component analysis
RMSEA	Root mean square error of approximation
RT	Response time
SEM	Structural equation model
SOA	Stimulus onset asynchrony
SR_i	Variance of the state residual i
SRT	Simple reaction time
T	Trait
t_0	Non-decision time
v	Drift rate
Y_{ij}	Variance of the variable Y

Appendix A1 – Manuscript 1

Decomposing the Relationship between Mental Speed and Mental Abilities

Anna-Lena Schubert, Dirk Hagemann, Andreas Voss, Andrea Schankin, and Katharina Bergmann

University of Heidelberg, Heidelberg, Germany

Author Note

Anna-Lena Schubert, Institute of Psychology, University of Heidelberg; Dirk Hagemann, Institute of Psychology, University of Heidelberg; Andreas Voss, Institute of Psychology, University of Heidelberg; Andrea Schankin, Institute of Psychology, University of Heidelberg; Katharina Bergmann, Institute of Psychology, University of Heidelberg.

Correspondence concerning this article should be addressed to Anna-Lena Schubert, University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, Phone: +49 (0) 6221-547354, Fax: +49 (0) 6221-547325, E-mail: anna-lena.schubert@psychologie.uni-heidelberg.de

Abstract

It is unclear whether different elementary cognitive tasks (ECTs) are associated with intelligence because these tasks tap the same basic cognitive process (suggesting a single mental speed factor) or different ones (suggesting several mental speed factors), as it is not known which specific cognitive processes are measured in ECTs and because the factor structure of these processes is unknown. To address these questions, 40 participants (50% males) between 18 and 75 years drawn from a community sample completed the Hick paradigm, the Sternberg memory scanning paradigm, and the Posner letter matching paradigm while an EEG was recorded. We applied a diffusion model approach to the response-time data, which allows the mathematical decomposition of different cognitive parameters involved in speeded binary decisions. Behavioral and electrophysiological results indicated that ECT conditions varied in different neuro-cognitive components of information processing. Further analyses revealed that all speed and latency variables had substantial loadings on a second-order general factor marked by general intelligence, and that the association between ERP latencies and general intelligence was mediated by reaction times. These results suggest that there is a general neuro-cognitive speed factor across different tasks and different levels of measurement that is associated with general intelligence.

Decomposing the Relationship between Mental Speed and Mental Abilities

After several decades of research, there is ample evidence of a moderate, but very consistent association between measures of intelligence and measures of mental speed. In a recent review of 172 studies, Sheppard and Vernon (2008) reported an average correlation of r = -.24 between different measures of intelligence and a variety of mental speed measures. This evidence indicates that more intelligent individuals have a higher speed of information processing. It is not yet known, however, if this association is driven by a general mental speed factor across different cognitive functions (e.g., information uptake, short-term memory, lexical access) or if there are several mental speed factors that are specific for cognitive functions and that are independently associated with general intelligence.

The aim of the present study was to address this question and to provide a rationale for a more refined analysis of the relationship between mental abilities and mental speed that may allow for a better understanding of the neuro-cognitive processes driving this association.

The study of mental speed

Almost all studies on the relationship between mental abilities and mental speed employ so-called elementary cognitive tasks (ECTs) when measuring reaction times (for a notable exception using pencil-and-paper tests see Neubauer & Knorr, 1998). These ECTs are tasks with very low cognitive demands that maximize the empirical control of task complexity and minimize unwanted sources of variance in individual differences. Because ECTs put only marginal cognitive requirements on participants, individual differences in strategy use and in previous experience with specific elements of the task are less likely to influence the association between RTs and intelligence than in more complex decision-making problems. Several of the often-used ECTs follow an idea in tradition of Donder’s *subtraction method* (Donders, 1969): The *subtraction method* presumes that when two reaction time tasks differ only in the number of stimulus or response alternatives while every other detail of the task remains the same across conditions, the difference between RTs is an indicator of a purely mental processing speed. Following this logic, difference parameters are often the theoretically most interesting variables in ECT research. There are several paradigms in which this idea is pursued.

In the *simple and choice reaction time task* based on the Hick paradigm (Hick, 1952), participants are presented between one and ten response buttons arranged in a semi-circle around a single home button and have to react when the light next to one of the response buttons is switched on. Because Hick showed that there is a linear relationship between the amount of information that has to be processed and reaction times (Hick, 1952), individual intercept and slope parameters can be computed when regressing RTs on the logarithm of stimulus-response alternatives. This way, individual slope parameters can be used as estimates for the "rate of gain of information" (Roth, 1964), which are theoretically (though seldom statistically) independent of motoric movement time, and can be correlated with measures of mental abilities. Another application of the general idea of the subtraction method can be found in the *Sternberg memory scanning task* (Sternberg, 1969). In this task, participants see memory sets of different sizes and are then asked if a single probe item was part of the previously presented memory set. Because RTs again increase linearly with memory set size, the slope parameter of the regression of RTs on memory set size can be used as an indicator of individual speed of short-term memory search. A similar idea is applied in the *letter matching paradigm* (Posner & Mitchell, 1967) where participants have to decide whether two letters are the same in accordance with their physical identity or in accordance with their name. The difference of RTs between these conditions is an estimate for the speed of lexical access (Hunt, 1983), because of the additionally required access to long-term memory in the name identity condition.

Associations between mental speed and mental abilities

Correlations between RTs of ECTs and mental abilities are moderate, but consistent. Jensen (1987) reviewed 26 studies with a total *N* of 2317 participants that investigated the relationship between different parameters of the Hick paradigm and mental abilities tests. He reported a multiple R^2 of .25 in a regression of IQ scores on different parameters derived from the Hick paradigm. In a review of ten studies using Sternberg's memory scanning task and psychometric intelligence tests, Neubauer (1997) reported a mean correlation of $r = -.27$ between mean RT and intelligence test scores. He also reviewed ten studies correlating RTs in the Posner letter matching task and mental abilities test scores and computed mean *N*-weighted correlations ranging between r = -.23 and -.33 for different parameters of the paradigm. In a recent review, Jensen (2006) reported canonical correlations ranging from $C =$ .55 to .72 between different measures of mental abilities and of mental speed. It should be noted that correlations including the difference measures and slope parameters are usually substantially lower (Jensen, 1998; Neubauer, 1997). Taken together, these results suggest that there is a consistent negative association between mental speed and mental abilities in the way that more intelligent individuals have a higher speed of information processing.

Cognitive processes in elementary cognitive tasks

The general idea that ECTs measure specific cognitive processes like *speed of short-term memory access* or *information processing speed* is appealing, because correlations between difference and slope parameters in ECTs and general intelligence would then be informative about the association between specific cognitive processes and general intelligence. This general idea should, however, be treated with caution. Although ECTs already have rather low task complexities, each ECT still requires several cognitive processes such as attention, perception, encoding, representation in working memory, decision making, and response preparation. Moreover, it can be argued that ECT conditions might differ in the demands they put on several cognitive processes simultaneously, so that difference and slope parameters might not only be indicators of a specific cognitive process, but might also include variance of other cognitive processes that differ between conditions. This would violate one of the assumptions of the subtraction method proposed by Donders (1969) and question the validity of difference and slope parameters. Because this often-implicated premise has to our knowledge never been tested empirically, the first aim of the present study was to investigate whether conditions in three ECTs differ only in one or in several cognitive processes.

Because not much is empirically known about which specific cognitive processes contribute to the distribution of reaction times in ECTs, even less is known about the origins of inter-individual differences in these RTs. One important question is whether these different tasks are related to general intelligence because they tap the *same* basic property of the cognitive system, or whether these tasks tap *different* cognitive system parameters. Many researchers tend to conclude from these findings that there is indeed one basic property at work, which is mental speed. According to this view, greater mental speed facilitates a better cognitive performance. Despite the great theoretical relevance of this concept, only few studies provided data that may help to answer the question whether there is one general factor of mental speed. Most studies include only one or two elementary cognitive tasks and are not focused on a systematic study of the factor structure itself. There are a few studies that report correlation matrices or factor analyses of ECTs that favor the hypothesis of a large general mental speed factor explaining more than 40% of variance (Burns & Nettelbeck, 2003; Hale & Jansen, 1994; Neubauer & Bucik, 1996; Neubauer, Spinath, Riemann, Borkenau, & Angleitner, 2000), while other studies, which employ not only classical ECTs but a more diverse range of information-processing tasks, report multi-factorial models of mental speed (O'Connor & Burns, 2003; Roberts & Stankov, 1999). Clearly these inconsistent results require further systematic study of the factor structure of mental speed, although the preliminary findings may suggest that there is a general mental speed factor, probably in addition to more task-specific speed factors. The second aim of the present study was to address this question by decomposing the information-processing components in three ECTs and testing whether a single general mental speed factor emerges in a factor analysis of different speed measures across the three tasks.

As long as we do not have enough knowledge about the factor structure of ECTs, we cannot know which cognitive processes might be responsible for individual differences in RTs. Therefore, we do not know whether more intelligent individuals have a generally faster speed of information processing or whether they differ in very specific facets of mental speed from less intelligent individuals. The behavioral data do not inform us which of these processes differ between individuals of different cognitive ability. The third aim of the present study was to address this problem using methods that allow the decomposition of the stream of information processing during reaction time tasks and to analyze the association between individual differences in these distinct information processing components and mental abilities.

Decomposing the stream of information processing in ECTs

In the present study, we used two methods to decompose the stream of information processing in ECTs: The first method is the *diffusion model*, which decomposes the stream of information-processing and decision making in RT tasks into distinct components based on RT distributions (Ratcliff, 1978). Another method that decomposes the stream of neuro-cognitive information processing are electrophysiological measures, namely *event-related potentials (ERPs)*, which allow to identify functionally distinct components in different time windows between the stimulus onset and the response execution. While diffusion models have only recently been applied in mental abilities research, ERPs are already used to a great extent.

Diffusion models are random walk-models used in the context of speeded binary decisions and provide a framework for analyzing the whole distribution of reaction time data (for recent reviews, see Ratcliff & McKoon, 2008, Wagenmakers, 2009; Voss, Nagler, & Lerche, 2013). They allow the identification of cognitive parameters by fitting predicted reaction time-distributions to empiric reaction time-distributions (Voss, Rothermund, & Voss, 2004). Diffusion models in their most basic form identify four distinct parameters: The first parameter, *drift rate (v)*, describes the strength of the systematic influence on the diffusion process with larger drift rates causing shorter reaction times and smaller amounts of errors. This parameter is most akin to the idea of 'speed of information processing' mentioned earlier, as it indicates the amount of information gathered per time unit. The second parameter, *boundary separation (a)*, is a measure for the distance between decision thresholds, i.e., an indicator for the conservatism of the decision criterion. The third parameter, *starting value (z)*, indicates whether a person is biased towards one of two decision thresholds. If z is closer to one threshold than the other, this threshold is reached more often due to random fluctuations, resulting in more and faster decisions associated with this threshold. The last parameter, *response-time constant* (t_0), encompasses processes unrelated to decision making, mainly stimulus encoding and response execution.

There are only a very small number of studies in which diffusion models were applied in intelligence research. In a study by Schmiedek, Oberauer, Wilhelm, Süß, and Wittmann (2007), university students had to complete several reasoning tasks and choice reaction tasks. They showed that a latent drift rate factor correlated positively with a latent reasoning ability factor (r = .79), whereas they reported a smaller negative association between a latent boundary separation factor (r = -.48) and reasoning ability. Ratcliff, Thapar, and McKoon (2010) asked participants in three different age groups (18-25, 60-74, 75-90 years) to complete different categorization tasks. They reported correlations ranging from r = .36 to .90 for the three age groups between a latent drift rate factor and intelligence, whereas they found no consistent association between other diffusion model parameters and intelligence. They found similar results in another study, where participants' drift rate in recognition tasks was the only diffusion model parameter consistently correlated with intelligence, r = .47 to .67 (Ratcliff, Thapar, & McKoon, 2011). Although these preliminary results are promising, it should be noted that none of these studies used ECTs that are normally used in intelligence research.

Another method suited to decompose cognitive components in the stream of information processing is the ERP. The ERP methodology allows identifying functionally distinct electrophysiological components (e.g. the N200 or P300) that might be affected differently by condition differences in ECTs. Moreover, according to the mental speed hypothesis, the latencies of ERP components should be negatively correlated with intelligence.

There are several electrophysiological studies that correlated ERP parameters with intelligence. In their review of 23 of these studies (N > 2400), Schulter and Neubauer (2005) concluded that there are no consistent associations between ERP latencies and intelligence. It should, however, be noted that most of these studies employed standard ERP paradigms such as the oddball paradigm and that behavioral data from these tasks is uncorrelated with intelligence. There are only a few studies in which classical ECTs were combined with ERP methodology. Houlihan, Stelmack, and Campbell (1998) and Pelosi et al. (1992) computed ERPs to probe stimuli in the Sternberg memory scanning tasks and found both weak and mostly insignificant associations between ERP latencies and intelligence test scores. McGarry-Roberts, Stelmack, and Campbell (1992) computed a factor analysis of P300 latencies recorded during six reaction time tasks including the Sternberg memory scanning task. They correlated this P300 factor with a general intelligence factor and reported a correlation of r = -.36 between these factors. All in all, these studies suggest that there may be a weak negative association between ERP latencies and mental abilities, but further studies are needed before any final conclusions can be drawn.

The present study

The goal of the present study was to decompose the information-processing components in three different ECTs (Hick paradigm, Sternberg memory scanning task, Posner letter matching task) by applying diffusion models to reaction time distributions and by monitoring the neuro-cognitive correlates of information processing with EEG methodology. We pursued three aims: First, we wanted to investigate whether differences between ECT conditions represent one or multiple cognitive processes, as the general idea of ECT implies that these differences represent a single process within each task. Contrary to this idea, we expected ECT conditions to represent a range of different processes such as attention, perception, encoding, representation in working memory, decision making, and response preparation, i.e. we anticipated that these tasks differ in several behavioral and electrophysiological parameters simultaneously. Our second aim was to investigate the factor structure of mental speed. We expected to identify a single general mental speed factor across all behavioral and electrophysiological measures and all tasks in addition to more specific factors. Our third aim was to investigate the association between mental speed and mental abilities across the different measures and tasks. We expected a) that a general mental speed factor is significantly associated with general intelligence, and b) that the association between ERP latencies and mental abilities is mediated by reaction times. This mediation model is based on the methodological framework of Baron and Kelley (1986), who suggested that mediation models are causal models. A proposed mediator variable Z mediates the relationship between an independent variable X and an outcome variable Y only if the independent variable has a causal effect on the mediator variable that in turn has a causal effect on the outcome variable (Baron & Kenny, 1986, p. 1176). While the mediation model allows for some part of the causal influence to take the direct path from the independent variable to the outcome variable (X->Y), it presumes that a substantial part of the causal influence is exerted through the indirect effect via the proposed mediator (X->Y->Z). In all ECTs of the present study, a stimulus has first to be processed visually and then relayed to frontal areas associated with executive functions and decision making before a motor response reflecting this decision can be initiated. Thus, there is a stream of processing that has some temporal order, with neuro-cognitive events taking place before behavioral events occur. Therefore, we expected that ERP latencies exert the majority of their influence on general intelligence indirectly through the proposed mediator reaction times.

Method

Participants

We recruited a sample of $N = 40$ participants (20 females, 20 males) between 18 and 75 years old ($M = 47.4$, $SD = 15.6$) from different educational and occupational backgrounds via local newspaper advertisement. All participants had normal or corrected to normal vision and no history of mental illness. They received 10€ as payment for their participation and could indicate whether they wanted to be informed about their personal results.

Measures

Elementary cognitive tasks

Hick paradigm. In order to control for visual attention effects, response bias effects, and top-down strategies associated with the classical Jensen apparatus and the use of a home button (Longstreth, 1984) and in order to ensure compatibility of this paradigm with EEG measurements, we adopted the modified Hick paradigm developed by Neubauer, Bauer, and Höller (1992). This modified paradigm is presented on a computer screen and does not employ a home button. Participants' middle and index fingers rested on four keys of a modified keyboard, on which all other keys irrelevant to the task were removed. Those keys were positioned directly underneath the squares relevant for the task, thus increasing stimulus-response compatibility as much as possible. Participants were instructed to always keep their fingers on the keys during the whole task. In the 2 bit condition, four squares arranged in a row with a fixation cross in their middle were shown on the screen for a time period varying between 1000 and 1500 ms. After this period, a cross appeared in one of the four squares and participants had to press the corresponding response-key. The screen remained unchanged for 1000 ms following the response, as we wanted to record post-decisional neuronal processes. After this time period, an ITI varying between 1000 and 1500ms was presented, followed by the next trial.

We implemented two 1 bit conditions: One condition (comparability low: 1 bit_{CL}) adopted from Neubauer et al. (1992) and a second one (comparability high: 1 bit_{CH}) designed to maximize stimulus comparability with the 2 bit condition. At the beginning of each trial in the 1 bit_{CL} condition, only two squares appeared on the screen with a fixation cross in the middle of the screen. These two squares appeared pseudo-randomly in two of the four locations used in the 2 bit condition. As in the 2 bit condition, a cross appeared in one of the two squares after 1000 to 1500 ms and participants had to press the corresponding key. In the 1 bit_{CH} condition, however, all four squares were presented on the screen, but participants were instructed to only pay attention to two of them, because the cross could only appear in one of these two squares. There were four blocks with 20 items each instructing participants to pay attention to the left/right/middle/outer two squares. We implemented this additional 1 bit condition because it shared all stimulus features with the 2 bit condition and only differed from this condition in the instruction participants were given. This is a necessary prerequisite for ruling out confounds in the interpretation of ERP effects, because small changes in physical stimulus features can result in sizeable changes in ERP amplitudes. For an overview over the different conditions, see Figure 1.

Participants were instructed to respond as quickly and accurately as possible. The order of conditions was the same for all participants. First they completed the 2 bit condition, then the 1 bit_{CL} condition followed by 1 bit_{CH} condition. Each condition consisted of a learning phase with ten sample items and direct feedback, followed by 80 test items. Participants made short breaks between blocks. There were two fixed sequences of the location of squares and crosses that were balanced across participants.

Sternberg memory scanning task. Participants were shown digits between 0 and 9 on a computer screen. There were three blocks of ten sample items each with feedback and 80 test items with a memory set size of 1, 3, and 5 digits. Each trial began with a fixation cross varying between 1000 and 1500 ms. Digits were presented sequentially for 1000ms with a blank screen of 400 to 600 ms between single digits. After the last digit 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 had to press one of two keys with their index fingers indicating whether the digit was part of the memory set seen immediately before. The probe item remained on screen for 1000 ms after the reaction was made and the intertrial interval was 1000 to 1500 ms. See Figure 1 for illustration.

All participants began with set size one and then progressed to set sizes three and five. They were given the option to make short breaks between blocks. There were two versions of stimulus material counterbalanced across participants. The probe item was previously presented in the memory set in 50% of the trials. The position of keys indicating whether the probe item was part of the memory set was counterbalanced across participants.

Posner letter matching task. After a fixation cross lasting between 1000 and 1500 ms, two letters were presented in the middle of the screen and participants had to decide whether this pair was identical or not by pressing the corresponding key. In the physical identity condition, participants were instructed to judge letters as identical only if they were identical in physical characteristics (thus, “AA” would be identical, while “Aa” or “AB” would be judged as different). In the name identity condition, participants were instructed to judge the name identity of stimuli (thus, “AA” and “Aa” would be judged as identical, while “AB” would not be). Afterwards, the pair of letters remained on the screen for 1000 ms and was followed by an ITI varying between 1000 and 1500 ms. See Figure 1 for illustration.

The two conditions were separated into blocks of 10 sample items with feedback and 200 test items each. All participants began with the physical identity condition and made a short break between blocks. There were two versions of stimulus material assorted to participants depending on their number. We used the upper- and lowercase letters A, B, F, H, and Q as stimulus material. The pair of letters was identical in 50% of the trials. The position of keys indicating whether the pair was identical was counterbalanced across participants.

Please insert Figure 1 about here

Intelligence tests

Fluid intelligence. We used a self-programmed computerized version of Raven’s Advanced Progressive Matrices (APM; Raven, Court, & Raven, 1994) to measure fluid intelligence. In this computer adapted version of the APM, one item was presented at a time with its eight possible alternatives and participants had to indicate their solution with a mouse click. They were able to go back and forth between the different items at any time with the exception that they could not go back to Item set I once they had started working on Item Set II. According to the test manual, the APM raw test score was computed as the number of correctly solved items and used in all further analyses. Cronbach’s alpha was $\alpha = .82$.

Crystallized intelligence. We constructed a short version of the knowledge test from the German Intelligenz-Struktur-Test 2000-R (IST 2000-R; Liepmann, Beauducel, Brocke, & Amthauer, 2001) as an indicator of crystallized intelligence. The knowledge test of the IST 2000-R consists of several knowledge questions tapping different fields of knowledge like "What does π (pi) mean?", "In which street is the New Yorker stock exchange?", or "What does the octane index indicate?". To create a short version, we chose the 20 items with the highest loadings on crystallized intelligence, although we lost some diversity in the content of test items. The knowledge test was administered according to the manual and the number of correctly solved items was used as the test score for all further analyses. We did not compute IQ scores because we had no normative data of our abbreviated version. Cronbach's alpha was $\alpha = .65$.

Procedure

Participants read and signed an informed consent. They were seated on a comfortable chair in a dimly lit EEG cabin in front of a computer screen. All participants completed the three ECTs in the same order with small breaks between the tasks: Hick paradigm, Sternberg memory scanning task, and Posner letter matching task. ECTs were followed by a short break, after which participants completed the APM and the knowledge test based on the IST 2000-R. Information about demographic variables was gathered at the end of the session.

EEG recording

The EEG was recorded with nine Ag-AgCl electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4) that were positioned according to the international 10-20 system. We used the aFz electrode as the ground electrode. Electrodes were initially referenced to Cz and later re-referenced to linked mastoids (TP9, TP10). To correct for ocular artifacts, we recorded the electrooculogram (EOG) bipolarly with two electrodes positioned above and below the right eye and two electrodes positioned at the outer canthi of the eyes. All electrode impedances were kept below 5kΩ. The EEG was recorded continuously with a sampling rate of 2500 Hz (band-pass 0.1-100 Hz). We applied an offline low-pass filter of 16 Hz for the determination of average activity within time windows and low-pass filters of 12 Hz (early ERPs) and 8 Hz (late ERPs) for latency detection.

Data analysis

Response times

Trials with extremely fast RTs (< 200 ms for the Hick paradigm and < 300 ms for the Sternberg memory scanning and the Posner letter matching task) or extremely slow RTs (> 3000 ms) were removed. We used the *fast-dm* program developed by Voss and Voss (2007) to fit diffusion models to RT distributions, which is free software that utilizes the Kolmogorov-Smirnov test statistic to estimate model parameters. The parameter z for mean starting point was set equal to $a/2$, presuming that participants had no response bias towards the correct or incorrect alternative. We computed separate diffusion models for each condition of the three ECTs in which the parameters a, v, t_0, and s_{t0} were allowed to vary freely. For correlational analyses, we averaged the respective parameters across all conditions for each ECT. Thus, we computed an average drift and an average response-time constant for each of the three ECTs in order to reduce the number of variables entered into the subsequent factor analysis while increasing their reliability. Intertrial-variability parameters of the diffusion model (s_v, s_z) were fixed to 0 to keep the model parsimonious with the exception of s_{t0}, because it led to a notable improvement of model fit.

To assess how well diffusion models fit the individual-level data, we conducted Monte-Carlo simulations and simulated 1000 data sets from the diffusion model that matched the characteristics of the empirical data (i.e., simulated parameter values were based on empirical parameters values, and the number of trials and conditions was equivalent to the tasks used in the present study). We then re-analyzed the simulated data sets with the diffusion model and used the 5% quantile of the distribution of fit-values in each ECT condition as the critical value to assess individual model fit in the empirical models.

EEG parameters

We calculated ERPs time-locked to the onset of probe items in all tasks, using the preceding 200 ms as baseline and including an interval from 200 ms before stimulus onset until 1000 ms afterward. Ocular artifacts were corrected using the regression procedure of Gratton, Coles, and Donchin (1983). Epochs with amplitudes exceeding ±70 µV, with amplitude changes exceeding 100 µV within 100 ms, or with lower activity than 0.5 µV were discarded as artifacts. We identified ERP components by visual inspection of the grand average waveforms (figure 3-5) for the three ECTs and computed mean EEG activity in the following time windows: In the Hick paradigm, we computed the P200 (175-215 ms), the N200 (210-240 ms), and the P300 (360-420 ms). In the Sternberg memory scanning paradigm, we computed the N150 (115-160 ms), the P200 (200-245 ms), the N300 (300-360 ms), and the P300 (400-600 ms). In the Posner letter matching paradigm, we computed the N140 (115-155 ms), the P210 (190-235 ms), the N300 (240-365 ms), and the P300 (465-580 ms). For ERP latencies, we inspected participants' individual averaged waveforms at all nine electrode positions for peaks during the time windows described above and used these peak latencies as individual latency measures. For correlational analyses, we inspected grand average waveforms and determined at which electrode position each ERP component was greatest and used the corresponding electrode position for all further analyses. We used the same electrode position for each ERP component for all participants.

Statistical analyses

In order to characterize ECTs in terms of information processing components, we ran repeated measures ANOVAs with the factor condition separately for median RTs, drift rates, and response-time constants for each ECT. In the following analyses on average EEG activity, we ran an omnibus repeated-measures ANOVA for each ECT with four within-subject factors: ERP component (with three levels for the Hick paradigm: P200, N200, P300; four levels for the Sternberg memory scanning task: N150, P200, N300, P300; with four levels for the Posner letter matching task: N140, P210, N300, P300), condition (with two levels for the Hick paradigm: 1 bit vs. 2 bit; three levels for the Sternberg memory scanning task: set size 1, set size 3, set size 5; two levels for the Posner letter matching task: PI vs. NI), caudality (with three levels for all tasks: frontal, central, parietal), and laterality (with three levels for all tasks: left, central, right) in order to test if condition effects differed between time windows. We then ran follow-up repeated measures ANOVAs for each ECT with the three within-subject factors condition, caudality, and laterality to test for condition differences in specific ERP components in each ECT. For these analyses, we dropped the fourth factor ERP component that was included in the omnibus ANOVAs, because we wanted to test for condition differences in specific time frames.

For factor analyses, we first computed principal component analyses (PCA) a) for intelligence test scores and b) for each of six time-domain variables across the three ECTs (Table 8 shows a list of variables). We included only time-domain variables that were available and comparable in at least two different ECTs, which led to the exclusion of slower ERP components because their time windows were not comparable across ECTs. Next, we computed individual factor scores of the first principal component of these seven PCAs that yielded individual factor scores for RT, v, t_0, and three ERP latencies. A hierarchical PCA was then run on the correlation matrix of these factor scores and the number of factors was determined by the scree plot (Cattell, 1966) and the parallel analysis criterion (Horn, 1965). Because of their intercorrelations, factors were rotated obliquely.

Finally, we ran mediation analyses to test whether the relationship between ERP latencies on intelligence test scores was mediated by reaction times and used the bootstrap procedure to test for indirect effect (Preacher & Hayes, 2004).

We repeated all correlational analyses controlling for age because of the heterogeneous sample. Since age had no influence on the pattern of results, we did not include these analyses in this report.

Results

Descriptive data

The median RT in the Hick paradigm was M = 447.22 (SD = 91.62) and the mean accuracy was M = 0.98 (SD = .01). In the Sternberg memory scanning paradigm, the median RT was M = 736.78 (SD = 133.17) and the mean accuracy was M = 0.96 (SD = .02). The median RT in the Posner letter matching task was M = 663.41 (SD = 104.89) and the mean accuracy was M = 0.98 (SD = .01). Please consult Table 1 for the descriptive data of the different ECT conditions. The mean score of correctly solved APM items was M = 24.55 (SD = 5.09), which corresponds to a mean IQ of M = 91.68 (SD = 14.6). IQ scores were normally distributed (skew = 0.16, kurtosis = -0.24) and ranged from 78 to 123 IQ points. The mean score of correctly solved knowledge test items was M = 15.49 (SD = 2.72). No corresponding IQ score could be computed, because we only used an abbreviated version of the full IST 2000-R knowledge test. Knowledge test scores were also normally distributed (skew = -0.17, kurtosis = -0.49).

Please insert Table 1 about here

Descriptive statistics for the diffusion model parameters are shown in Table 1. Model fits were satisfactory for all three ECTs. In the Hick paradigm, 5% of the models in the 1 bit condition and 2.5% of the models in the 2 bit condition had *p*-values smaller than the critical *p*-values of p_{crit} = .794 and .843, respectively. In the set size 1 condition of the Sternberg memory scanning paradigm, there were no models with *p*-values below the critical value of p_{crit} = .834, while 2.5% and 7.5% of the models in the set size 3 and in the set size 5 condition had *p*-values smaller than p_{crit} = .839 and p_{crit} = .836. Model fits were slightly worse in the Posner letter matching paradigm with 10% of the models falling short of the critical *p*-value p_{crit} = .833 in the physical identity condition and 5% of the models falling short of the critical *p*-value p_{crit} = .824 in the name identity condition. We kept the models with bad model fits in our analyses, because excluding these models did not change the pattern of results.

Characterization of ECTs in terms of neuro-cognitive processing

One aim of this study was to identify neuro-cognitive parameters differing between conditions of the three ECTs. The main purpose of the analyses reported in this section was to test whether there are specific RT and ERP parameters that differ between conditions or whether we have to assume that ECT conditions differ in several steps in the course of neuro-cognitive information processing.

RT characterization and diffusion model results of ECTs

As expected, median RTs increased with increasing task difficulty in all ECTs (Figure 2). *In the Hick paradigm*, median RTs were significantly larger in the 2 bit than in the 1 bit_{CH}

condition, $F(1,38) = 92.73$, $p < .001$, $\omega^2 = .71$. In the 1 bit_{CL} condition, however, median RTs were significantly larger than in the 2 bit condition, $F(1,38) = 4.62$, $p = .038$, $\omega^2 = .09$, which was unexpected as less information (only two alternatives) had to be processed in the 1 bit_{CL} than in the 2 bit condition (four alternatives). As we did not know which cognitive processes were responsible for this unexpected increase in RTs, we dropped the 1 bit_{CL} condition from all further analyses and renamed the "1 bit_{CH}" condition to "1 bit" condition for the remainder of this report. When we analyzed the effects of condition on drift rates and response-time constants separately, we observed no change in drift rates with increasing number of stimulus alternatives, $F(1,38) = 2.02$, $p = .163$, $\omega^2 = .03$, but an increase in response-time constants, $F(1,38) = 58.62$, $p < .001$, $\omega^2 = .60$.

In the Sternberg memory scanning paradigm, RTs increased with set size, $F(2,78) = 113.17$, $p < .001$, $\omega^2 = .74$, $\varepsilon = .68$, in a strictly linear way, $F(1,39) = 133.53$, $p < .001$, $\omega^2 = .77$, for the linear trend of effect size. As expected, drift rates decreased with memory set size, $F(2,76) = 18.47$, $p < .001$, $\omega^2 = .31$, $\varepsilon = .91$, also following a linear pattern, $F(1,38) = 31.16$, $p < .001$, $\omega^2 = .44$. t_0 also differed between conditions, $F(2,76) = 35.57$, $p < .001$, $\omega^2 = .48$, $\varepsilon = .82$, and increased linearly with memory set size, $F(1,38) = 46.76$, $p < .001$, $\omega^2 = .55$.

In the Posner letter matching paradigm, median RTs were higher in the name identity than in the physical identity condition, $F(1,38) = 70.36$, $p < .001$, $\omega^2 = .64$. When v and t_0 were compared between conditions, we found that drift rates decreased in the NI condition, $F(1,38) = 35.76$, $p < .001$, $\omega^2 = .48$, and that there was no significant difference in response-time constants between conditions, $F(1,38) = 2.64$, $p = .112$, $\omega^2 = .04$. Overall, these results indicated that there was substantial variation between tasks in which diffusion model parameters varied significantly between ECT conditions. Only in the Sternberg memory scanning paradigm did different conditions differ not only in their information processing demands, but also in their sensomotoric difficulties.

Please insert Figure 2 about here

ERP characterization of ECTs

In order to investigate whether electrophysiological activity differed between conditions within each of the three ECTs, we compared average activity and peak latencies across different time windows in the course of information processing. Our main aim was not to identify specific processes differing between conditions, but to test if ECT conditions differed in only one or several electrophysiological components. We only reported main effect and interactions including the factor ECT condition, as we were only interested in condition effects on ERP; additional topographical information on the ERP components can be found in the tables detailing the complete ANOVA results in the supplementary material.

In the Hick paradigm, we compared mean activity and peak latencies between conditions in three different time windows. First, we computed an omnibus ANOVA with the four within-subject factors ERP component (P200: 175-215 ms, N200: 210-240 ms, P300: 360-420 ms), condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) to test whether condition effects differed between time windows. For mean activity, we observed a significant interaction between component and condition, $F(2,70) = 20.26$, $\varepsilon = .71$, $p < .001$, $\omega^2 = .55$, as well as significant three-way interactions between ERP component, condition and caudality, $F(4,140) = 5.12$, $\varepsilon = .39$, $p = .014$, $\omega^2 = .11$, and between ERP component, condition and laterality, $F(4,140) = 3.16$, $\varepsilon = .53$, $p = .045$, $\omega^2 = .06$. For ERP latencies we observed a significant interaction between ERP component and caudality, $F(4,140) = 3.96$, $\varepsilon = .64$, $p = .032$, $\omega^2 = .08$, and a significant three-way interaction between ERP component, condition and caudality, $F(4,140) = 8.96$, $\varepsilon = .56$, $p < .001$, $\omega^2 = .18$. See Figure 3 for the ERPs elicited by stimuli in the Hick paradigm.

Please insert Table 2 and 3 about here

In a next step, we compared mean activity and ERP latencies between conditions in each of the three different time windows. Please see Table 2 for detailed results of the main effects and interactions including the factor condition on ERP amplitudes and Table 3 for detailed results on ERP peak latencies. We found a significant difference in mean P200 and N200 activity with amplitudes being greater in the 1 bit than in the 2 bit condition for the P200, $\omega^2 = .41$, and with amplitudes being greater in the 2 bit than in the 1 bit condition for the N200, $\omega^2 = .42$. The significant interactions between condition and caudality, $\omega^2 = .25$, and between condition and laterality, $\omega^2 = .22$ and .30, indicated a specific topography of this effect. In particular, condition differences were largest at central and central parietal electrode sites for both ERP components. Moreover, P200 latencies were shorter in the 2 bit than in the 1 bit condition, $\omega^2 = .27$, but the significant interaction between condition and caudality, $\omega^2 = .17$, suggested that this was mostly the case for frontal electrode sites, as P200 latencies were slightly shorter in the 1 bit than in the 2 bit condition at parietal electrode sites, $F(1,35) = 4.26$, $p = .046$, $\omega^2 = .09$. Furthermore, we found significant interactions between condition and caudality, $\omega^2 = .11$, between condition and laterality, $\omega^2 = .08$, and between condition, caudality and laterality, $\omega^2 = .11$, for the N200 peak latencies. These interactions indicated that N200 latencies were shorter in the 2 bit condition than in the 1 bit condition at Fz and F4, $F(1,35) = 6.78$, $p = .013$, $\omega^2 = .14$, and marginally larger in the 2 bit than in the 1 bit condition at Cz and Pz, $F(1,35) = 3.14$, $p = .085$, $\omega^2 = .06$.

We observed no main effect of condition on average activity in the P300 component, $\omega^2 = .00$. The significant interactions (see Table 2c)) between condition and caudality, $\omega^2 = .25$, and between condition and laterality, $\omega^2 = .13$, indicated that amplitudes in the 1 bit were greater than in the 2 bit condition at central electrode sites, $F(1,35) = 4.36$, $p = .044$, $\omega^2 = .09$, and tended to be smaller at frontal electrode sites in comparison to the 2 bit condition, $F(1,35) = 3.58$, $p = .067$, $\omega^2 = .07$. Moreover, condition effects could only be observed at central and left electrode sites. For P300 latencies we found a pattern of results that again indicated that P300 were marginally shorter in the 2 bit than in the 1 bit condition at frontal electrode sites, $F(1,35) = 3.44$, $p = .072$, $\omega^2 = .07$, and shorter in the 1 bit than in the 2 bit condition at parietal electrode sites, $F(1,35) = 5.20$, $p = .029$, $\omega^2 = .11$. It should be noted that mean RTs in both condition were close to the P300 time window (391 and 467 ms) and might therefore account for condition effects in terms of differently timed response preparation processes.

Together, the topography effects described for the mean activity in each ERP time window and the significant interactions of the omnibus analysis suggested that conditions in the Hick paradigm differ in several electrophysiological components of information processing. For ERP latencies, however, there seemed to be a caudality-specific pattern of results that is consistent across all ERP components and that suggests that condition differences in ERP latencies are not specific for ERP components.

Please insert Figure 3 about here

In the Sternberg memory scanning task, we compared mean activity and ERP latencies between conditions in four different time windows. First, we computed another omnibus analysis following the previously described logic with the four within-subject factors ERP component (N150: 115-160 ms, P200: 200-245 ms, N300: 300-360 ms, P300: 400-600 ms), condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) to test whether condition effects differed between time windows. We observed a significant interaction between ERP component and condition on average activity, $F(6,228) = 8.52$, $\varepsilon = .59$, $p < .001$, $\omega^2 = .17$, as well as significant three-way interactions between ERP component, condition, and caudality, $F(12,456) = 18.92$, $\varepsilon = .36$, $p < .001$, $\omega^2 = .32$, and between ERP component, condition, and laterality, $F(12,456) = 3.79$, $\varepsilon = .23$, $p = .015$, $\omega^2 = .07$. For ERP latencies, we observed a significant main effect of condition, $F(2,76) = 8.87$, $\varepsilon = .91$, $p = .001$, $\omega^2 = .17$, and a significant interaction between condition and ERP component, $F(6,228) = 4.52$, $\varepsilon = .37$, $p = .011$, $\omega^2 = .08$. See Figure 4 for the ERPs elicited by stimuli in the Sternberg memory scanning paradigm.

Please insert Table 4 and 5 about here

Please see Table 4 for detailed results of the ANOVAs on the average activity for the specific time windows and Table 5 for the detailed results on peak latencies. We found no significant main effects or interactions including condition on the amplitudes or peak latencies of the N150 and P200 component, all Fs < 3.15, all ps > .065, all ω^2s < .06 (see Table 4 and Table 5 a) and b)).

We observed a significant main effect of condition on average N300 activity, $\omega^2 = .26$, with greater amplitudes in the set size 3 and set size 5 conditions than in the set size 1 condition, $F(1,38) = 26.31$, $p < .001$, $\omega^2 = .40$, and no difference between amplitudes in the more difficult conditions, $F < 1$. Moreover, a significant interaction between condition and caudality indicated a specific topography of this effect, $\omega^2 = .15$. The condition effects were greatest at central and parietal electrode sites. We also observed a significant interaction between condition and caudality for N300 peak latencies, $\omega^2 = .06$, indicating that N300 latencies became longer with increasing memory set size at central and parietal electrodes, $F(2,76) = 5.07$, $\varepsilon = .97$, $p = .009$, $\omega^2 = 10$.

Next, we compared conditions and electrode sites for P300 activity. We observed a main effect of condition, $\omega^2 = .36$, that indicated that P300 amplitudes decreased linearly with increasing memory set size, $F(1,38) = 30.84$, $p < .001$, $\omega^2 = .44$. There was also a significant interaction between condition and caudality, $\omega^2 = .31$, as P300 amplitudes only increased at central, $F(2,76) = 15.85$, $\varepsilon = .92$, $p < .001$, $\omega^2 = .28$, and parietal electrode sites with increasing memory set size, $F(2,76) = 49.44$, $\varepsilon = .90$, $p < .001$, $\omega^2 = .56$, but not at frontal electrode sites, $F(2,76) < 1$, $\varepsilon = .80$, $p = .856$, $\omega^2 = .00$. Moreover, P300 peak latencies became longer with increasing memory set size.

The results of these analyses indicated that conditions differ systematically in average activity and suggest together with the specific topographic interactions for each ERP that the neural processing of probe items in different memory set sizes differs in more than one electrophysiological component.

Please insert Figure 4 about here

In the Posner letter matching paradigm, we compared mean activity and ERP latencies between conditions in four different time windows. Again, we first computed an omnibus ANOVA with the four within-subject factors ERP component (N140: 115-155 ms, P210: 190-235 ms, N300: 240-365 ms, P300: 465-580 ms), condition (PI vs. NI), caudality (frontal, central, parietal), and laterality (left, central, right) to test if condition effects differed between ERPs. The effect of the interaction between ERP component and condition on average activity was marginally significant, $F(3,99) = 3.36$, $\varepsilon = .36$, $p = .072$, $\omega^2 = .06$. We also observed a significant three-way interaction between ERP component, condition and caudality, $F(6,198) = 5.05$, $\varepsilon = .36$, $p < .01$, $\omega^2 = .11$, and a marginally significant three-way interaction between ERP component, condition and laterality, $F(6,198) = 2.55$, $\varepsilon = .44$, $p = .068$, $\omega^2 = .04$. There was no significant main effect or interaction including condition on ERP peak latencies. See Figure 5 for the ERPs elicited by stimuli in the Posner letter matching paradigm.

Please insert Table 6 and 7 about here

Next, we computed several ANOVAs for the different time windows in the Posner letter matching paradigm. Please see Table 6 and Table 7 for all main effects and interactions including the factor condition. There were no significant main effects of condition on ERP amplitudes, all ω^2s $< .05$, but several interactions between condition and caudality and condition and laterality. Condition differences were most pronounced at frontal electrode sites for the N140, P210 and N300 component. These interactions were not further unraveled, as follow-up tests of condition differences at frontal electrode sites yielded no significant effects, all Fs < 1.2, all ps $> .282$, all $\omega^2 = .00$. For the P300 component, we observed a specific topography of condition effects that indicated that P300 amplitudes were greater in the PI than in the NI condition and that this effect was greatest at central electrode sites, $\omega^2 = .07$. Moreover, the significant three-way interaction suggested that condition differences were greatest at Cz, $\omega^2 = .05$. As in the overall analyses, there were no main effects or interactions including condition on any of the ERP peak latencies.

Again, the topography differences between condition differences in ERPs and the significant interactions in the omnibus analysis indicated that the PI and NI condition differ in more than one ERP component. These differences were only manifest in average activity, but not in peak latencies.

Please insert Figure 5 about here

Factor structure of mental speed

In the next step, we analyzed the factor structure of mental speed and its relation to general intelligence. In order to investigate the factor structure of mental speed, we computed six principal component analyses separately for the six time-domain variables (RT, v, t_0, P100 latency, N150 latency, P200 latency) across the three ECTs. We then computed individual component scores of the first principal component for all participants to generate six new variables that capture the greatest amount of variance in each set of time-domain variables across ECTs. We repeated this procedure for intelligence test scores and extracted a general intelligence factor. Table 8 shows the variables entered into each PCA and the amount of variance explained by the respective first principal component. We then entered these seven component score variables into further analyses to explore the factor structure of mental speed. Correlations between these seven component scores are shown in Table 9. If the factor structure of mental speed is unitary, all component score variables should load onto one mental speed variable that should have a great eigenvalue and explain a substantial amount of variance in speed and latency parameters.

Please insert Table 8 and Table 9 about here

To explore this idea, we conducted a hierarchical PCA of the six time-domain component scores and identified two components explaining 76% of variance based on the scree plot (Cattell, 1966) and the parallel analysis criterion (Horn, 1965). These two components had eigenvalues of 3.32 and 1.21. Component loadings after an oblique rotation of the two factors are shown in Table 10. All ERP latencies loaded strongly onto the first rotated component that was also loaded by drift rates. All behavioral time domain component scores loaded more strongly on the second component that was marked by RT component scores. Because greater (i.e., slower) ERP latencies were associated with greater component scores in the first component, we reversed the polarity of the first component so that greater component scores indicated smaller (i.e., faster) ERP latencies. Subsequently, we labeled the two components ‘processing speed’ and ‘behavioral speed’, respectively. The two components were correlated, $r = .36$.

Please insert Table 10 about here

We extracted individual component scores for these two hierarchical components and computed a hierarchical second-order PCA of these two components and the intelligence component scores. Correlations between the three variables were $r = .54$, $p < .001$ (*g* and processing speed), $r = .52$, $p = .001$ (*g* and behavioral speed), and $r = .36$, $p = .028$ (processing and behavioral speed). The PCA of these correlates yielded a single second-order component based on the scree plot (Cattell, 1966) and the parallel analysis criterion (Horn, 1965) with an eigenvalue of 1.98 onto which all hierarchical first-order components loaded (see Table 11 for factor loadings). This component explained 66% of variance in first-order factor scores. *g* component scores had the greatest loadings on this component, followed by neural and behavioral speed with highly similar loadings.

Please insert Table 11 about here

In the last step, we computed correlations between the two speed components and APM and knowledge test scores to investigate whether correlations were greater for g*f* or g*c*. Correlations were generally greater for *gf* than for *gc*: Correlations between APM scores and speed components ranged from r = .53 to .54, while correlations between knowledge test scores and speed components ranged from r = .35 to .39.

The effects of latencies on *g* are mediated by RTs

Next, we analyzed if reaction times mediate the relationship between ERP latencies and intelligence test scores. For all mediation analyses, we used the component scores we computed in the PCA.

As Figure 6 illustrates, the relationship between ERP latencies and intelligence was mediated by reaction times. A bootstrap analysis with m = 5000 resamples yielded a significant indirect effect of P100 latencies through RTs on intelligence test scores, CI 95% (-0.44, -0.01). We found that RTs also partially mediated the effect of N150 latencies on intelligence test scores. We observed a significant indirect effect when we computed a bootstrap analysis with m = 5000 resamples, CI 95% (-0.41, -0.02). Lastly, we tested if RTs also mediated the effect of P200 latencies on *g*. Again, the bootstrap analysis with $m = 5000$ resamples indicated a significant indirect effect, CI 95% (-0.43, -0.04).

Please insert Figure 6 about here

Discussion

The present study sheds light on the neuro-cognitive processes of mental speed. The results indicate that so-called elementary cognitive tasks (ECTs) are not as elementary as presumed but that they tap several functionally different neuro-cognitive processes. As expected, we found that there is no unitary construct of mental speed, but that there are several distinct speeded processes involved in elementary cognitive tasks. Moreover, our results show that an increase in the difficulty and complexity of these ECTs affects several of these processes simultaneously. If we consider, for example, our results for the Sternberg memory scanning paradigm, we see that conditions in this task differed in several behavioral and electrophysiological parameters. As expected, diffusion model analyses revealed that drift rates decreased and response-time constants increased with increasing memory set size, which indicates that conditions differ both in the speed of decision making (reflected in the changes in the v parameter) and in the speed of encoding, memory access, and/or in the speed of movement times (reflected in the changes in the t_0 parameter). Moreover, changes in memory set size also had an effect on several ERP components in the stream of information processing, namely the N300 component, which is associated with spatial, structural and categorical incongruences of visual stimuli (Demiral, Malcolm, & Henderson, 2012; Hamm, Johnson, & Kirk, 2002), and the P300 component, which is associated with stimulus evaluation and categorization and is known to be sensitive to stimulus probability, subjective uncertainty and resource allocation (Luck, 2005). All in all, we can conclude that the traditional difference and slope measures of ECTs do not only capture variance from a single cognitive process, but that they reflect a multitude of different processes.

Condition differences in ERP amplitudes and latencies were mostly consistent with previous research on these tasks, although there are only few studies with comparable designs. In the Hick paradigm, we observed significantly greater P200 amplitudes for the 1 bit than for the 2 bit condition, which is consistent with the results reported by Falkenstein, Hohnsbein, and Hoormann (1994) who analyzed ERPs in 2- and 4-choice RT tasks and found that P200 amplitudes were greater in the 2-choice than in 4-choice condition. Moreover, they reported that P390 amplitudes were greater in the 2-choice than in the 4-choice condition for all electrode sites, while we found this effect only at central electrode sites and observed a reversed effect at frontal electrode sites. McGarry-Roberts, Stelmack, and Campbell (1992) reported greater P300 amplitudes in a choice reaction time task than in a simple reaction time task, which may indicate that the more complex RT task resulted in greater P300 amplitudes. As McGarry-Roberts et al. (1992) only used a 2-choice CRT and no 4-choice CRT and only entered the Pz electrode into the statistical analyses, their results are not directly comparable to our results that showed a very specific topography. In the present study, we found an effect of choice alternatives on P300 latencies with a specific topography in the way that P300 latencies were larger for the 2 bit than for the 1 bit condition at parietal electrode sites, while this effect was reversed at frontal electrode sites. Falkenstein et al. (1994) found a similar effect with a very specific topography as the P390 component peaked later in the 4-choice than in the 2-choice task at Pz, and McGarry-Roberts et al. (1992) also reported longer P300 latencies for the CRT task in comparison to the SRT task at Pz. There was also a latency shift in the N200 peak reported by Falkenstein et al. (1994), but it did not display the specific topography effects of the present study.

In the Sternberg memory scanning paradigm, we found that P300 amplitudes decreased and P300 latencies increased with increasing memory set size which is consistent with the majority of the studies analyzing the electrophysiological activity in this paradigm (Brookhuis, Mulder, Mulder, & Gloerich, 1983; Ford, Roth, Mohs, Hopkins, & Kopell, 1979; Gomer, Spicuzza, & O'Donnell, 1979; Houlihan et al., 1998; Pelosi, Hayward, & Blumhardt, 1998), although some studies found no difference in P300 latencies across conditions (Pelosi, Hally, Slade, Hayward, Barrett, & Blumhardt, 1992) or substantial interindividual differences in condition effects on P300 latencies (Pelosi, Hayward, & Blumhardt, 1995).

To our knowledge, there are no previous EEG-studies specifically aimed at analyzing the Posner letter matching paradigm. McGarry-Roberts et al. (1992) used a comparable paired-stimuli task, in which two words were presented subsequently and participants had to decide whether the target stimulus was a) physically or b) semantically the same (i.e., a synonym) as the prior presented first stimulus. Please note that the experimental setup (presenting subsequent instead of parallel stimuli) as well as the stimulus material (words instead of letters) and the depth of semantic processing (word meaning instead of letter identification) varied substantially from the present study. Still, the authors reported greater P300 amplitudes to the target stimulus for the physical similarity task than for the semantic similarity task, which is consistent with the result of the present study as we also found greater P300 amplitudes in the physical identity condition than in the name identity condition. McGarry-Roberts et al. (1992) also reported longer P300 latencies in the semantic similarity task than in the physical similarity task, while we found no latency shift in the data. This discrepancy may be due to a multitude of reasons as their paradigms varied substantially from ours.

Nearly all of these studies analyzed a smaller number of time windows and fewer ERP components than the present study and generally focused on one or two major components (often the P300). Therefore, it is not possible to relate our results for all time windows to previous research, as the stream of information processing during ECTs has not yet been analyzed electrophysiologically in such detail. Moreover, in several of these previous studies only very few electrodes were used, often only the midline electrodes (Fz, Cz, Pz), making it difficult to compare condition effects with a specific topography to these studies, as in many of the previous studies condition effects were only analyzed at one electrode (e.g., Pz for the P300 component) and the topographic characteristics of condition effects were not considered.

Furthermore, we could show that a single broad general mental speed factor is substantially associated with general intelligence, because a second order factor analysis of two more specific speed factors and general intelligence yielded a single broad factor marked by general intelligence. Thus our results indicate that although there are several functionally distinct processes involved in ECTs, it is the common time-domain variance shared by all these components that is associated with general intelligence. This does not imply that more specific speed components do not share unique variance with intelligence, but it does imply that the association between mental speed and mental abilities could in most part be due to a single shared source of variance. This result is consistent with the few studies that reported associations between RTs in different elementary cognitive tasks (Burns & Nettelbeck, 2003; Hale & Jansen, 1994; Neubauer & Bucik, 1996; Neubauer et al., 2000). In his reanalysis of the reaction time data reported by Kranzler and Jensen (1991), Carroll (1991) also found a broad general factor in addition to narrower task-specific factors with substantial variable loadings reflecting *decision time* (in contrast to an orthogonal factor of *movement time*). In our study, movement times (captured in the t_0 parameter) did not span a distinct factor, but loaded onto the behavioral speed factor that showed substantial loadings on the second-order mental speed factor. One difference between the movement speed factor in Carroll's (1991) reanalysis and movement speed measured as t_0 might be that t_0 does not only capture movement speed, but also stimulus encoding and memory-related processes (Ratcliff &

McKoon, 2008), which might be more closely related to a general mental speed factor. Taken together, our findings suggest that there is indeed a single broad mental speed factor across different tasks and across both behavioral and electrophysiological measurements, a general factor that is significantly associated with general intelligence.

The associations between RTs, ERP latencies and general intelligence in this study are substantially greater than the initially quoted average correlation of r = -.24 between RTs and mental abilities in Sheppard's and Vernon's (2008) review or the weak negative associations between ERP latencies and intelligence reported in the literature (Houlihan et al., 1998; McGarry-Roberts, Stelmack, & Campbell, 1992; Pelosi et al., 1992; Schulter & Neubauer, 2005). There may be two reasons why the magnitude of the associations in the present study is greater than in the literature: Jensen (2006) argued that characteristics of the participant sample may affect the size of the association between RTs and mental abilities. We deliberately recruited a heterogeneous community sample to avoid any restriction in the variance of the cognitive variables, because a lack of variation in one or more variables may decrease the co-variation between variables. Moreover, the number of trials used in the three paradigms was higher than most trial numbers in the literature, which may have increased the reliability of the ERP latencies that are known to sometimes have low to moderate reliabilities even when the number of trials is relatively large (Cassidy, Robertson, & O'Connell, 2012).

What is intriguing about our findings is that the association between ERP latencies and mental abilities was mediated by reaction times. This mediation supports the hypothesis that individual differences in psychophysiological information processing speed are manifested in behavioral reaction times and may in this way contribute to individual difference in general intelligence. In other words, individual differences in the onset of early ERP components such as the P100 or P200, which are components that occur nearly immediately after stimulus presentation in the chronometry of neurocognitive information-processing, predict individual difference in reaction times that occur about half a second later. This result suggests that higher speed of neurocognitive information-processing may contribute to advantages in the speed of cognitive information-processing, decision, and memory processes. These advantages in the speed of different cognitive processes may then enhance performance on psychometric intelligence tests and explain the association between early ERP latencies and mental abilities.

Limitations

Some limitations have to be considered before strong conclusions may be drawn. First, the sample size with $N = 40$ participants is rather large for an electrophysiological study, but it is too small for complex multivariate analyses such as multiple regression and structural equation modeling. Moreover, the stability of the factor structure we extracted has to be replicated in further studies before drawing any final conclusions, although our results are generally consistent with earlier studies on the factor structure of RTs in elementary cognitive tasks.

Second, it is unclear if ERP latencies and diffusion model parameters show enough stability over measurement occasions to qualify as trait-like variables. There are no systematic studies on the stability of diffusion model parameters except for one study that reported between-session stabilities of $r \geq .65$ for drift rate and non-decision time parameters in a lexical decision task (Yap, Balota, Sibley, & Ratcliff, 2012). A first study on the temporal stability of ERP components reported strong test-retest correlations for ERP amplitudes ranging from $r = .63$ to .89 and varying test-retest correlations for ERP latencies ranging from $r = .19$ to .89 (Cassidy et al., 2012). Both measures can only explain inter-individual differences in intelligence if they show sufficient psychometric stability.

Third, the RTs of the 1 bit condition we adopted from Neubauer et al. (1992) did not follow Hick's law, but were instead significantly larger than the RTs in the 2 bit condition. We therefore did not include behavioral and electrophysiological data from this condition in further analyses. Still, this divergence from the data reported by Neubauer et al. (1992) is surprising. The standard deviation in this condition was twice as large as the standard deviations in the other conditions, which indicates a great increase in difficulty or complexity. Moreover, individual differences in the understanding of the rather complex instructions of the task or in the ability to adapt to position changes might have affected RTs to a great degree. It should be noted that the sample in the original study by Neubauer et al. (1992) consisted only of children (11 to 15 years old) who got feedback immediately after each trial. Therefore, either the age difference or the direct feedback might explain why no such phenomenon was reported in the original study. A thorough validation of the modified paradigm with several control conditions would be needed to understand which cognitive processes are involved in the strangely behaving original 1 bit condition.

Conclusion

The aim of the present study was to decompose the stream of information processing in elementary cognitive tasks in order to identify processes that might contribute to the association between mental abilities and mental speed. By combining diffusion model analysis with ERP methodology, we showed that ECT conditions differ in several neuro-cognitive parameters. Therefore, we would not recommend the use of difference scores in further studies, not only because they suffer from severe psychometric problems such as low reliabilities (Jensen, 1998), but also because they do not seem to measure what they are supposed to. According to our results, difference parameters are likely to capture several different sources of variance and are probably *not* singling out specific cognitive processes such as the speed of information uptake. Instead, we propose using diffusion models and electroencephalography in order to single out specific components of information processing for further analyses.

Future studies should include several measurement occasions to test whether a general mental speed factor qualifies as a trait-like variable. Only a factor that captures a certain amount of trait-like performance is suited to be considered as an explanation of general intelligence. Moreover, future studies should also include a broader battery of intelligence tests to investigate if mental speed is more strongly associated with general intelligence (as our data would suggest) or with specific mental abilities, which we could not test in the present study.

Our study is one of the few studies that reported consistent negative correlations between ERP latencies and intelligence across different tasks and different time windows. In contrast to most other studies, we recruited a community sample in order to avoid restricted variance in the cognitive performance variables and their electrophysiological correlates. Moreover, each of our tasks had a large number of trials to increase the reliability of the notoriously unreliable ERP latencies. We could show that there is a general mental speed factor across different tasks and different measures of speed that is associated with general intelligence. Moreover, we found that the association between ERP latencies and intelligence is mediated by reaction times. These results illustrate the benefits of the application of diffusion models and ERPs in research on the chronometry of mental abilities. All in all, we found that more intelligent individuals do not only show faster behavioral reactions, but that they have a general advantage in all neuro-cognitive speed-related processes.

References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182. doi: 10.1037/0022-3514.51.6.1173

Brookhuis, K. A., Mulder, G. G., Mulder, L. J., & Gloerich, A. B. (1983). The P3 complex as an index of information processing: The effects of response probability. *Biological Psychology, 17*, 277-296. doi:10.1016/0301-0511(83)90004-2

Burns, N. R., & Nettelbeck, T. (2003). Inspection time in the structure of cognitive abilities. *Intelligence*, *31*, 237–255. doi:10.1016/S0160-2896(02)00120-4

Carroll, J. B. (1991). No demonstration that g is not unitary, but there's more to the story: Comment on Kranzler and Jensen. *Intelligence*, *15*, 423–436. doi:10.1016/0160-2896(91)90004-W

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*, 659–664. doi:10.1037/t03589-000;

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*, 245–276. doi:10.1207/s15327906mbr0102_10

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia*, *50*, 1271–1285.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Falkenstein, M. M., Hohnsbein, J. J., & Hoormann, J. J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography & Clinical Neurophysiology: Evoked Potentials, 92*, 148-160. doi:10.1016/0168-5597(94)90055-8

Ford, J. M., Roth, W. T., Mohs, R. C., Hopkins, III, W. F., & Kopell, B. S. (1979). The effects of age on event-related potentials in a memory retrieval task. *Electroencephalography and Clinical Neurophysiology, 47*, 450-459.

Gomer, F. E., Spicuzza, R. J., & O'Donnell, R. D. (1976). Evoked potential correlates of visual item recognition during memory-scanning tasks. *Physiological Psychology, 4*, 61-65.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484. doi:10.1016/0013-4694(83)90135-9

Hale, S., & Jansen, J. (1994). Global Processing-Time Coefficients Characterize Individual and Group Differences in Cognitive Speed. *Psychological Science*, *5*, 384–389. doi:10.1111/j.1467-9280.1994.tb00290.x

Hamm, J. P., Johnsin, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*, 1339–1350.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26. doi:10.1080/17470215208416600

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185. doi:10.1007/BF02289447

Houlihan, M., Stelmack, R., & Campbell, K. (1998). Intelligence and the effects of perceptual processing demands, task difficulty and processing speed on P300, reaction time and movement time. *Intelligence*, *26*, 9–25. doi:10.1016/S0160-2896(99)80049-X

Hunt, E. (1983). On the nature of intelligence. *Science*, *219*, 141–146.

Jensen, A. R. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence.* Norwood, N.J: Ablex.

Jensen, A. R. (1998). The suppressed relationship between IQ and the reaction time slope parameter of the Hick function. *Intelligence*, *26*, 43–52. doi:10.1016/S0160-2896(99)80051-8

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences* (1st ed.). Amsterdam, Boston, London: Elsevier.

Kranzler, J. H., & Jensen, A. R. (1991). The nature of psychometric g: Unitary process or a number of independent processes? *Intelligence*, *15*, 397–422. doi:10.1016/0160-2896(91)90003-V

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer R. (2007). *Intelligenz-Struktur-Test 2000-R*. Göttingen: Hogrefe.

Longstreth, L. E. (1984). Jensen's reaction-time investigations of intelligence: A critique. *Intelligence*, *8*, 139–160.

Luck, S. J. (2005). *An Introduction To The Event-Related Potential Technique*. Cambridge, London: The MIT Press.

McGarry-Roberts, P. A., Stelmack, R. M., & Campbell, K. B. (1992). Intelligence, reaction time, and event-related potentials. *Special Issue: Biology and Intelligence*, *16*, 289–313. doi:10.1016/0160-2896(92)90011-F

Neubauer, A. C. (1997). The mental speed approach to the assessment of intelligence. In J. Kingma & W. Tomic (Eds.), *Advances in cognition and educational practice: Reflections on the concept of intelligence, Vol. 4* (pp. 149–173). US: Elsevier Science/JAI Press.

Neubauer, A. C., Bauer, C., & Höller, G. (1992). Intelligence, attention, motivation and speed-accuracy trade-off in the hick paradigm. *Personality and Individual Differences*, *13*(12), 1325–1332. doi:10.1016/0191-8869(92)90175-O

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

Neubauer, A. C., & Knorr, E. (1998). Three paper-and-pencil tests for speed of information processing: Psychometric properties and correlations with intelligence. *Intelligence*, *26*(2), 123-151.

Neubauer, A. C., Spinath, F. M., Riemann, R., Borkenau, P., & Angleitner, A. (2000). Genetic and environmental influences on two measures of speed of information processing and their relation to psychometric intelligence: Evidence from the German Observational Study of Adult Twins. *Intelligence, 28*, 267-289. doi:10.1016/S0160-2896(00)00036-2

O'Connor, T. A., & Burns, N. R. (2003). Inspection time and general speed of processing. *Personality and Individual Differences*, *35*, 713–724. doi:10.1016/S0191-8869(02)00264-7

Pelosi, L., Hayward, M. M., & Blumhardt, L. D. (1995). Is 'memory scanning' time in the Sternberg paradigm reflected in the latency of event-related potentials?. *Electroencephalography & Clinical Neurophysiology: Evoked Potentials, 96*(1), 44-55. doi:10.1016/0013-4694(94)00163-F

Pelosi, L., Hayward, M., & Blumhardt, L. (1998). Which event-related potentials reflect memory processing in a digit-probe identification task*?. Cognitive Brain Research, 6*, 205-218. doi:10.1016/S0926-6410(97)00032-3

Pelosi, L., Holly, M., Slade, T., Hayward, M., Barrett, G., & Blumhardt, L. D. (1992). Event-related potential (ERP) correlates of performance of intelligence tests. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *84*, 515–520. doi:10.1016/0168-5597(92)90040-I

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers, 36*, 717-731. doi: 10.3758/bf03206553

Posner, M. I.., & Mitchell, R. F. (1967). Chronometric Analysis of Classification. *Psychological Review*, *74*, 392–409. doi:10.1037/h0024913

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

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

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

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

Raven, J. C., Court, J. H., & Raven, J. (1994). *Manual for Raven's progressive matrices and mill hill vocabulary scales. Advanced progressive matrices*. Oxford: Oxford University Press.

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences*, *11*, 1–120. doi:10.1016/S1041-6080(00)80007-2

Roth, E. (1964). Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *11*, 616–622.

Schmiedek, F., Oberauer, K., Wilhelm, O., Suss, 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*, *136*, 414–429. doi:10.1037/0096-3445.136.3.414

Schulter, G., & Neubauer, A. C. (2005). Zentralnervensystem und Persönlichkeit. In J. Henning & P. Netter (Eds.), *Biopsychologische Grundlagen der Persönlichkeit* (pp. 35–190). München: Elsevier.

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

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction time experiments. *American Scientist*, *57*, 421–457.

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

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

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*, 767-775. doi:10.3758/BF03192967

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

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

Table 1

Median RTs, mean accuracies, mean drift rates and mean response-time constants for the different conditions in the three ECTs (SD in parentheses).

ECT	Condition	Median RT	Accuracy	v	t_0	a	s_{t0}
Hick paradigm	1 bit_{CL}	483 (172.41)	.97 (.03)	3.53 (1.57)	0.31 (0.10)	1.47 (0.56)	0.13 (0.11)
	1 bit_{CH}	380 (73.59)	1.00 (.01)	5.26 (1.34)	0.30 (0.04)	1.18 (0.33)	0.11 (0.06)
	2 bit	461 (98.73)	.99 (.02)	4.89 (1.44)	0.37 (0.06)	1.17 (0.26)	0.15 (0.08)
Sternberg memory scanning task	Set size 1	599.25 (105.60)	.96 (.03)	3.15 (0.81)	0.40 (0.08)	1.35 (0.35)	0.18 (0.13)
	Set size 3	732.75 (146.69)	.97 (.03)	2.86 (0.79)	0.51 (0.10)	1.54 (0.36)	0.21 (0.16)
	Set size 5	851.25 (187.74)	.96 (.03)	2.36 (0.72)	0.56 (0.13)	1.68 (0.41)	0.20 (0.14)
Posner letter matching task	PI	618 (96.44)	.98 (.01)	3.98 (0.94)	0.46 (0.07)	1.37 (0.35)	0.13 (0.07)
	NI	683.50 (120.78)	.98 (.02)	3.03 (0.84)	0.47 (0.08)	1.64 (0.31)	0.15 (0.14)

Note. 1 bit_{CL} = 1 bit condition with low comparability; 1 bit_{CH} = 1 bit condition with high comparability; PI = Physical Identity; NI = Name Identity; v = drift rate; t_0 = response-time constant; a = boundary separation; s_{t0} = intertrial-variability of the response-time constant.

Table 2

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on a) P200 (175-215 ms) amplitude, b) N200 (210-240 ms) amplitude, and c) P300 (360-420 ms) amplitude in the Hick paradigm. (n = 36)

ERP component	Variable	*df*	*F*	*p*	ε	ω²
a) P200	Condition	1,35	25.18	<.001	-	.41
	Condition x Caudality	2,70	12.40	<.001	.67	.25
	Condition x Laterality	2,70	11.08	<.001	.90	.22
	Condition x Caudality x Laterality	4,140	<1	.765	.66	.00
b) N200	Condition	1,35	25.89	<.001	-	.42
	Condition x Caudality	2,70	27.79	<.001	.43	.25
	Condition x Laterality	2,70	16.17	<.001	.82	.30
	Condition x Caudality x Laterality	4,140	<1	.654	.72	.00
c) P300	Condition	1,35	<1	.497	-	.00
	Condition x Caudality	2,70	12.90	<.001	.69	.25
	Condition x Laterality	2,70	8.21	<.001	.71	.13
	Condition x Caudality x Laterality	4,140	<1	.526	.59	.00

Table 3

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on a) P200 (175-215 ms) peak latencies, b) N200 (210-240 ms) peak latencies, and c) P300 (360-420 ms) peak latencies in the Hick paradigm. (n = 36)

ERP component	Variable	*df*	*F*	*p*	ε	ω^2
a) P200	Condition	1,35	13.91	.001	-	.27
	Condition x Caudality	2,70	8.42	.003	.65	.17
	Condition x Laterality	2,70	<1	.631	.93	.00
	Condition x Caudality x Laterality	4,140	<1	.932	.69	.00
b) N200	Condition	1,35	<1	.813	-	.00
	Condition x Caudality	2,70	5.35	.009	.90	.11
	Condition x Laterality	2,70	3.84	.026	.82	.08
	Condition x Caudality x Laterality	4,140	5.29	.002	.73	.11
c) P300	Condition	1,35	<1	.759	-	.00
	Condition x Caudality	2,70	7.88	.004	.64	.16
	Condition x Laterality	2,70	1.50	.233	.81	.01
	Condition x Caudality x Laterality	4,140	<1	.40	.68	.00

Table 4

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on a) N150 (115-160 ms) amplitude, b) P200 (200-245 ms) amplitude, c) N300 (300-360 ms) amplitude, and d) P300 (400-600 ms) amplitude in the Sternberg memory scanning paradigm. (n = 39)

ERP component	Variable	*df*	*F*	*p*	ε	ω^2
a) N150	Condition	2,76	<1	.95	.66	.00
	Condition x Caudality	4,152	2.08	.151	.32	.03
	Condition x Laterality	4,152	<1	.813	.39	.00
	Condition x Caudality x Laterality	8,304	1.04	.334	.16	.00
b) P200	Condition	2,76	2.13	.134	.86	.03
	Condition x Caudality	4,152	1.48	.235	.34	.01
	Condition x Laterality	4,152	2.05	.149	.37	.03
	Condition x Caudality x Laterality	8,304	1.19	.289	.15	.00
c) N300	Condition	2,76	14.41	<.001	.93	.26
	Condition x Caudality	4,152	7.46	.002	.42	.15
	Condition x Laterality	4,152	<1	.595	.55	.00
	Condition x Caudality x Laterality	8,304	1.20	.300	.19	.01
d) P300	Condition	2,76	22.33	<.001	.93	.36
	Condition x Caudality	4,152	18.32	<.001	.45	.31
	Condition x Laterality	4,152	<1	.409	.49	.00
	Condition x Caudality x Laterality	8,304	<1	.387	.23	.00

Table 5

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on a) N150 (115-160 ms) peak latencies, b) P200 (200-245 ms) peak latencies, c) N300 (300-360 ms) peak latencies, and d) P300 (400-600 ms) peak latencies in the Sternberg memory scanning paradigm. (n = 39)

ERP component	Variable	*df*	*F*	*p*	ε	ω²
a) N150	Condition	2,76	1.55	.22	.85	.01
	Condition x Caudality	4,152	<1	.758	.72	.00
	Condition x Laterality	4,152	<1	.695	.89	.00
	Condition x Caudality x Laterality	8,304	1.12	.351	.60	.00
b) P200	Condition	2,76	<1	.433	.80	.00
	Condition x Caudality	4,152	<1	.485	.51	.00
	Condition x Laterality	4,152	<1	.782	.70	.00
	Condition x Caudality x Laterality	8,304	<1	.578	.57	.00
c) N300	Condition	2,76	2.01	.143	.96	.03
	Condition x Caudality	4,152	3.37	.019	.78	.06
	Condition x Laterality	4,152	2.43	.064	.81	.04
	Condition x Caudality x Laterality	8,304	1.69	.123	.76	.02
d) P300	Condition	2,76	6.43	.005	.82	.13
	Condition x Caudality	4,152	1.84	.139	.82	.02
	Condition x Laterality	4,152	2.06	.100	.85	.03
	Condition x Caudality x Laterality	8,304	<1	.767	.76	.00

Table 6

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on a) N140 (115-155 ms) amplitude, b) P210 (190-235 ms) amplitude, c) N300 (240-365 ms) amplitude, and d) P300 (465-580 ms) amplitude in the Posner letter matching paradigm. (n = 35)

ERP component	Variable	*df*	*F*	*p*	ε	ω^2
a) N140	Condition	1,33	<1	.712	-	.00
	Condition x Caudality	2,66	2.22	.139	.62	.04
	Condition x Laterality	2,66	<1	.377	.95	.00
	Condition x Caudality x Laterality	4,132	2.68	.045	.84	0.05
b) P210	Condition	1,33	<1	.535	-	.00
	Condition x Caudality	2,66	7.37	.007	.58	.16
	Condition x Laterality	2,66	<1	.480	.77	.00
	Condition x Caudality x Laterality	4,132	2.04	.122	.67	.03
c) N300	Condition	1,33	1.15	.292	-	.00
	Condition x Caudality	2,66	5.91	.015	.61	.13
	Condition x Laterality	2,66	<1	.573	.69	.00
	Condition x Caudality x Laterality	4,132	2.79	.043	.77	.05
d) P300	Condition	1,33	2.53	.121	-	.04
	Condition x Caudality	2,66	<1	.424	.61	.00
	Condition x Laterality	2,66	3.64	.038	.88	.07
	Condition x Caudality x Laterality	4,132	2.84	.037	.83	.05

Table 7

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on a) N140 (115-155 ms) peak latencies, b) P210 (190-235 ms) peak latencies, c) N300 (240-365 ms) peak latencies, and d) P300 (465-580 ms) peak latencies in the Posner letter matching paradigm. (n = 35)

ERP component	Variable	*df*	*F*	*p*	ε	ω²
a) N140	Condition	1,33	<1	.580	-	.00
	Condition x Caudality	2,66	<1	.794	.61	.00
	Condition x Laterality	2,66	1.07	.321	.61	.00
	Condition x Caudality x Laterality	4,132	<1	.413	.62	.00
b) P210	Condition	1,33	<1	.471	-	.00
	Condition x Caudality	2,66	3.14	.066	.75	.06
	Condition x Laterality	2,66	2.69	.086	.84	.05
	Condition x Caudality x Laterality	4,132	1.41	.243	.82	.01
c) N300	Condition	1,33	1.17	.288	-	.00
	Condition x Caudality	2,66	<1	.629	.98	.00
	Condition x Laterality	2,66	1.04	.356	.94	.00
	Condition x Caudality x Laterality	4,132	1.87	.148	.65	.03
d) P300	Condition	1,33	3.01	.092	-	.06
	Condition x Caudality	2,66	<1	.443	.73	.00
	Condition x Laterality	2,66	<1	.614	.85	.00
	Condition x Caudality x Laterality	4,132	<1	.832	.80	.00

Table 8

Sources of entered variables for the six speed, latency, and intelligence variables and the amount of variance explained by the first principal components of each PCA.

Variable name	Source of entered variables	% of variance explained by first principal component
g	APM, knowledge test	78%
Median RT	all ECTs and conditions	62%
v	all ECTs, estimated across conditions	42%
t_0	all ECTs, estimated across conditions	51%
P100 latency	Hick paradigm and Sternberg memory scanning task, all conditions	37%
N150 latency	all ECTs and conditions	45%
P200 latency	Sternberg memory scanning and Posner letter matching task, all conditions	69%

Note. v = drift rate; t_0 = response-time constant.

Table 9

Product-moment correlations (rank correlation coefficients in parentheses) between the six mental speed component scores (RT, v, t0, P100 latency, N150 latency, P200 latency) and g.

ECT	RT	*v*	*t*0	P100	N150	P200	*g*
RT component scores	1						
v component scores	-.74*** (-.79***)	1					
t_0 component scores	.72*** (.65***)	-.34* (-.37*)	1				
P100 latency component scores	.53** (.38**)	-.38* (-.41*)	.11 (.02)	1			
N150 latency component scores	.50** (.35*)	-.38* (-.32*)	.19 (.06)	.54*** (.48**)	1		
P200 latency component scores	.46** (.40*)	-.46** (-.44**)	.18 (.11)	.25 (.25)	.80*** (.80***)	1	
g	-.55*** (-.56***)	.50** (.56***)	-.42* (-.38*)	-.49** (-.44**)	-.53** (-.50**)	-.45** -(.45**)	1

Note. *v* = drift rate; t_0 = response-time constant. *p < .05. **p < .01. ***p < .001.

Table 10

Component loadings for the principal component analysis with oblimin rotation of time-domain component scores.

	First-order component	
	1	2
RT component scores	-.29	-.83
v component scores	.40	.55
t_0 component scores	.20	-.94
P100 latency component scores	-.59	-.20
N150 latency component scores	-.96	.04
P200 latency component scores	-.93	.10

Note. Because greater (i.e., slower) ERP latencies were associated with greater component scores in the first component, we reversed the polarity of the first component so that greater component scores indicated smaller (i.e., faster) ERP latencies

Table 11

Component loadings for the principal component analysis of the two hierarchical mental speed factors and g.

	G
g	.87
processing speed	.80
behavioral speed	.77

Note. Lowercase *g* designates general intelligence extracted from the PCA of APM and knowledge test scores, whereas uppercase *G* is the second-order component derived from speed and intelligence components.

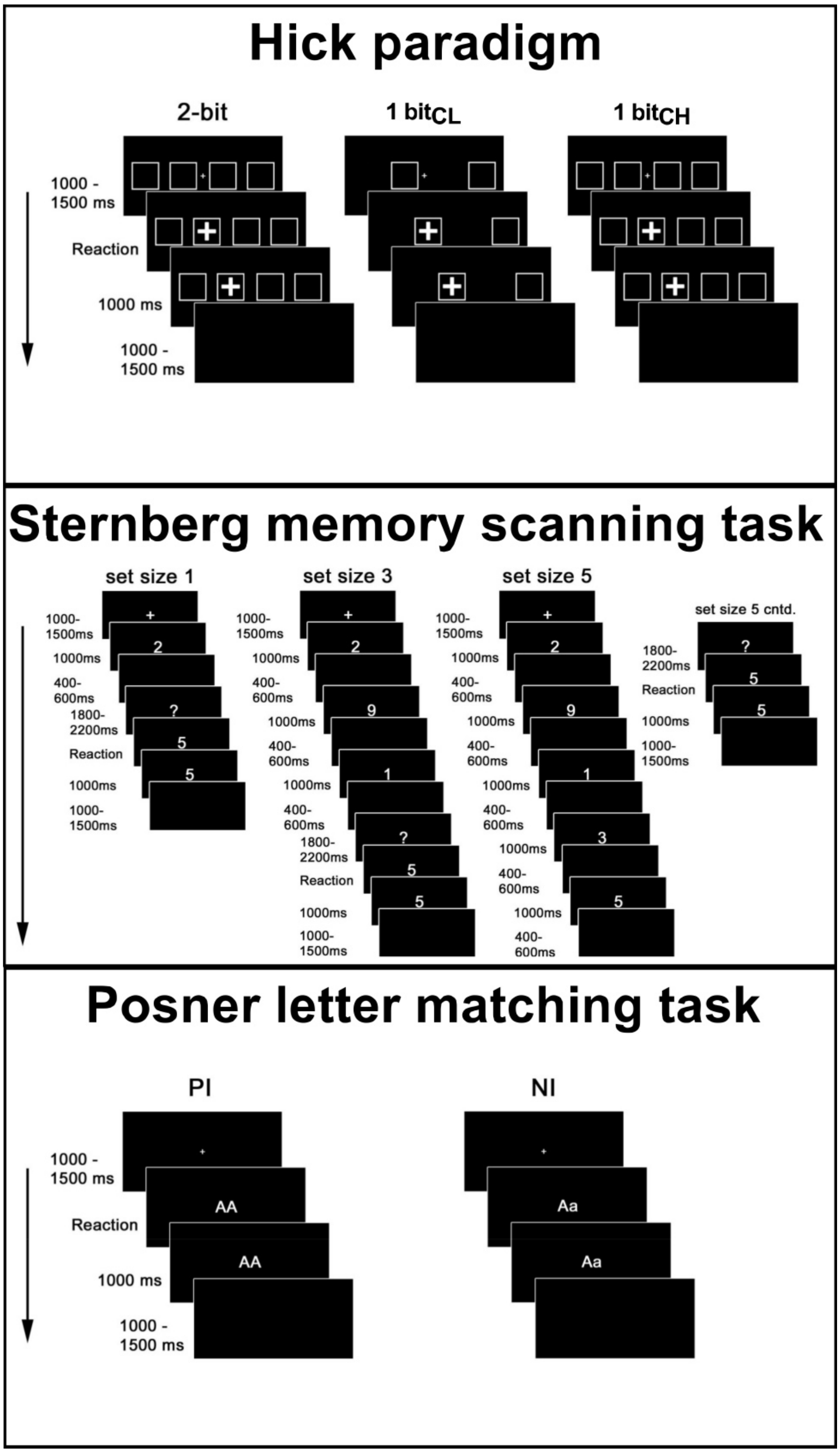

Figure 1. Stimulus material for the three ECTs. A: The three conditions of the modified Hick paradigm. 2-bit = 2-bit condition, 1-bit$_{CL}$ = 1-bit condition with low stimulus comparability to the 2-bit condition, 1-bit$_{CH}$ = 1-bit condition with high stimulus comparability to the 2-bit condition. B: Different memory set sizes in the Sternberg memory scanning task. Physical identity (PI) and name identity (NI) condition in the Posner letter matching task.

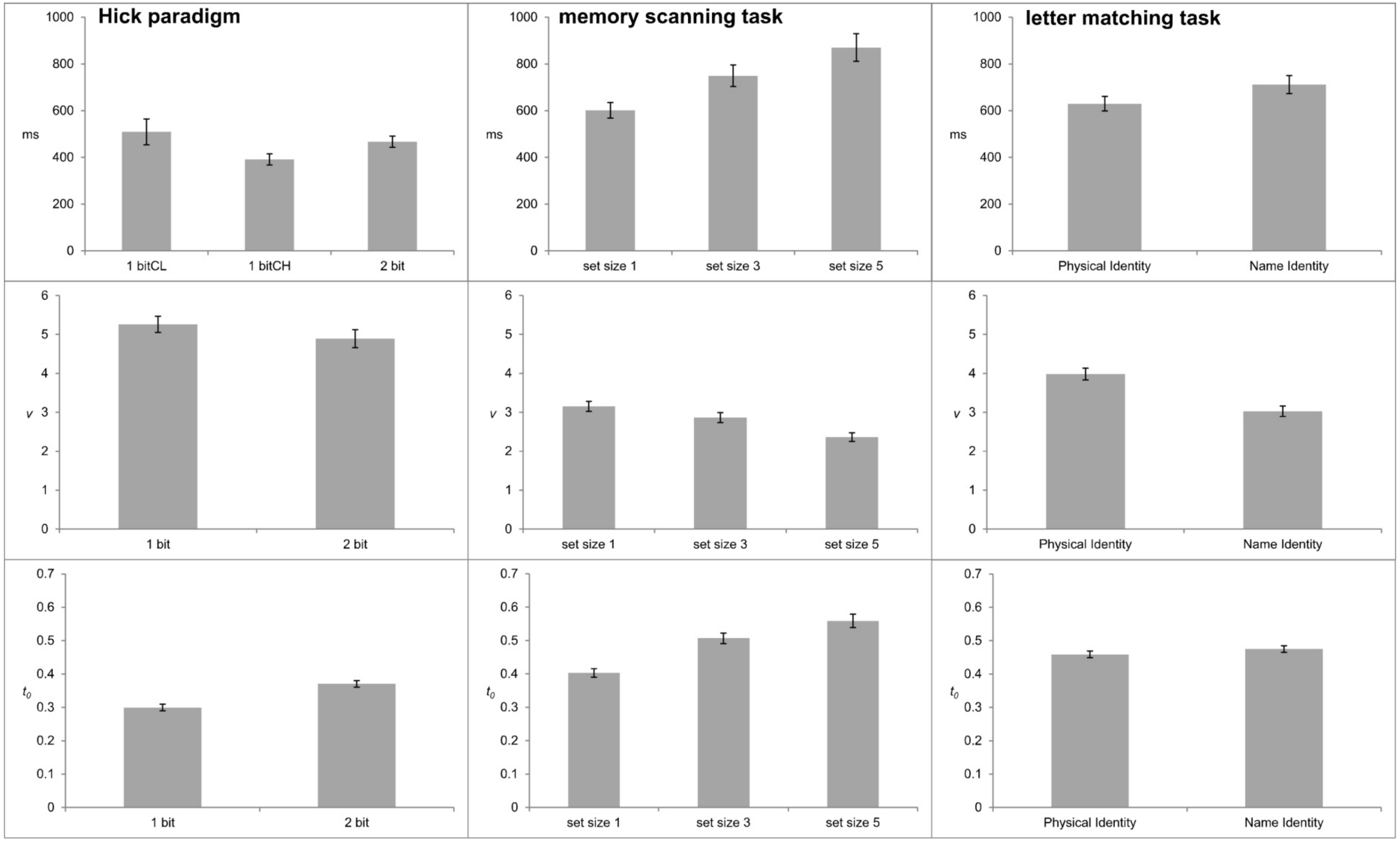

Figure 2. Median RTs, mean drift rates, and mean response time-constants for the different conditions of the three ECTs. Error bars represent doubled standard errors.

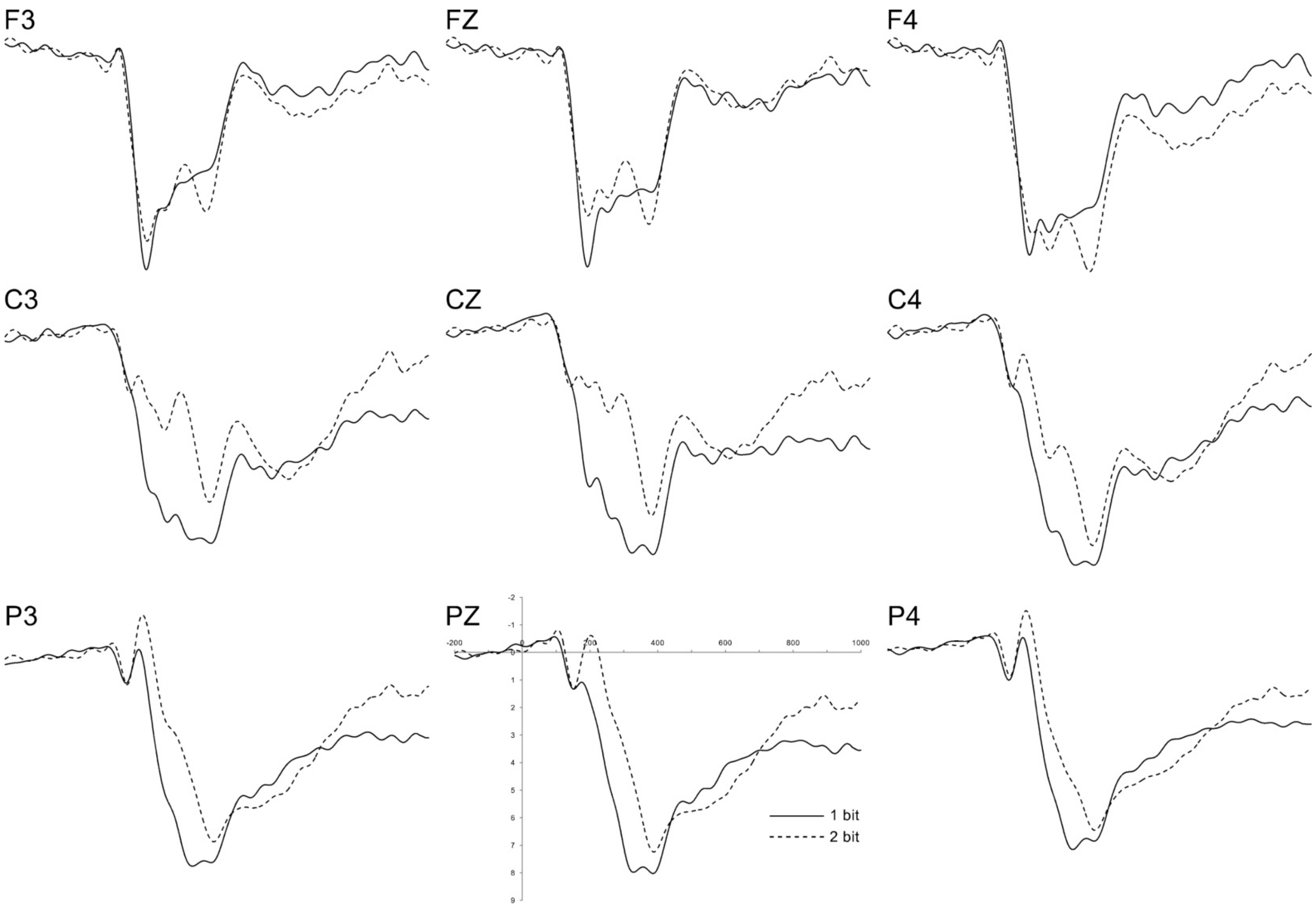

Figure 3. Event-related potentials elicited by the onset of the cross in the 1 bit condition (solid lines) and 2 bit condition (broken lines) in the Hick paradigm. Electrodes are arrayed from most anterior (top) to most posterior (bottom) and from left to right.

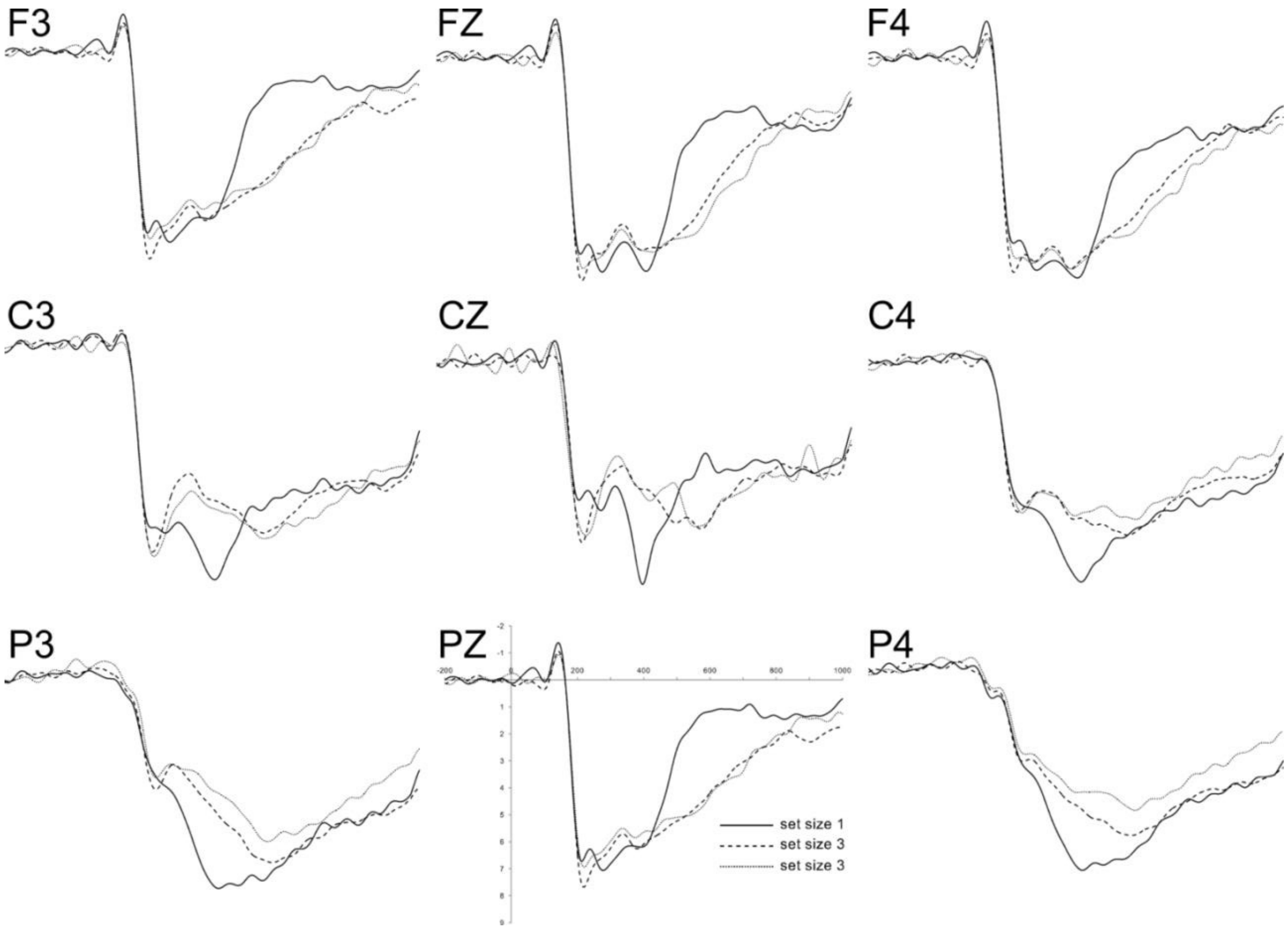

Figure 4. Event-related potentials elicited by the onset of the memory probe for the different memory set sizes in the Sternberg letter matching task. Electrodes are arrayed from most anterior (top) to most posterior (bottom) and from left to right.

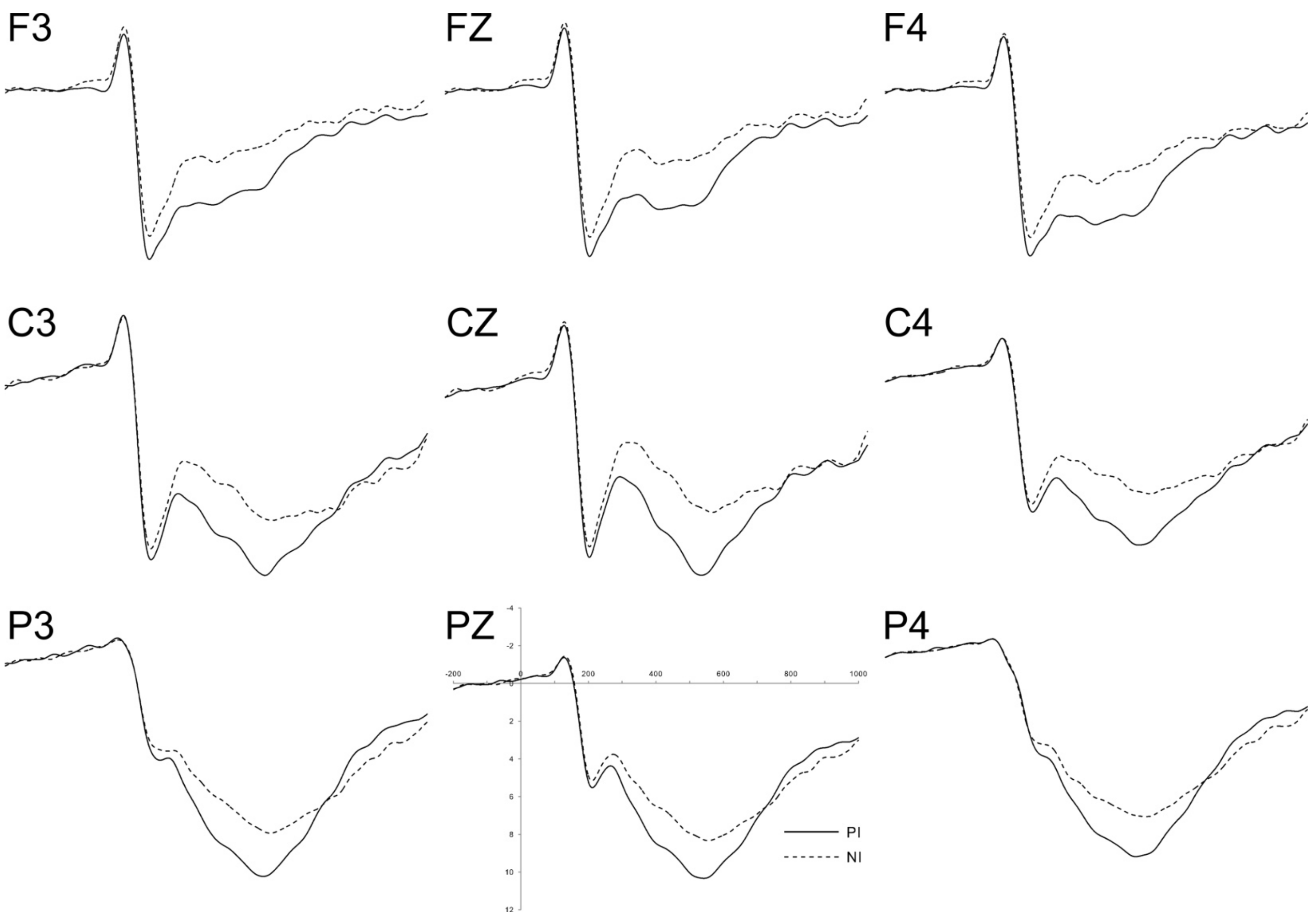

Figure 5. Event-related potentials elicited by the onset of the letter pair in the PI condition (PI = physical identity; solid lines) and the NI condition (NI = name identity; broken lines) in the Posner letter matching task. Electrodes are arrayed from most anterior (top) to most posterior (bottom) and from left to right.

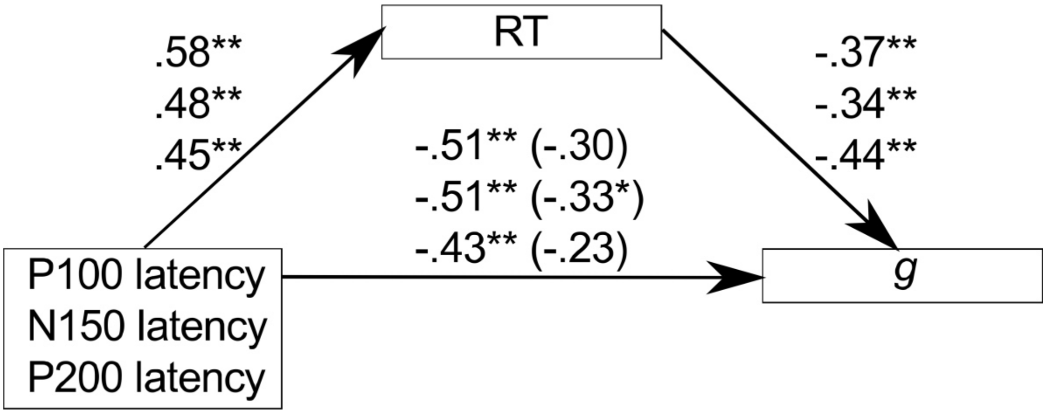

Figure 6. Standardized regression coefficients for the association between ERP latencies and general intelligence mediated by reaction times. The standardized regression coefficients between ERP latencies and general intelligence controlling for reaction times are in parantheses.

* $p < .05$. ** $p < .01$.

Supplementary Information

Table 12

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on P200 (175-215 ms) activity in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,35	25.18	<.001	-	.41
Caudality	2,70	53.15	<.001	.60	.60
Laterality	2,70	4.98	.012	.94	.10
Condition x Caudality	2,70	12.40	<.001	.67	.25
Condition x Laterality	2,70	11.08	<.001	.90	.22
Caudality x Laterality	4,140	9.57	<.001	.72	.20
Condition x Caudality x Laterality	4,140	<1	.765	.66	.00

Table 13

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on N200 (210-240 ms) activity in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,35	25.89	<.001	-	.42
Caudality	2,70	16.80	<.001	.61	.31
Laterality	2,70	1.68	.197	.92	.02
Condition x Caudality	2,70	27.79	<.001	.43	.25
Condition x Laterality	2,70	16.17	<.001	.82	.30
Caudality x Laterality	4,140	1.20	.313	.82	.00
Condition x Caudality x Laterality	4,140	<1	.654	.72	.00

Table 14

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (360-420 ms) activity in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,35	<1	.497	-	.00
Caudality	2,70	18.08	<.001	.63	.33
Laterality	2,70	19.51	<.001	.89	.35
Condition x Caudality	2,70	12.90	<.001	.69	.25
Condition x Laterality	2,70	8.21	<.001	.71	.13
Caudality x Laterality	4,140	4.89	.004	.71	.10
Condition x Caudality x Laterality	4,140	<1	.526	.59	.00

Table 15

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on P200 (175-215 ms) peak latencies in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω²
Condition	1,35	13.91	.001	-	.27
Caudality	2,70	1.26	.284	.78	.00
Laterality	2,70	5.69	.008	.84	.12
Condition x Caudality	2,70	8.42	.003	.65	.17
Condition x Laterality	2,70	<1	.631	.93	.00
Caudality x Laterality	4,140	3.68	.020	.63	.07
Condition x Caudality x Laterality	4,140	<1	.932	.69	.00

Table 16

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on N200 (210-240 ms) peak latencies in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,35	<1	.813	-	.00
Caudality	2,70	22.99	<.001	.74	.39
Laterality	2,70	16.25	.197	.99	.30
Condition x Caudality	2,70	5.35	.009	.90	.11
Condition x Laterality	2,70	3.84	.026	.82	.08
Caudality x Laterality	4,140	5.64	.001	.89	.12
Condition x Caudality x Laterality	4,140	5.29	.002	.73	.11

Table 17

Results of the ANOVA with the three within-subject factors condition (1 bit vs. 2 bit), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (360-420 ms) peak latencies in the Hick paradigm. (n = 36)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,35	<1	.759	-	.00
Caudality	2,70	9.95	.001	.64	.20
Laterality	2,70	1.75	.191	.75	.02
Condition x Caudality	2,70	7.88	.004	.64	.16
Condition x Laterality	2,70	1.50	.233	.81	.01
Caudality x Laterality	4,140	<1	.44	.70	.00
Condition x Caudality x Laterality	4,140	<1	.40	.68	.00

Table 18

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on N150 (115-160 ms) activity in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	<1	.95	.66	.00
Caudality	2,76	30.26	<.001	.59	.44
Laterality	2,76	2.18	<.001	.89	.03
Condition x Caudality	4,152	2.08	.151	.32	.03
Condition x Laterality	4,152	<1	.813	.39	.00
Caudality x Laterality	4,152	2.92	.032	.83	.05
Condition x Caudality x Laterality	8,304	1.04	.334	.16	.00

Table 19

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on P200 (200-245 ms) activity in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	2.13	.134	.86	.03
Caudality	2,76	33.41	<.001	.62	.46
Laterality	2,76	23.82	<.001	1.00	.38
Condition x Caudality	4,152	1.48	.235	.34	.01
Condition x Laterality	4,152	2.05	.149	.37	.03
Caudality x Laterality	4,152	13.87	<.001	.84	.26
Condition x Caudality x Laterality	8,304	1.19	.289	.15	.00

Table 20

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on N300 (300-360 ms) activity in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω²
Condition	2,76	14.41	<.001	.93	.26
Caudality	2,76	3.92	0.046	.61	.07
Laterality	2,76	14.59	<.001	.87	.26
Condition x Caudality	4,152	7.46	.002	.42	.15
Condition x Laterality	4,152	<1	.595	.55	.00
Caudality x Laterality	4,152	1.78	.163	.65	.02
Condition x Caudality x Laterality	8,304	1.20	.300	.19	.01

Table 21

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (400-600 ms) activity in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	22.33	<.001	.93	.36
Caudality	2,76	<1	.584	.59	.00
Laterality	2,76	18.32	<.001	.95	.31
Condition x Caudality	4,152	18.32	<.001	.45	.31
Condition x Laterality	4,152	<1	.409	.49	.00
Caudality x Laterality	4,152	1.78	.137	.98	.02
Condition x Caudality x Laterality	8,304	<1	.387	.23	.00

Table 22

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on N150 (115-160 ms) peak latencies in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	1.55	.22	.85	.01
Caudality	2,76	15.40	<.001	.65	.27
Laterality	2,76	6.82	.004	.84	.13
Condition x Caudality	4,152	<1	.758	.72	.00
Condition x Laterality	4,152	<1	.695	.89	.00
Caudality x Laterality	4,152	<1	.479	.75	.00
Condition x Caudality x Laterality	8,304	1.12	.351	.60	.00

Table 23

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on P200 (200-245 ms) peak latencies in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	<1	.433	.80	.00
Caudality	2,76	12.21	<.001	.75	.23
Laterality	2,76	10.96	<.001	.89	.21
Condition x Caudality	4,152	<1	.485	.51	.00
Condition x Laterality	4,152	<1	.782	.70	.00
Caudality x Laterality	4,152	10.44	<.001	.82	.20
Condition x Caudality x Laterality	8,304	<1	.578	.57	.00

Table 24

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on N300 (300-360 ms) peak latencies in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	2.01	.143	.96	.03
Caudality	2,76	9.05	.002	.67	.17
Laterality	2,76	3.18	.058	.82	.05
Condition x Caudality	4,152	3.37	.019	.78	.06
Condition x Laterality	4,152	2.43	.064	.81	.04
Caudality x Laterality	4,152	1.72	.162	.83	.02
Condition x Caudality x Laterality	8,304	1.69	.123	.76	.02

Table 25

Results of the ANOVA with the three within-subject factors condition (set size 1, set size 3, set size 5), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (400-600 ms) peak latencies in the Sternberg memory scanning paradigm. (n = 39)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	2,76	6.43	.005	.82	.13
Caudality	2,76	26.96	<.001	.80	.41
Laterality	2,76	<1	.645	.85	.00
Condition x Caudality	4,152	1.84	.139	.82	.02
Condition x Laterality	4,152	2.06	.100	.85	.03
Caudality x Laterality	4,152	2.11	.104	.75	.03
Condition x Caudality x Laterality	8,304	<1	.767	.76	.00

Table 26

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on N140 (115-155 ms) activity in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	<1	.712	-	.00
Caudality	2,66	35.63	<.001	.54	.51
Laterality	2,66	39.63	<.001	.76	.54
Condition x Caudality	2,66	2.22	.139	.62	.04
Condition x Laterality	2,66	<1	.377	.95	.00
Caudality x Laterality	4,132	14.85	<.001	.78	.30
Condition x Caudality x Laterality	4,132	2.68	.045	.84	.05

Table 27

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on P210 (190-235 ms) activity in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	<1	.535	-	.00
Caudality	2,66	16.96	<.001	.56	.33
Laterality	2,66	6.81	.002	.97	.15
Condition x Caudality	2,66	7.37	.007	.58	.16
Condition x Laterality	2,66	<1	.480	.77	.00
Caudality x Laterality	4,132	20.00	<.001	.68	.37
Condition x Caudality x Laterality	4,132	2.04	.122	.67	.03

Table 28

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on N300 (240-365 ms) activity in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	1.15	.292	-	.00
Caudality	2,66	3.69	.052	.63	.08
Laterality	2,66	11.07	<.001	.90	.23
Condition x Caudality	2,66	5.91	.015	.61	.13
Condition x Laterality	2,66	<1	.573	.69	.00
Caudality x Laterality	4,132	5.91	.003	.54	.13
Condition x Caudality x Laterality	4,132	2.79	.043	.77	.05

Table 29

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (465-580 ms) activity in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	2.53	.121	-	.04
Caudality	2,66	30.78	<.001	.59	.47
Laterality	2,66	12.78	<.001	1.00	.26
Condition x Caudality	2,66	<1	.424	.61	.00
Condition x Laterality	2,66	3.64	0.38	.88	.07
Caudality x Laterality	4,132	6.08	.001	.80	.13
Condition x Caudality x Laterality	4,132	2.84	.037	.83	.05

Table 30

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on N140 (115-155 ms) peak latencies in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	<1	.580	-	.00
Caudality	2,66	4.18	.040	.61	.09
Laterality	2,66	2.72	.088	.77	.05
Condition x Caudality	2,66	<1	.794	.61	.00
Condition x Laterality	2,66	1.07	.321	.61	.00
Caudality x Laterality	4,132	<1	.467	.43	.00
Condition x Caudality x Laterality	4,132	<1	.413	.62	.00

Table 31

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on P210 (190-235 ms) peak latencies in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	<1	.471	-	.00
Caudality	2,66	4.06	.044	.59	.09
Laterality	2,66	12.92	<.001	.92	.05
Condition x Caudality	2,66	3.14	.066	.75	.06
Condition x Laterality	2,66	2.69	.086	.84	.05
Caudality x Laterality	4,132	4.99	.007	.57	.11
Condition x Caudality x Laterality	4,132	1.41	.243	.82	.01

Table 32

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on N300 (240-365 ms) peak latencies in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω^2
Condition	1,33	1.17	.288	-	.00
Caudality	2,66	7.09	.005	.74	.16
Laterality	2,66	<1	.616	.92	.00
Condition x Caudality	2,66	<1	.629	.98	.00
Condition x Laterality	2,66	1.04	.356	.94	.00
Caudality x Laterality	4,132	<1	.630	.67	.00
Condition x Caudality x Laterality	4,132	1.87	.148	.65	.03

Table 33

Results of the ANOVA with the three within-subject factors condition (Physical Identity vs. Name Identity), caudality (frontal, central, parietal), and laterality (left, central, right) on P300 (465-580 ms) peak latencies in the Posner letter matching paradigm. (n = 35)

Variable	*df*	*F*	*p*	ε	ω²
Condition	1,33	3.01	.092	-	.06
Caudality	2,66	5.23	.014	.78	.11
Laterality	2,66	6.05	.006	.89	.13
Condition x Caudality	2,66	<1	.443	.73	.00
Condition x Laterality	2,66	<1	.614	.85	.00
Caudality x Laterality	4,132	<1	.936	.73	.00
Condition x Caudality x Laterality	4,132	<1	.832	.80	.00

Appendix A2 – Manuscript 2

Evaluating the model fit of diffusion models with the root mean square error of approximation

Anna-Lena Schubert[a], Dirk Hagemann[b], Andreas Voss[c], and Katharina Bergmann[d]

University of Heidelberg

Author Note

[a]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: anna-lena.schubert@psychologie.uni-heidelberg.de

[b]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: dirk.hagemann@psychologie.uni-heidelberg.de

[c]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: andreas.voss@psychologie.uni-heidelberg.de

[d]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: katharina.bergmann@psychologie.uni-heidelberg.de

Correspondence concerning this article should be addressed to Anna-Lena Schubert, University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, Phone: +49 (0) 6221-547354, Fax: +49 (0) 6221-547325, E-mail: anna-lena.schubert@psychologie.uni-heidelberg.de

Abstract

The statistical evaluation of model fit is one of the greatest challenges in the application of diffusion modeling in research on individual differences. Relative model fit indices such as the AIC and BIC are often used for model comparison, but they provide no information about absolute model fit. Statistical and graphical tests can be used to identify individuals whose data cannot be accounted for by the diffusion model, but they become overly sensitive when trial numbers are large, and are subjective and time-consuming. We propose that the evaluation of model fit may be supplemented with the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), which is one of the most popular goodness-of-fit indices in structural equation modeling. It is largely invariant to trial numbers, and allows identifying cases with poor model fit, calculating confidence intervals, and conducting power analyses. In two simulation studies, we evaluated whether the RMSEA correctly rejects badly-fitting models irrespective of trial numbers. Moreover, we evaluated how variation in the number of trials, the degree of measurement noise, the presence of contaminant outliers, and the number of estimated parameters affects RMSEA values. The RMSEA correctly distinguished between well- and badly-fitting models unless trial numbers were very small. Moreover, RMSEA values were in a value range expected from structural equation modeling. Finally, we computed cut-off values as heuristics for model acceptance or rejection. In a third simulation study we assessed how the RMSEA performs in model selection in comparison to the AIC and BIC. The RMSEA correctly identified the generating model in the majority of cases, but was outperformed by the AIC and BIC. All in all, we showed that the RMSEA can be of great value in the evaluation of absolute model fit, but that it should only be used in addition to other fit indices in model selection scenarios.

Evaluating the model fit of diffusion models with the root mean square error of approximation

1. Introduction

In recent years, the diffusion model for binary responses (Ratcliff, 1978) has seen a huge rise of popularity in a wide area of research fields (Voss, Nagler, & Lerche, 2013). Within the diffusion model framework, response time distributions can be decomposed in terms of different parameters associated with specific cognitive processes. While traditionally researchers interested in mental chronometry drew inferences based on mean (and sometimes SDs of) response times and could therefore only infer *whether* response times differ between experimental conditions and/or individuals, they are now able to infer which processing components may be responsible for the observed response time differences.

The diffusion model has been successfully applied in the context of social cognitive research (e.g., Germar, Schlemmer, Krug, Voss, & Mojzisch, 2014; Klauer, Stahl, & Voss, 2011; Voss, Rothermund, & Brandstädter, 2008; Voss & Schwieren, 2015), prospective memory (e.g., Boywitt & Rummel, 2012), aging (e.g., McKoon & Ratcliff, 2013; Spaniol, Madden, & Voss, 2006), individual differences (e.g., Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Schubert, Hagemann, Voss, Schankin, & Bergmann, 2015), and in many more areas of research. It is, however, not always a priori known whether the diffusion model is an adequate process model for the cognitive processes that resulted in a specific distribution of response times and response frequencies in an experimental paradigm. A necessary (but not sufficient) precondition for the interpretation of diffusion model results is the model fit, that is, the degree of match between predicted and observed response time data. Even if it is undisputed that a specific task is principally suited for diffusion modeling, researchers still have to make an informed decision about the specific implementation of the model (e.g., the coding of responses or the number of estimated parameters). Therefore, it is crucial to evaluate how well the estimated model parameters can account for the actual accuracy and response time data, and to identify cases in which the diffusion model fails to describe the data.

Currently, there is no universally accepted gold standard for evaluating model fit in the diffusion model framework. Model fit indices such as the Akaike Information Criterion (AIC; Akaike, 1973) or the Bayes Information Criterion (BIC; Schwarz, 1978) are often used for model comparison and selection purposes, but they provide no information about absolute model fits. Therefore, in diffusion modeling these criteria only convey information about which of several models accounts for the empirical data best, but they do not help to decide whether the models should be accepted or rejected. Moreover, such relative fit indices do not allow identifying and possibly excluding individuals whose data cannot be accounted for within the diffusion model framework, as general suggestions when a model should be rejected cannot be given. In order to identify individuals with badly-fitting model parameters, different strategies are usually pursued. Statistical tests of model fit, such as the χ^2 test, are very common, but sensitivity of these tests is closely tied to the amount of data that is available. For small data sets this leads to a power problem as the test power may be too small to reject the model, and for larger data the test will become overly sensitive. Statistical tests are also biased in favor of more complex models, as a model with higher degrees of freedom that most often provides a better account for the data is not punished in comparison to a more parsimonious model.

To overcome the problems associated with null-hypothesis testing of model fit, Clauset, Shalizi, and Newman (2009), and Voss, Nagler, and Lerche (2013) suggested simulating a large number of synthetic data sets based on the estimated model parameters and deriving critical *p*-

values from these subsequently re-fitted data sets. The 5% or 10% quantile of the distribution of *p*-values can then be taken as a critical value for the evaluation of the empirical models. This method overcomes some of the problems associated with the statistical testing of model fits, but does not specifically consider trial numbers and the parsimony of a model. Moreover, models get accepted with an unknown error probability, as this procedure does not offer a method to estimate the statistical power of the test.

Graphical methods provide an alternative approach to the evaluation of model fit. For this approach the deviation of the predicted response times from the empirical response times are displayed either individually, or for a complete sample (e.g., Schmitz and Voss, 2014; Voss, Rothermund, Gast, & Wentura, 2013). Decisions based on graphical tests are, however, subjective and may therefore lead to spurious conclusions (D'Agostino, 1986). Moreover, extensive graphical model tests for each individual can quickly become time-consuming in large samples.

An ideal goodness-of-fit (GOF) index that can be used to identify data sets that the diffusion model is not able to account well for should have the properties of the very popular AIC and BIC (i.e., reward parsimonious models and not be strongly affected by variations in trial numbers), but would be an absolute measure of model fit, not a relative one. As such, it would presume that a perfectly fitting model has a fit value of zero and that a deviation from zero indicates how far the model deviates from perfect fit. Then, this deviation from perfect model fit could be compared across different models as well as between cases, and standards for acceptable model-fit could be defined. Moreover, conventions for cut-off values could be suggested indicating when a model should be rejected, and these cut-off values would be invariant across different applications (and therefore, different trial numbers and different degrees of parsimony) of the model.

Another field of research that has been concerned with the performance of GOF indices is structural equations modeling (SEM), which is a statistical technique for testing the structural relations within a multivariate data space. To maximize model fit, the discrepancy between the empirical covariance matrix of all measured variables and the covariance matrix implied by the model specifications gets minimized. As in diffusion modeling, this minimization process does not yield a GOF value that can reasonably be used to decide about model acceptance due to the same problems as listed above. Because participant numbers are typically very large in SEM studies (> 200 participants), model predictions often deviate significantly from the empirical data, although the model fit is actually quite good. Therefore, there have been several suggestions how to evaluate model fit in the SEM framework (see Jackson, Gillaspy, & Stephenson, 2009, for a review). Several of these GOF indices are not easily transferable to the diffusion model framework, because they compare the performance of the estimated model to a baseline model (in which all variables are presumed to be uncorrelated). Within the diffusion model framework, no such baseline model could be easily specified without further debatable assumptions, because it is entirely unclear which configuration of parameter values might reflect an appropriate baseline model.

One very popular absolute GOF index, the root mean square error of approximation (RMSEA; Steiger & Lind, 1980), however, does not require the assumption of a baseline model, but is based on the noncentrality parameter of the χ^2 distribution. The RMSEA is relatively unaffected by variations in sample size and rewards parsimonious models. Moreover, as the RMSEA is an absolute fit index with a minimum of zero, conventions for cut-off values have been suggested and are frequently used in the SEM framework. Because the RMSEA is based on the noncentrality parameter of the χ^2 distribution, it could be easily reported and used as an evaluation criterion in addition to the χ^2 statistic and its corresponding *p*-value in the context of diffusion modeling. In the present paper, we propose to use the RMSEA as an index of absolute model fit within the diffusion model framework and discuss its benefits in comparison to standard methods of model evaluation.

1.1 The diffusion model

The diffusion model makes the basic assumption that during a decision process with two alternatives, information is accumulated continuously until the diffusion process reaches one of two thresholds. Specifically, this information accumulation process consists of a constant systematic component, the *drift*, and normally distributed random noise. The basic diffusion model estimates four parameters from empirical response time distributions: The drift rate (v), which describes the strength and direction of the systematic influence on the diffusion process, the threshold separation (a), which maps the amount of information that is used for a decision, the starting point (z), which indicates possible biases towards one of the two decision thresholds, and the non-decision time (t_0 or t_{er}), which encompasses all processes unrelated to the decision such as encoding and motor reaction time. In the full diffusion model, inter-trial variabilities of drift (s_v or η), starting point (s_z), and non-decision time (s_{t0}) can be estimated, which increases model fit by accounting for different shapes of response time distributions for correct responses and errors (Ratcliff & Rouder, 1998; Ratcliff & Tuerlinckx, 2002).

Different optimization criteria are available to minimize the deviation of the predicted response latencies from the empirical data (Voss, Voss, & Lerche, 2015). Most commonly, optimization processes are based on the χ^2 statistic (e.g., Ratcliff & McKoon, 2008; Ratcliff &

Tuerlinckx, 2002; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Separately for the upper and lower threshold, responses are grouped into six bins that are defined by the .1, .3, .5, .7, and .9 quantiles of the empirical RT distribution (Ratcliff & Tuerlinckx, 2002). Then the expected number of responses per bin under the predicted cumulative distribution function (CDF) is compared with the number of observed responses per bin *i*, and the resulting χ^2 statistic is minimized in iterative steps:

$$\chi^2 = \sum_{i=1}^{t} \frac{(n_{iobserved} - n_{ipredicted})^2}{n_{ipredicted}} \quad (1)$$

The greatest advantages of the χ^2 statistic as an optimization criterion are the high computation speed of the optimization algorithm and its robustness against outliers. One major disadvantage is that the χ^2 criterion yields biased parameter estimates if trial numbers are small (Lerche, Voss, & Nagler, submitted). Other optimization criteria that are not covered here are based on a maximum likelihood approach or on the Kolmogorov-Smirnov statistic and are more successful at fitting small trial numbers (for an overview of the different optimization criteria see Van Zandt, 2000; Voss, Nagler, & Lerche, 2013).

1.2 The root mean square error of approximation

We propose to supplement the established approaches to assess the fit of the diffusion model with the root mean square error of approximation initially proposed by Steiger and Lind (1980) and later discussed in more detail and popularized by Browne and Cudeck (1993). Nowadays, it is one of the most frequently used goodness-of-fit measures in structural equation modeling (SEM). Jackson et al. (2009) recently reviewed 194 SEM studies published between 1998 and 2006 and reported that the RMSEA was the second most popular goodness-of-fit index reported in 64.9% of the reviewed studies. The great popularity of the RMSEA is based on its properties: First, simulation studies have shown that it is largely invariant with regard to sample size. Second, it rewards parsimonious models. Third, the RMSEA has a minimum of zero with suggested conventions for cut-off criteria for excellent, good and mediocre model fit (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996). Fourth, a confidence interval around the point estimate of the RMSEA can be computed to assess the degree of uncertainty, which allows to hypothesis testing (Browne & Cudeck, 1993; MacCallum et al., 1996). Fifth, a power analysis can be conducted fairly easily to determine the trial number necessary for adequate power in these tests, (MacCallum et al., 1996).

The RMSEA is a goodness-of-fit measure based on the noncentrality parameter of the noncentral χ^2 distribution. The noncentrality parameter is defined as $\delta = \chi^2 - df$, where the χ^2 statistic indicates the deviation of the model predictions from the observed data (see equation 1), and *df* are the degrees of freedom ($df = 2 * b - 1 - p$ with *b* response time bins and *p* free parameters). The RMSEA is then computed as the square root of the normalized noncentrality parameter per degree of freedom

$$\varepsilon = \sqrt{\frac{\max(\delta,0)}{df(N-1)}}, \quad (2)$$

where *N* is the number of trials (in the SEM framework, *N* is the number of participants). Smaller values indicate better model fit. If the χ^2 statistic is smaller than the model's degrees of freedom, ε is set to zero.

As is immediately evident from equation 2, both the nominator (due to a higher sensitivity to small deviations in model fit) and the denominator increase with increasing trial numbers. Moreover, liberal models are punished more strongly in comparison to more parsimonious models because of the χ^2 to *df* ratio inherent to equation 2. Thus, the RMSEA tries to reward model parsimony.

1.2.1 Interpretation of the RMSEA

Values of ε smaller than .05 are typically considered to indicate good fit (Browne & Cudeck, 1993). Browne and Cudeck (1993) further recommended interpreting values ranging from .05 to .08 as fair model fit, and values greater than .10 as poor fit. MacCallum et al. (1996) suggested that values in the range from .08 to .10 indicate mediocre fit. These conventions for cut-off values are based on over 20 years of experience and simulations in the SEM framework. It is unclear whether the same cut-off values are appropriate within the diffusion model framework, as too many or too few models might be rejected when they are used as a decision criterion. Based on these cut-off values, specific hypothesis of fit can be tested such as the hypothesis of exact fit (i.e., H_0: RMSEA = 0), or the hypothesis of close fit, (i.e., H_0: RMSEA ≤ .05; see Browne & Cudeck, 1993; MacCallum et al., 1996).

1.2.2 Absolute indices of model fit and model selection

Absolute indices of model fit such as the RMSEA quantify the discrepancy between perfect fit, i.e. no deviation between the predicted and the empirical response latencies, and the fit of a specific instantiation of a model. For this purpose, it is assumed that a perfectly fitting model has a χ^2 value that is not larger than the model's degrees of freedom, resulting in a RMSEA value of zero for perfectly fitting models. This fixed minimum value allows identifying badly-fitting models that deviate too far from perfect model fit. At the same time it limits the severity with which model complexity can be penalized. Assuming that the generating model has a perfect fit with $\delta = 0$ (i.e., a χ^2 value smaller than or equal to the model's degrees of freedom), a nested model with additional model parameters also has a perfect fit. In this case, the nominator of the

RMSEA and as well as ε will be zero irrespective of model complexity. Consequently, the RMSEA is only able to penalize model complexity reasonably well when the difference between the χ^2 value and the degrees of freedom is substantially large.

When the RMSEA is used in to identify and subsequently exclude data sets for which the diffusion model does not provide a good account, this limitation may lead to an acceptance of over-parameterized models even though they should be rejected in favor of a more parsimonious model under considerations of parsimony. Nevertheless, the over-parameterized model is still a well-fitting model and model parameters can be interpreted and used for subsequent correlational analyses, although the reliability of these parameters may suffer due to the over-parameterization. However, if the RMSEA is used for model selection purposes, it may select the more complex model even when the generating model or the model with the best predictive validity has fewer model parameters. Moreover, it is also unclear whether the penalty for model complexity is sufficient for model selection even in cases where the χ^2 value exceeds the degrees of freedom. In a recent evaluation of model selection performance of different fit indices in the SEM framework, the RMSEA had the least success in identifying the generating model even at large sample sizes (Bollen, Harden, Ray, & Zavisca, 2014). Therefore, we anticipate that the RMSEA cannot keep up in model selection accuracy in the diffusion model framework with relative fit indices such as the AIC or BIC. Nevertheless, we believe that it can be a useful addition to other model evaluation procedures when the absolute goodness-of-fit has to be evaluated prior to further multivariate analyses of diffusion model parameters.

An alternative approach to the evaluation of model fit in model selection scenarios could be based on the likelihood ratio statistic $G^2(a,b) = -2\ln[L_a(X,\theta_a)/L_b(X,\theta_b)]$, with $L_a(X,\theta_a)$ corresponding to the maximum likelihood of model a and $L_b(X,\theta_b)$ corresponding to the maximum likelihood of model *b*. Because the likelihood ratio statistic is χ^2 distributed (Wilks, 1938), it can be subsequently used to calculate the noncentrality parameter of the χ^2 distribution and the RMSEA with degrees of freedom equal to the difference in estimated parameters. Model complexity would still be punished by a decrease in the nominator and by a smaller χ^2 to *df* ratio in equation 2. Larger RMSEA values would then reflect a better model fit for the nested in comparison to the less complex model. This approach has never been used in structural equation modeling.

1.3 The present study

In order to evaluate how the RMSEA performs in the diffusion model framework compared to the SEM framework, we conducted three simulation studies. In the first two simulation studies we simulated data sets from diffusion models with different numbers of free parameters and varied the number of trials, the degree of measurement noise, the presence of contaminated trials, and the number of estimated parameters. We pursued three general aims in these two studies: First, we wanted to evaluate how the RMSEA behaves in comparison to statistical tests of model fit when trial numbers were varied. Specifically, we wanted to show that rejection decisions based on the RMSEA are largely invariant with regard to variation in trial numbers, whereas rejection rates based on the χ^2 statistic increase with increasing trial numbers. Second, we wanted to investigate whether RMSEA values for well-fitting models in the diffusion model framework were comparable to RMSEA values typically observed in structural equation modeling. Third, we wanted to establish cut-off values for acceptance and rejection decisions in the diffusion model framework.

In the third simulation study, we compared the RMSEA to the AIC and BIC in its ability to select the generating model from among other models. For this purpose, we simulated data sets from diffusion models with two, one, or no parameters varying between conditions, and varied the number of estimated parameters varying between conditions to evaluate how often each fit index identified the generating model as the correct model from over- and under-parameterized alternatives.

2. Simulation Study 1

The aim of the first simulation study was to evaluate how the RMSEA performs when response time data are simulated from the diffusion model. In particular, we wanted to compare rejection rates based on the RMSEA to rejection rates based on statistical tests of model fit in their dependency on trial numbers. For this purpose, we varied the number of parameters in the generator models and the number of trials. Moreover, we added different degrees of measurement noise to the response time data to evaluate how it affected the RMSEA. For this contamination scenario, we added normally distributed, random noise to reflect a variety of processes such as fatigue, motivation, technical problems, or disruptions that may lead to a general decrease in the signal-to-noise ratio in response times. We then re-fitted these data sets, varied the number of parameters in the estimated models, and computed the corresponding RMSEA values and χ^2 statistics. To pursue the three aims of the present study, we then a) compared rejection decisions based on the RMSEA to rejection decisions based on the χ^2 statistic, b) compared RMSEA values for well-fitting diffusion models to RMSEA values for well-fitting models in the SEM framework, and c) computed cut-off values indicating good model fit. For the data sets with only sampling noise, we expected RMSEA values to indicate good model fit (i.e., $\varepsilon \leq .05$) regardless of trial numbers or the number of parameters in the estimated model. Moreover, we expected that estimated models with lower complexity do better when evaluated with the RMSEA than with the corresponding χ^2 statistic. We could not make clear predictions about the effect of trial numbers on RMSEA values. RMSEA values should be less affected by increasing trial numbers than the χ^2 statistic, but the RMSEA is also known to reject models too often when both sample sizes and degrees of freedom are small (Hu & Bentler, 1999; Kenny, Kaniskan, & McCoach, 2014). Moreover, when inter-trial variabilities were estimated, we expected the RMSEA to be only marginally affected when slight or moderate measurement random noise was added to the data, and to increase substantially when substantial measurement noise was added to the data, thus reflecting that the data generating process differed considerably from the diffusion model. When inter-trial variabilities were not estimated, we expected the RMSEA to be strongly affected by all levels of measurement noise.

2.1 Method

2.1.1 Data simulation

We generated 10,000 random diffusion model parameter sets (v, a, z, t_0, s_v, s_z, s_{t0}) from uniform distributions in the range of typically observed parameters values (see Table 1). For each of these parameter sets, we simulated three data sets of response times and accuracies with different numbers of trials (100; 500; 1000) using all seven diffusion model parameters (v, a, z, t_0, s_v, s_z, s_{t0}) for the generating model. In addition, we repeated this simulation using only four diffusion model parameters (v, a, z, t_0) for the generating model (the inter-trial variability parameters, s_v, s_z, s_{t0}, were fixed to zero). All data sets were simulated with the *construct-samples* routine from *fast-dm*, which samples random data sets from a multinomial distribution defined by the diffusion model (Voss & Voss, 2007; Voss, Voss, & Lerche, 2015). These are the data sets with only sampling noise. Note that the variation of trial numbers is also an indirect manipulation of sampling noise, as sampling noise is inversely related to trial numbers.

Subsequently, we added different degrees of measurement noise to the response times of each of these predicted data sets in the form of a normally distributed error variable with different standard deviations, noise ~ $N(0, \sigma_n)$. We created slightly noisy data sets with $\sigma_n = 0.05$ (seconds), moderately noisy data sets $\sigma_n = 0.2$, and substantially noisy data sets with $\sigma_n = 0.4$. For each of the three noise conditions, this normally distributed error variable was randomly sampled for each trial of the simulated data sets and added to the simulated response time in this trial, leaving response frequencies to the thresholds unaffected. All in all, we simulated 240,000 data sets of response times and accuracies.

2.1.2 Estimation of diffusion model parameters

Diffusion models were fitted to the data sets with *fast-dm-30* (Voss & Voss, 2007; Voss, Voss, & Lerche, 2015). For each of the 240,000 data sets, we estimated a full seven-parameter model (v, a, z, t_0, s_v, s_z, s_{t0}) and a four-parameter model (v, a, z, t_0) in which the inter-trial variability parameters (s_v, s_z, s_{t0}) were fixed to zero. The χ^2 statistic was used as optimization criterion with responses for each decision threshold grouped into six bins that were defined by the .1, .3, .5, .7, and .9 quantiles of the empirical RT distribution.

2.1.3 Evaluation of the RMSEA

We computed corresponding RMSEA values for each minimized χ^2 value. Then, we evaluated how RMSEA values were affected by four factors: Number of trials (100 vs. 500 vs. 1000), degree of measurement noise (no noise vs. slight vs. moderate noise vs. substantial noise), number of parameters in the generator model (four vs. seven), and number of estimated parameters (four vs. seven). We compared the number of rejected models when evaluating model fits with the RMSEA (rejection when $\varepsilon > .10$) to the number of rejected models when evaluating model fits with the χ^2 statistic (rejection when $p < .05$). Moreover, we computed the average RMSEA and the number of well-fitting models ($\varepsilon < .05$) for each of these combinations.

2.2 Results

First, we compared how many models were rejected with the RMSEA as a goodness-of-fit index ($\varepsilon > .10$) and with the χ^2 statistic ($p < .05$). Rejection rates for all models are shown in Figure 1. When *no or slight measurement noise* was added to the data and the *seven parameter model* was estimated, both decision criteria were largely invariant to increasing trial numbers and rejection rates were low. In comparison, when the *four parameter model* was estimated, rejection rates increased with increasing trial numbers for the χ^2 statistic, while the RMSEA was largely invariant to an increase in trial numbers above 500. Moreover, we observed a decrease in rejection rates based on the RMSEA from 100 to 500 and from 100 to 1,000 trials. There was no difference in rejection rates based on the RMSEA between 500 and 1,000 trials.

The difference in rejection rates between the two decision criterions increased with increasing measurement noise. When *moderate and strong measurement noise* was added to the data and the *seven parameter model* was estimated, we observed a huge increase in rejection rates based on the χ^2 statistic, whereas rejection rates based on the RMSEA were largely invariant against an increase in trial numbers. In comparison, when the *four parameter model* was estimated, we found high rejection rates that were relatively invariant to the number of trials for both statistics.

Taken together, we observed an interaction between the number of estimated parameters (four vs. seven), the degree of measurement noise, and the number of trials on the discrepancy between rejection rates based on the two statistics. Rejection rates based on the χ^2 statistic increased with increasing trial numbers only when no or slight measurement noise was added to the data and the four parameter model was estimated, or when moderate or strong measurement noise was added to the data and the seven parameter model was estimated. In these cases the RMSEA was superior to the χ^2 statistic, because rejection rates based on the RMSEA did not increase with increasing trial numbers. In all other cases there were bottom or ceiling effects that prevented any discrepancy between the two criteria.

In the second step, we evaluated if average RMSEA values for well-fitting diffusion models are comparable in magnitude to well-fitting RMSEA values in the SEM framework. When *no measurement noise* was added to the data, average RMSEA values were ≤ 0.05 and the number of well-fitting models was great. Table 2 and Table 3 show the average RMSEA values, percentage of well-fitting models ($\varepsilon < .05$) and the percentage of poorly fitting models ($\varepsilon >.10$) for all models. As expected, models fitted best when the number of estimated parameters corresponded to the number of simulated parameters (average $\varepsilon \leq 0.02$, 89.8% models with $\varepsilon < 0.05$, 2.7% models with $\varepsilon > 0.10$). When only *four parameters* were estimated for data generated with the *seven parameter model*, fits deteriorated notably (average $\varepsilon \leq 0.05$, 64.2% models with $\varepsilon < 0.05$, 18.4% models with $\varepsilon > 0.10$). In comparison, data generated with the *four parameter model* could be well described by the *seven parameter model* (average $\varepsilon \leq 0.02$, 91.3% models with $\varepsilon < 0.05$, 1.7% models with $\varepsilon > 0.10$).

When *slight measurement noise* was added to the RT distributions, RMSEA values were still in an acceptable range (average $\varepsilon \leq 0.08$, 69.2% models with $\varepsilon < 0.05$, 17.2% models with $\varepsilon > 0.10$) for all models. When *moderate or strong measurement noise* was added to the RT distributions and *seven parameters* were estimated, RMSEA values were still acceptable for data generated from both models (average $\varepsilon \leq 0.08$, 56.0% models with $\varepsilon < 0.05$, 21.0% models with $\varepsilon > 0.10$), but they were no longer indicating acceptable fit when *four parameters* were estimated regardless of the generator model ($0.13 \leq$ average $\varepsilon \leq 0.19$, 7.8% models with $\varepsilon < 0.05$, 84.5% models with $\varepsilon > 0.10$).

All in all, as long as no large amounts of measurement noise were added to the data and an adequate model was fitted to the data, RMSEA values were low and comparable in magnitude to values typically observed for well-fitting models in the SEM framework.

In the third step, we computed critical cut-off values indicating good model fit by choosing the RMSEA value at which only 5% and 10% of the correct models were incorrectly rejected. For this purpose, we considered only the models with only sampling noise where the number of estimated parameters corresponded to the number of simulated parameters. As can be seen from Table 4 and from the full distribution of RMSEA values in the six correct models in Figure 2, cut-off values ranged from 0.04 to 0.13 for the *four parameter model* and from 0.03 to 0.10 for the *seven parameter model*.

2.3 Discussion

In the first simulation study, we simulated data sets in the diffusion model framework and estimated the corresponding RMSEAs of the re-fitted models. First, we compared the rejection rates based on the RMSEA with the rejection rates based on a significant χ^2 statistic as a decision criterion in order to show that the χ^2 statistic is sometimes not an appropriate decision criterion as it is not invariant with regard to trial numbers and that rejection decisions improve by using the RMSEA instead. We observed that rejection decisions based on the χ^2 statistic varied with variation in trial numbers only when no or slight measurement noise was added to the data and the four parameter model was estimated, or when moderate or strong measurement noise was added to the data and the seven parameter model was estimated.

The dependency of the χ^2 statistic on trial numbers and the number of estimated parameter makes sense when we compare the absolute model fit for the two models: The seven parameter model accounts for the data extremely well and therefore the discrepancy between the empirical and the estimated RT distributions is so small that even when trial numbers are large there are only few models for which the empirical χ^2 value exceeds the critical χ^2 value. Because the majority of χ^2 values were very small, we did not observe an effect of increasing trial numbers on rejection rates in the conditions with no and little measurement noise; instead, we only observed this effect when moderate and substantial measurement noise was added to the data and the empirical χ^2 values were no longer very small. In contrast, the four parameter model is not as powerful as the seven parameter model when describing the data, and therefore we observed rejection rates dependent on the trial number already when little measurement noise was added to the data or when the data was generated by the seven parameter model (note that we also did not observe any dependency on trial numbers when the four parameter model was fitted to the data with only sampling noise generated by the same model). When the amount of measurement noise was too high to be accounted for in the four parameter model, we observed a ceiling effect with rejection rates close to 100% for both the RMSEA and the χ^2 statistic.

All in all, the RMSEA can only play to its strength regarding its invariance to trial numbers when model fits are neither perfect nor terrible, because in the case of perfect or terrible model fit the χ^2 statistic is likely to accept or reject the majority of models regardless of trial numbers. We believe, however, that most real-life applications of the diffusion model are somewhere in between these two extremes, because models fitted to empirical data reach hopefully acceptable or good, but hardly perfect model fit. As long as model fit in empirical applications is only satisfactory but not perfect, the χ^2 value will increase with increasing trial numbers, while the RMSEA allows for decisions independent of trial numbers and would thus be the better choice.

Next, we compared the magnitude of RMSEA values for well-fitting models with typically observed RMSEA values in the SEM framework. When the number of estimated parameters corresponded to the number of simulated parameters and no or only very little measurement noise was added to the data, RMSEA values for all trial numbers were low and comparable to those typically observed for well-fitting models in the SEM framework, where values smaller than .05 are considered to indicate good fit (e.g., Browne & Cudeck, 1993; Steiger, 1989).

Finally, we aimed to establish critical cut-off values indicating good model fit in the diffusion model framework. RMSEA values at which only 5% of the correct models were incorrectly rejected were all ≤ .06 for larger trial numbers, suggesting that a cut-off criterion of $\varepsilon = .06$ for good fit is appropriate in the diffusion model framework. This value is very close to the cut-off value of $\varepsilon = .05$ that is typically used in the SEM framework. When the trial number was small, the computed cut-off values were notably higher.

This increase in RMSEA values for the models with only 100 trials may be due to two reasons: First, the χ^2 optimization criterion used for parameter estimation struggles to recover the correct parameter values when trial numbers are small (Voss et al., 2013; White, Ratcliff, Vasey, & McKoon, 2010). Due to the binning procedure, information about the RT distribution is aggregated and thus some information is inevitably lost, which is especially problematic if error rates are low. Because the CDF is compared with the number of observed responses per bin separately for the upper and the lower threshold, there will be only very few (if any) RTs in each bin of the error distribution. This leads to an inaccuracy in the identification of the empirical quantiles and thus to a biased estimation of model parameters. Second, simulation studies in the SEM framework have shown that the RMSEA too often rejects models when both the degrees of freedom and the sample size are small (Chen, Curran, & Bollen, 2008; Curran; Bollen, Chen, Paxton, & Kirby, 2003; Kenny et al., 2014).

One key characteristic of the RMSEA is that it considers model parsimony in the evaluation of model fit and rewards parsimonious models. Therefore, the four parameter model should fit better to data generated with the four parameter model than the less parsimonious seven parameter model. We did, however, not observe any difference in average RMSEA values regardless of which of the two models was fitted to the data generated with the four parameter model. Moreover, the percentage of accepted and rejected models differed only by a small amount. This result suggests that the RMSEA may not be able to punish the less parsimonious seven parameter model sufficiently.

In Study 1, we could show that RMSEA parameter values in the diffusion model framework are comparable to the parameter values typically observed in the SEM framework. Moreover, we showed that the RMSEA was invariant with regard to increasing trial numbers, while the χ^2 statistic sometimes tended to reject too many models when the trial number was high. Because all response time data in Study 1 was simulated from diffusion models, the data could be described well by the diffusion model even when some measurement noise was added. However, this measurement noise was added to assess how a general decrease in the signal-to-noise ratio in response times affected the RMSEA, but it did not correspond to a specific process contamination scenario. Moreover, the simulation scenarios only included a single condition, but multiple condition scenarios are prevalent in the literature outside the area of individual differences research. In Study 2, we hence investigate how the RMSEA performs in more realistic scenarios, e.g., when data are generated from multiple conditions and contaminations affecting the process model are added to the simulated data.

3. Simulation Study 2

The aim of the second simulation study was to evaluate how the RMSEA performs in more realistic scenarios. For this purpose, we simulated data sets from the diffusion model with no parameters, drift rate or boundary separation varying between conditions, reflecting specific experimental manipulations. Moreover, we simulated two process contamination scenarios reflecting random guessing and distraction. We then re-fitted these data sets and computed the corresponding RMSEA values and χ^2 statistics. As in Study 1, we a) compared rejection decisions based on the RMSEA to rejection decisions based on the χ^2 statistic, b) compared RMSEA values for well-fitting diffusion models to RMSEA values for well-fitting models in the SEM framework, and c) computed cut-off values indicating good model fit. We expected RMSEA values to behave similarly across single and multiple condition models, and to stay robust in the process contamination scenarios

3.1 Method

3.1.1 Data simulation

For Study 2, we simulated data sets with two experimental conditions. We generated 1,000 random diffusion model parameter sets (v, a, z, t_0, s_v, s_z, s_{t0}) from uniform distributions in the range of typically observed parameters values (see Table 1) with none of these parameters varying between conditions. In addition, we simulated 1,000 random diffusion model parameter sets (v_1~U(0,2), v_2~U(2,4), a, z, t_0, s_v, s_z, s_{t0}) with drift rate varying between two conditions, and 1,000 random diffusion model parameter sets (v, a_1~U(0.5,1.25), a_2~U(1.25,2), z, t_0, s_v, s_z, s_{t0})

with boundary separation varying between two conditions. For each of these parameter sets, we simulated three data sets of response times and accuracies with different numbers of trials per condition (50; 250; 500). All data sets were simulated with the *construct-samples* routine from *fast-dm* by sampling random data sets from a multinomial distribution defined by the diffusion model. These are the uncontaminated data sets.

Next, we simulated two process contamination scenarios. For the *fast guessing contamination*, we randomly selected 5% of the trials in each uncontaminated data set and changed response frequencies so that each threshold was reached with a 50% probability. Moreover, we changed the response times of these trials to response times sampled from a uniform distribution in a range from 50 to 250 ms. For the *distraction contamination*, we randomly selected 5% of the trials in each uncontaminated data set and added a random delay that was uniformly distributed in a range from 10 to 2,000 ms. All in all, we simulated 27,000 data sets of response times and accuracies.

3.1.2 Estimation of diffusion model parameters

Diffusion models were fitted to the data sets with *fast-dm-30*. For each of the 27,000 data sets, we fitted a diffusion model that had the same parameters as the data generating model of the uncontaminated data set. When a parameter varied between conditions in the generating model, a common diffusion model was estimated for the data from both conditions with the specific parameter allowed to vary between conditions. The χ^2 statistic was used as optimization criterion.

3.1.3 Evaluation of the RMSEA

We computed corresponding RMSEA values for each minimized χ^2 value. Then, we evaluated how RMSEA values were affected by four factors: Number of trials per condition (50 vs. 250 vs. 500), data contamination (uncontaminated vs. fast guessing vs. distraction), and parameters varying between conditions (none vs. drift rate vs. boundary separation).We compared the number of rejected models when evaluating model fits with the RMSEA (rejection when $\varepsilon > .10$) to the number of rejected models when evaluating model fits with the χ^2 statistic (rejection when $p < .05$). Moreover, we computed the average RMSEA and the number of well-fitting models ($\varepsilon < .05$) for each of these combinations. Prior to calculating the average RMSEAs, we discarded any outlier RMSEA values > 1 to prevent a distortion of mean values.

3.2 Results

First, we compared how many models were rejected with the RMSEA as a goodness-of-fit index ($\varepsilon > .10$) and with the χ^2 statistic ($p < .05$). Rejection rates for all models are shown in Figure 3. For the uncontaminated data we found that rejection rates based on the χ^2 statistic tended to increase with increasing trial numbers per conditions. In comparison, rejection rates based on the RMSEA were close to zero even at larger trial numbers with a small peak (5.1% - 6.4%) in rejection rates when there were only 50 trials per condition. This pattern was consistent across the single and multiple condition models.

When data were contaminated with fast guessing trials, rejection rates increased for both decision criteria. They increased more strongly for rejection decisions based on the χ^2 statistic than for the RMSEA, and they increased more strongly when a model parameter varied between conditions and when trial numbers per condition were larger. *When data sets contaminated with distraction trials*, we observed a small increase in rejection rates for the χ^2 statistic with larger trial numbers, whereas we observed no increase in rejection rates based on the RMSEA in comparison to the uncontaminated data.

In the second step, we evaluated if average RMSEA values for well-fitting diffusion models are comparable in magnitude to well-fitting RMSEA values in the SEM framework. For the *uncontaminated data*, average RMSEA values were ≤ 0.03 and the number of well-fitting models was high. Table 5 shows the average RMSEA values, percentage of well-fitting models ($\varepsilon < .05$) and the percentage of poorly fitting models ($\varepsilon >.10$) for all models. As expected, models fitted best *when data were uncontaminated* (average $\varepsilon \leq 0.02$, 86.9% models with $\varepsilon < 0.05$, 2.6% models with $\varepsilon > 0.10$), but also *when data were contaminated with distraction trials* (average $\varepsilon \leq 0.03$, 81.6% models with $\varepsilon < 0.05$, 3.5% models with $\varepsilon > 0.10$). *When data were contaminated with fast guessing trials*, about half of the models were still identified as well-fitting ($0.05 \leq$ average $\varepsilon \leq 0.13$, 41.6% models with $\varepsilon < 0.05$, 40.5% models with $\varepsilon > 0.10$) with slightly more badly-fitting models when a model parameter was estimated separately per condition.

In the third step, we computed critical cut-off values indicating good model fit by choosing the RMSEA value at which only 5% and 10% of the correct models were incorrectly rejected (see Figure 4 for the whole distribution of RMSEA values). 5% cut-off values ranged from 0.03 to 0.11 for the uncontaminated data, from 0.23 to 0.31 for the data contaminated with fast guessing trials, and from 0.05 to 0.12 for the data contaminated with distraction trials as shown in Figure 4 and Table 6. Unexpectedly, cut-off values were slightly larger when boundary separation varied between conditions than when no parameter or drift rate varied between conditions.

3.3 Discussion

In the second simulation study, we simulated data corresponding to several realistic scenarios, such as the experimental manipulation of diffusion model parameters and two different contamination scenarios. First, we evaluated how many models were rejected with the RMSEA in comparison to the χ^2 statistic as a decision criterion. As in Study 1, the RMSEA was relatively invariant with regard to trial numbers, but it tended to reject too many models when the trial number was as low as 50 trials per condition. When the χ^2 statistic was used as a decision criterion, rejection rates again increased with increasing trial numbers per condition. This dependency on trial numbers was amplified when the data were contaminated with outliers. Thus, the results of Study 2 illustrate that the RMSEA is better suited to evaluate the goodness-of-fit than the χ^2 statistic if trial numbers are not extremely small.

Estimating multiple conditions simultaneously had little effect on average RMSEA values and rejection rates. Overall, average RMSEA values were smaller than the critical value of .05 typically considered indicating good fit (e.g., Browne & Cudeck, 1993; Steiger, 1989). The number of well-fitting models was surprisingly higher when one model parameter varied between conditions at 50 trials per condition than when no parameter varied between conditions. At higher trial numbers, the number of well-fitting models was similarly high when no parameter or when drift rate varied between conditions, and decreased slightly when boundary separation varied between conditions. This pattern of results also emerged for the suggested cut-off values, which were all smaller than or equal to .06 for larger trial numbers (≤ 0.12 for all trial numbers) when no parameter or drift rate varied between conditions, replicating the results regarding cut-off values in Study 1. When boundary separation varied between conditions, cut-off values were slightly higher (≤ 0.10 for larger trial numbers, ≤ 0.11 for all trial numbers). Overall, these results indicate that the RMSEA allows a stable evaluation of goodness-of-fit even when there is an experimental manipulation of model parameters and some model parameters are free to vary between conditions.

The two contamination scenarios had different effects on the RMSEA: While the fast guessing contamination led to deterioration in fits and to an increase in rejection rates up to between 20 and 60 percent, we observed only marginal effects on model fits for the distraction contamination. Similarly, 5% cut-off values were far higher in the fast guessing contamination condition (≥ 0.18) than in any of the other conditions (≤ 0.12). There were no systematic interactions between the variation of parameters across conditions and the effects of contamination. All in all, these results highlight that long outlier response times due to distraction need not be specifically considered when evaluating the goodness-of-fit with the RMSEA, which is an important result, because slow outliers cannot be easily identified by typical outlier detection procedures as they can hardly be distinguished from the normal tail of the response time distribution (Ratcliff, 1993; Ratcliff & Tuerlinckx, 2002).Very fast, anticipatory response times that occur before the diffusion process is terminated, however, affect goodness-of-fit evaluations. Therefore, we recommend applying standard outlier detection techniques before evaluating model fit with the RMSEA, and adhering to the suggestions by Ratcliff and Tuerlinckx (2002) considering the elimination of fast outliers.

In Study 2, we could show that the RMSEA is suited to evaluate goodness-of-fit in the diffusion model framework when model parameters are manipulated experimentally. Moreover, we evaluated how different contamination scenarios influenced model evaluation. One aspect of model evaluation that we have not analyzed yet is how the RMSEA performs at model selection, i.e. at selecting the true model (or the one with the highest predictive validity) out of a set of alternate models. In Study 3, we will therefore compare model selection performance of the RMSEA to two very popular model fit indices, the AIC and the BIC.

4. Simulation Study 3

The aim of the third simulation study was to evaluate how the RMSEA performs in comparison to the AIC and BIC at model selection. For this purpose, we simulated data sets from the diffusion model with no parameters, or drift rate, boundary separation, or drift rate and boundary separation varying between conditions, reflecting specific experimental manipulations. We then fitted different diffusion models to these data sets that were either the generating model or under- or over-parameterized models. In order to compare the different indices in terms of model selection, we evaluated how often each index identified the generating model as the correct model. Because the RMSEA may not be able to reward parsimony sufficiently when the overall model fit is very high, we predicted that it would identify the true model less often than the AIC and BIC.

4.1 Method

4.1.1 Data simulation

Data sets were simulated the same way as in Study 2 with the exception that no contamination scenarios were added. Moreover, we simulated additional 1,000 random diffusion model parameter sets (v_1~U(0,2), v_2 ~U(2,4), a_1~U(0.5,1.25), a_2~U(1.25,2), z, t_0, s_v, s_z, s_{t0}) with drift rate and boundary separation varying between the two conditions and simulated 3,000 data sets of response times and accuracies with varying trial numbers per condition (100; 250; 500) from these. All in all, we simulated 12,000 data sets of response times and accuracies with no parameter, or drift rate, boundary separation, or drift rate *and* boundary separation varying between the two conditions, and either 50, 250, or 500 trials per condition.

4.1.2 Estimation of diffusion model parameters

Diffusion models were fitted to the data sets with *fast-dm-30*. For each of the 12,000 data sets, we fitted four different diffusion models: One with no parameter varying between conditions, one with drift rate varying between conditions, one with boundary separation varying between conditions, and one with both drift and boundary separation varying between conditions. The maximum likelihood method was used as the optimization criterion.

4.1.3 Evaluation of model selection performance

We considered estimated models that had the same parameters as the corresponding generating model as correct models. We compared each correct estimated model to estimated models in which more (over-parameterized models) or less (under-parameterized models) parameters vary between conditions than in the correct model by calculating the likelihood ratio (please see Table 7 for an overview over the different model selection scenarios). We then calculated the AIC, BIC, and RMSEA based on the likelihood ratio to decide whether the more complex model provided a better model fit than the less complex one. If AIC or BIC values were larger than ten we took this as evidence that the more complex model provided a substantially better description for the data (Burnham & Anderson, 2002; Raftery, 1995). For the RMSEA, we used cut-off values dependent on trial numbers as suggested by Studies 1 and 2, i.e. $\varepsilon > .10$ for 50 trials per condition, $\varepsilon > .05$ for 250 trials per condition, and $\varepsilon > .03$ for 500 trials per condition. Thus, we assessed the number of correct decisions (i.e., decisions for the model with parameters corresponding to the generating model) as a function of specific model parameter varying between conditions (none vs. drift rate vs. boundary separation vs. drift rate and boundary separation), misspecification (over-parameterized vs. under-parameterized), and trial numbers per condition (50 vs. 250 vs. 500).

4.2 Results

As shown in Figure 5, the general pattern of results was the same across all three fit indices. The correct model was identified as the best-fitting model in most cases irrespective of trial numbers per condition when the alternative model was over-parameterized, i.e. when more parameters were allowed to vary between conditions in the alternative than in the generating model. In contrast, when the correct model was compared to an under-parameterized alternative, all three fit indices only decided in favor of the correct model when trial numbers per condition were small, whereas they tended to decide in favor of the under-parameterized model when trial numbers got larger.

When the correct model was compared to over-parameterized alternatives, the RMSEA (on average: 85.2 %) identified fewer correct models than the AIC (on average: 97.6 %) and BIC (on average: 98.1 %). When the correct model was compared to under-parameterized alternatives, however, the RMSEA (on average: 38.6 %) identified more correct models than the AIC (on average: 19. 6%) and the BIC (on average: 16.1 %).

4.3 Discussion

In Study 3, we evaluated how the RMSEA performs in comparison to the AIC and BIC at model comparison in different scenarios. Overall, all three fit indices preferred more parsimonious models regardless of trial numbers and model specifications. Only at small trial numbers (50 per condition, 100 in total) was the correct generating model with model parameters varying between conditions favored instead of a less complex, under-parameterized model. Because this result was consistent across all fit indices and did not differ for the AIC, which is independent of trial numbers, it suggests that this divergence at small trial numbers was likely due to the maximum likelihood optimization at the model fitting stage.

Contrary to what we expected based on its performance at model selection in structural equation modeling (Bollen et al., 2014), the RMSEA performed relatively well at model selection. Our more favorable evaluation may be due to the fact that the we calculated it based on the likelihood ratio χ^2, whereas Bollen et al. (2014) calculated it separately for each model based on the asymptotically χ^2 distributed value of the discrepancy function and chose the model with the smallest RMSEA value as the best-fitting one. Therefore, the problems associated with using an absolute index of model fit with a fixed minimum value for model selection may not be relevant when this index is calculated based on the likelihood ratio χ^2, and thus becomes larger if the more complex model is better suited to describe the data in comparison to the less complex one.

Across all model selection cases and trial numbers, the RMSEA identified the correct model in 61.8 % of the cases, which is a higher success rate than either the AIC (58.6 %) or thee BIC (57.2 %) had. Nevertheless, although the RMSEA did unexpectedly well at model selection, it was always outperformed by the AIC and BIC when the correct model was compared to an over-parameterized alternative. Therefore, we would only suggest using it in addition to established fit indices for model selection, but not as a replacement for them.

5. General Discussion

The statistical evaluation of model fit is one of the greatest challenges in the application of diffusion modeling in a wider context. While model comparison is straightforward thanks to several relative fit indices, the rejection of specific models or the exclusion of specific participants is more ambiguous.

As we have shown in two simulation studies, the χ^2 statistic rejected more models with increasing trial numbers irrespective of model fit. In the first simulation study we found that rejection rates based on the χ^2 statistic increased substantially with increasing trial numbers as soon as some measurement noise was added to the data. In the second simulation study we could show a comparable increase in rejection rates with increasing trial numbers when diffusion models where fitted to realistic response time data. Therefore, the χ^2 statistic is clearly not an ideal candidate for an objective evaluation of model fit due to its dependency on trial numbers.

Thus, we suggest that instead of relying solely on graphical tests or on the χ^2 statistic for the evaluation of model fit, the evaluation of model fit should be supplemented with the RMSEA. The RMSEA is one of the most popular and most frequently used GOF indices in structural equation modeling. In two simulation studies we showed that the RMSEA can be applied successfully in the diffusion model framework and that the observed RMSEA values were comparable to the values typically observed in structural equation modeling for well-fitting models.

Across both studies we found that acceptance and rejection decisions based on the RMSEA were largely invariant to trial numbers, which is one of the key benefits of the RMSEA as a GOF index in comparison to statistical tests of model fit. In particular, where the χ^2 statistic indicated an incorrect rejection of a large amount of well-fitting models when the trial number was large, we observed no such increase for the RMSEA. Moreover, model selection decisions in the third simulation study were also largely invariant with regard to trial numbers with the exception that decision accuracies decreased with increasing trial numbers when the correct model was a more complex model that was compared to under-parameterized alternatives.

The second key property of the RMSEA is that it rewards parsimonious models. We did, however, not observe the RMSEA favoring the more parsimonious four parameter model over the seven parameter model in our first simulation study. This leads us to the conclusion that whenever the inter-trial variability parameters are estimated, model fit increases to such a great degree that the RMSEA is not able to compensate for the loss in model parsimony. Because the RMSEA accounts for model parsimony by considering the degrees of freedom, a comparatively small loss of degrees of freedom in comparison to a large increase in absolute model fit must inevitably lead to an enhancement of model fit. Moreover, because the RMSEA is an absolute index of model fit with a fixed minimum value of zero, a model nested within the perfectly fitting generating model will also always have a perfect fit. Note that the observed improvement of model fit based on the intra-trial-variability parameters can also lead to a spurious acceptance of models. This is evident from the noise conditions in Study 1. When data was generated with a four parameter model, the seven parameter model was able to fit the data even if there was considerable measurement noise added to the data. Thus, the seven parameter model accounted well for these extremely noisy data according to the RMSEA, although the generating process for these data was no longer the diffusion model. It has to be tested in further studies whether the parameter estimation is biased if such measurement noise is mapped by the inter-trial variability parameters.

In the third simulation study, we evaluated how the RMSEA performs at model selection when calculated based on the likelihood ratio of two nested models. In the majority of cases we observed that the RMSEA strongly favored the more parsimonious model over the nested alternative. Even when the generating model was the more complex model, the RMSEA often decided incorrectly in favor of the more parsimonious alternative one. This strong reward of parsimony was, however, not unique to the RMSEA: We observed the same preference for parsimonious models even more strongly for the AIC and BIC.

5.1 Conditions for a successful application of the RMSEA in the diffusion model framework

The results of the first two simulation studies suggest that the application of the RMSEA makes most sense when trial numbers are not extremely small. Both the χ^2 optimization criterion and the RMSEA itself are known to behave not optimally when trial numbers and the degrees of freedom are small (Chen et al., 2008; Bollen et al., 2003; Kenny et al., 2014, Voss et al., 2013, White et al., 2010). It is important to note that the degrees of freedom entering the computation of the RMSEA are often rather small in the diffusion model framework in comparison to the typical degrees of freedom resulting in the SEM framework. Because the degrees of freedom in structural equation modeling do not only depend on the specified model, but also on the number of sample moments, a multitrait-multimethod model may easily have over a hundred degrees of freedom (e.g., Marsh & Bailey, 1991), although latent growth curve or specific path models may have fewer degrees of freedom (e.g. Curran, 2000; Williams & Hazer, 1986). In comparison, degrees of freedom ranged from four in the first simulation study when no parameter varied between conditions and inter-trial variabilities were estimated to fourteen in the second simulation study when drift rate or boundary separation were allowed to vary between two conditions. In the first simulation study, the RMSEA tended to reject too many correct models when there were only 100 trials in each simulated data set. In comparison, when the trial number was ≥ 500, none of the correct models were rejected. Therefore we suggest that the RMSEA should only be used as a GOF index when the trial number is sufficiently high. In the case of small trial numbers, it might be a good idea to adapt the numbers of bins used for the computation of χ^2. For example, if there are fewer than twelve error trials, the calculation of χ^2 – and, thus, of the RMSEA – will become more stable if only one bin for all errors is used (Ratcliff & Childers, 2015).

Moreover, the estimation of the RMSEA presupposes the estimation of the χ^2 statistic. Therefore, either the χ^2 statistic has to be used directly as an optimization criterion in the modeling process, or it has to be computed post-hoc after the modeling process if a different optimization criterion was used. In the latter case, the χ^2 statistic needs to be estimated as the deviation of the predicted from the observed response time distributions given the model parameters optimized by a different optimization criterion.

5.2 Cut-off values

One aim of the present study was to determine appropriate cut-off values for the application of the RMSEA in the diffusion model framework. Cut-off values are always arbitrary and should only be treated as heuristics when evaluating model fit. We do, however, believe that they can also provide a valuable first orientation when deciding about the acceptance or rejection of a model. To recommend appropriate cut-off values, we considered only those simulation results where the trial number was ≥ 500 for the reasons explained above. Moreover, we decided to suggest different cut-off values for models with and without inter-trial variabilities due to the huge discrepancy in model fits. Based on the first and second simulation study, we suggest that RMSEA values ≤ 0.05 indicate good model fit for all kinds of models. For models without inter-trial variabilities, we consider RMSEA values of ≤ 0.14 to indicate acceptable model fit. For models with inter-trial variabilities, we suggest that RMSEA values of ≤ 0.08 to indicate acceptable model fit. Again, we would like to stress that these cut-off values are only heuristics that need to be tested and refined in future applications of the RMSEA in the diffusion model framework.

5.3 Alternative approaches to the simultaneous analysis of psychometric and modeling data

There are already proposals for the solution of problems associated with the combined analysis of psychometric and modeling data, of which the problem of identifying individuals whose data are not adequately described by the diffusion model is only one problem of many. Vandekerckhove, Tuerlinckx, and Lee (2011) suggested a Bayesian hierarchical framework in which diffusion model parameters can be regressed on psychometric variables. Moreover, Vandekerckhove (2014) developed a Bayesian cognitive latent variable framework for the simultaneous latent analysis of behavioral and personality data that circumvents any problem of measurement or estimation errors by the simultaneous estimation procedure. Both psychometric modeling frameworks are very sophisticated and elegant, but they also require basic programming knowledge and some experience in Bayesian modeling in comparison to easy-to-use existing software solutions such as *EZ-diffusion model* (Wagenmakers, van der Maas, & Grasman, 2007; Wagenmakers, van der Maas, Dolan, & Grasman, 2008; Grasman, Wagenmakers, & van der Maas, 2009), *DMAT* (Vandekerckhove & Tuerlinckx, 2007, 2008), and *fast-dm* (Voss & Voss, 2007; Voss, Voss, & Lerche, 2015). Therefore, we believe that a method of assessing individual model fit outside of a hierarchical Bayesian framework is of great value for researchers resorting to one of these existing software solutions.

5.4 Conclusion

Given the recent rise in popularity of the diffusion model, and given its more frequent application in studies of individual differences (e.g., Ratcliff, Thapar & McKoon, 2010, 2011; Schmiedek et al., 2007), we believe that an objective criterion for the evaluation of goodness-of-fit for individual participants may be useful for a wider audience. As we have shown in two simulation studies, the χ^2 statistic is not suited as an absolute index of model fit, because it tends to reject all models regardless of actual model fit when the trial numbers are large.

Therefore, we suggest supplementing the evaluation of model fit with the RMSEA, which can be calculated easily from the χ^2 statistic and which offers several advantages over other methods: The RMSEA is not only largely invariant to trial numbers, but it also allows to identify cases with poor model fit, to calculate confidence intervals, and to conduct power analyses. For these reasons, it is one of the most popular goodness-of-fit indices used in structural equation modeling. We hope that our demonstrations of its successful application in the diffusion model framework will aid its further dissemination in the modeling community.

6. References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Bollen, K. A., Harden, J. J., Ray, S., & Zavisca, J. (2014). BIC and alternative Bayesian information criteria in the selection of structural equation models. *Structural Equation Modeling*, *21*, 1–19. doi:10.1080/10705511.2014.856691

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *4*, 70–82. doi: 10.3758/s13421-011-0128-6

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Sage focus editions: Vol. 154. Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed). New York: Springer.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, *51*, 661–703. doi:10.1137/070710111

Curran, Patrick J. 2000. A Latent Curve Framework for Studying Developmental Trajectories of Adolescent Substance Use. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate Applications in Substance Use Research* (pp. 1-42). Hillsdale, NJ: Erlbaum.

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite Sampling Properties of the Point Estimates and Confidence Intervals of the RMSEA. *Sociological Methods & Research*, *32*, 208–252. doi:10.1177/0049124103256130

D'Agostino, R. B. (1986). Graphical Analysis. In R. B. D'Agostino & M. A. Stephens (Eds.), *Goodness-of-fit techniques* (pp. 7–62). New York, NY: Marcel Dekker.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models. *Sociological Methods & Research*, *36*, 462–494. doi:10.1177/0049124108314720

Germar, M., Schlemmer, A., Krug., K, Voss, A., & Mojzisch, A. (2014). Social Influence and Perceptual Decision-Making: A Diffusion Model Analysis. *Personality and Social Psychology Bulletin, 40*, 217-231. doi: 10.1177/0146167213508985

Grasman, R. P. P. P., Wagenmakers, E.-J., & van der Maas, H. L. J. (2009). On the mean and variance of response times under the diffusion model with an application to parameter estimation. *Journal of Mathematical Psychology*, *53*, 55–68. doi:10.1016/j.jmp.2009.01.006

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

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–23. doi:10.1037/a0014694

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research.* doi:10.1177/0049124114543236

Klauer, K. C., Stahl, C., & Voss, A. (2011). Multinomial models and diffusion models. In K. C. Klauer, A. Voss, & C. Stahl (Eds.), *Cognitive methods in social psychology* (pp. 367–390). New York, NY, US: Guilford Press.

Lerche, V., Voss, A., & Nagler, M. (submitted). How Many Trials are Required for Robust Parameter Estimation in Diffusion Modeling?

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.

Marsh, H. W., & Bailey, M. (1991). Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models. *Applied Psychological Measurement*, *15*, 47–70. doi:10.1177/014662169101500106

McKoon, G., & Ratcliff, R. (2013). Aging and predicting inferences: A diffusion model analysis. *Journal of Memory and Language*, *68*(3), 240–254. doi:10.1037/t02942-000.

Petrov, B. N., & Caski, F. (Eds.). (1973). *Proceedings of the Second International Symposium on Information Theory*: Budapest: Akademiai Kiado.

Raftery, A. E. (1995). *Bayesian Model Selection in Social Research. Sociological Methodology, 25*, 111-163.

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

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. doi:10.1037/0033-2909.114.3.510

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

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

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

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

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

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaching to dealing with contaminant reaction and parameter variability. *Psychonomic Bulletin & Review*, *9*, 438–481. doi:10.3758/BF03196302

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., & Voss, A. (2014). Components of task switching: A closer look at task switching and cue switching. *Acta Psychologica*, *151*, 184–196. doi:10.1016/j.actpsy.2014.06.009

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

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*, 461–464. doi:10.2307/2958889

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

Steiger, J. H. & Lind, J. (1980). *Statistically-based tests for the number of common factors*. Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City.

Vandekerckhove, J. (2014). A cognitive latent variable model for the simultaneous analysis of behavioral and personality data. *Journal of Mathematical Psychology*, *60*, 58–71. doi:10.1016/j.jmp.2014.06.004

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, *14*, 1011–1026. doi:10.3758/BF03193087

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, *40*, 61–72. doi:10.3758/BRM.40.1.61

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*, 44–62. doi:10.1037/a0021765

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin and Review, 7*, 424-465. doi:10.3758/BF03214357

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion Models in Experimental Psychology. *Experimental Psychology (formerly Zeitschrift für Experimentelle Psychologie)*, *60*, 385–402. doi:10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Brandtstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology*, *44*, 1048–1056. doi:10.1016/j.jesp.2007.10.009

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in associative and categorical priming: A diffusion model analysis. *Journal of Experimental Psychology: General*, *142*, 536–559. doi:10.1037/a0029459

Voss, A., & Schwieren, C. (2015). *The Dynamics of Motivated Perception: Effects of Control and Status on the Perception of Ambivalent Stimuli. Cognition & Emotion, 29*, 1411-1423. doi: 10.1080/02699931.2014.984660

Voss, A., & Voss, J. (2007). Fast-dm: A Free Program for Efficient Diffusion Model Analysis. Behavioral Research Methods, 39, 767-775. doi:10.3758/BF03192967

Voss, A., Voss, J. & Lerche, V. (2015). Assessing Cognitive Processes with Diffusion Model Analyses: A Tutorial based on fast-dm-30. Frontiers in Psychology, 6:336. doi: 10.3389/fpsyg.2015.00336

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

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, *15*, 1229–1235. doi:10.3758/PBR.15.6.1229

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3–22. doi:10.3758/BF03194023

White, C., Ratcliff, R., Vasey, M., & McKoon, G. (2009). Dysphoria and memory for emotional material: A diffusion-model analysis. *Cognition and Emotion*, *23*, 181–205. doi:10.1080/0269993080

Wilks, S. S. (1938). The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. *Ann. Math. Statist.*, 60–62. doi:10.1214/aoms/1177732360

Williams, L. J. & Hazer, J. T. (1986). *Antecedents and Consequences of Satisfaction and Commitment in Turnover Models: A Reanalysis Using Latent Variable Structural Equation Methods. Journal of Applied Psychology, 71*, 219-31.

Table 1

Value ranges for all diffusion model parameters used for data simulation

parameter	minimum	maximum
v	-4.0	4.0
a	0.5	2.0
z	0.3	0.7
t_0	0.2	0.5
s_v	0.0	1.0
s_z	0.0	0.5
s_{t0}	0.0	0.2

Note: These values refer to a diffusion constant (intra-trial-variability) of s=1.

Table 2

Average RMSEA values and percentage of good-fitting models (ε < .05) as a function of number of trials, degree of noise, and number of estimated parameters for the four parameter generator model.

	100 trials				500 trials				1000 trials			
Noise	No	Little	Moderate	Substantial	No	Little	Moderate	Substantial	No	Little	Moderate	Substantial
Estimated model: 4 parameter model												
Mean ε	0.02	0.05	0.13	0.19	0.01	0.06	0.15	0.18	0.01	0.06	0.15	0.18
$\varepsilon < .05$	79.9%	68.0%	33.0%	13.5%	89.3%	53.1%	4.3%	0.0%	97.3%	50.1%	1.0%	0.0%
$\varepsilon > .10$	10.7%	22.9%	55.1%	76.1%	0.2%	25.8%	80.3%	98.7%	0.0%	26.4%	85.3%	99.5%
Estimated model: 7 parameter model												
Mean ε	0.02	0.02	0.03	0.06	0.01	0.01	0.03	0.07	0.01	0.01	0.03	0.08
$\varepsilon < .05$	78.5%	79.2%	70.9%	49.5%	95.8%	95.9%	82.0%	36.2%	99.5%	99.7%	86.6%	31.2%
$\varepsilon > .10$	5.2%	5.1%	10.7%	29.7%	0.0%	0.0%	6.7%	35.4%	0.0%	0.0%	6.9%	36.7%

Table 3

Average RMSEA values and percentages of good-fitting models (ε < .05) and bad-fitting models (ε > .10) as a function of number of trials, degree of noise, and number of estimated parameters for the seven parameter generator model.

	100 trials				500 trials				1000 trials			
Noise	No	Slight	Moderate	Substantial	No	Slight	Moderate	Substantial	No	Slight	Moderate	Substantial
Estimated model: 4 parameter model												
Mean ε	0.04	0.06	0.15	0.19	0.05	0.08	0.16	0.18	0.05	0.08	0.16	0.18
ε < .05	67.1%	60.9%	30.7%	13.1%	61.7%	42.4%	2.7%	0.0%	63.8%	40.0%	0.5%	0.0%
ε > .10	21.3%	29.2%	58.4%	76.8%	17.1%	34.3%	84.7%	98.9%	16.9%	33.9%	88.4%	99.6%
Estimated model: 7 parameter model												
Mean ε	0.02	0.02	0.04	0.07	0.01	0.02	0.03	0.08	0.01	0.01	0.03	0.08
ε < .05	76.0%	76.6%	67.4%	46.4%	96.4%	95.6%	80.3%	29.6%	99.9%	99.6%	85.3%	27.1%
ε > .10	5.4%	5.8%	12.6%	33.2%	0.0%	0.0%	7.8%	40.7%	0.0%	0.0%	7.7%	40.7%

Table 4

Critical cut-off values at which only 5% or 10% of the correct models were incorrectly rejected. The number of estimated parameters corresponded to the number of simulated parameters.

	100 trials	500 trials	1000 trials
Four parameter model			
5% error rate	0.13	0.06	0.04
10 % error rate	0.10	0.05	0.04
Seven parameter model			
5% error rate	0.10	0.05	0.03
10 % error rate	0.08	0.04	0.03

Table 5

Average RMSEA values and percentages of good-fitting models (ε < .05) and bad-fitting models (ε > .10) as a function of number of trials per condition, data contamination, and number of parameters varying between conditions

	50 trials			250 trials			500 trials		
Contamination	none	fast guessing	distraction	none	fast guessing	distraction	none	fast guessing	distraction
				No parameter varies between conditions					
Mean ε	0.02	0.06	0.03	0.01	0.10	0.02	0.01	0.10	0.02
$\varepsilon < .05$	75.6%	62.9%	76.2%	95.3%	39.7%	87.9%	99.8%	36.8%	94.4%
$\varepsilon > .10$	5.5%	22.6%	6.4%	0.0%	39.2%	0.0%	0.0%	41.8%	0.0%
				v varies between conditions					
Mean ε	0.01	0.05	0.02	0.02	0.13	0.03	0.02	0.12	0.03
$\varepsilon < .05$	86.6%	68.3%	84.6%	91.3%	23.7%	79.9%	97.6%	25.9%	89.7%
$\varepsilon > .10$	5.1%	22.7%	6.4%	0.0%	57.5%	0.0%	0.0%	29.3%	0.0%
				a varies between conditions					
Mean ε	0.02	0.05	0.02	0.03	0.11	0.04	0.03	0.11	0.03
$\varepsilon < .05$	85.3%	69.8%	81.4%	71.5%	25.3%	61.2%	78.8%	21.9%	78.8%
$\varepsilon > .10$	6.4%	23.2%	9.1%	4.1%	50.4%	7.2%	2.3%	51.4%	2.3%

Table 6

Critical cut-off values at which only 5% or 10% of the correct models were incorrectly rejected.

	50 trials			250 trials			500 trials		
Contamination	no	fast guessing	distraction	no	fast guessing	distraction	no	fast guessing	distraction
				No parameter varies between conditions					
5 % error rate	0.10	0.29	0.11	0.05	0.30	0.06	0.03	0.29	0.05
10 % error rate	0.08	0.18	0.09	0.04	0.23	0.05	0.03	0.24	0.05
				Drift rate varies between conditions					
5 % error rate	0.10	0.26	0.12	0.06	0.31	0.07	0.04	0.28	0.06
10 % error rate	0.07	0.20	0.08	0.05	0.25	0.06	0.04	0.24	0.05
				Boundary separation varies between conditions					
5 % error rate	0.11	0.26	0.12	0.10	0.23	0.11	0.08	0.23	0.08
10 % error rate	0.07	0.20	0.09	0.08	0.21	0.09	0.07	0.20	0.07

Table 7

Overview over the different model selection scenarios.

Parameter varying between conditions in the generating model	**Parameter varying between conditions in the estimated model** None	Drift rate	Boundary seperation	Drift rate, boundary seperation
None	perfect fit	overfit	overfit	overfit
Drift rate	underfit	perfect fit	misfit[1)]	overfit
Boundary seperation	underfit	misfit[1)]	perfect fit	overfit
Drift rate, boundary seperation	underfit	underfit	underfit	perfect fit

[1)] These models were not considered in this model selection study, as they are not nested versions of the perfectly fitting model and can thus not be compared based on the likelihood ratio.

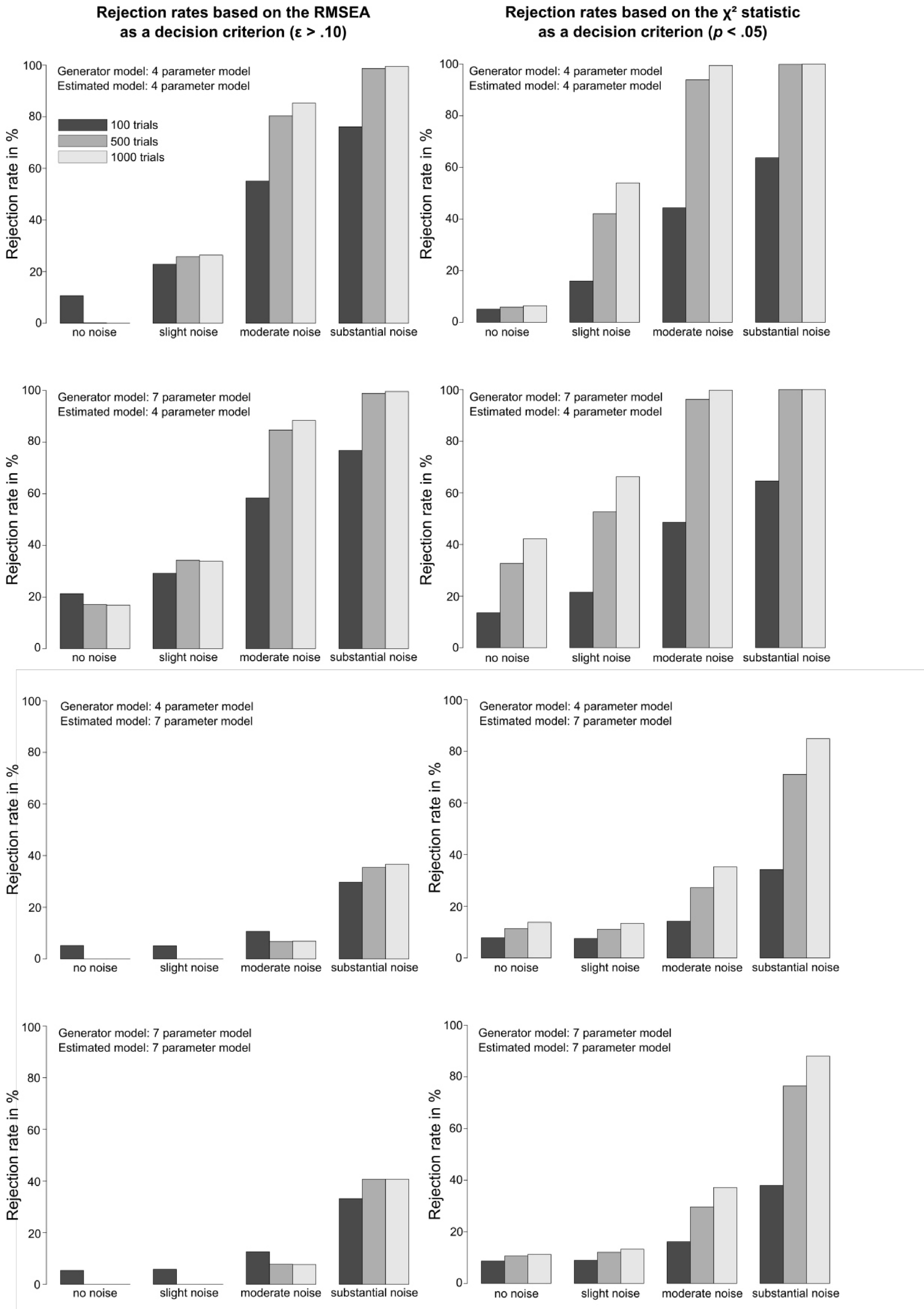

Figure 1. Rejection rates for all models with the RMSEA as a goodness-of-fit index ($\varepsilon > .10$) are presented on the left side and with the χ^2 statistic ($p < .05$) on the right side.

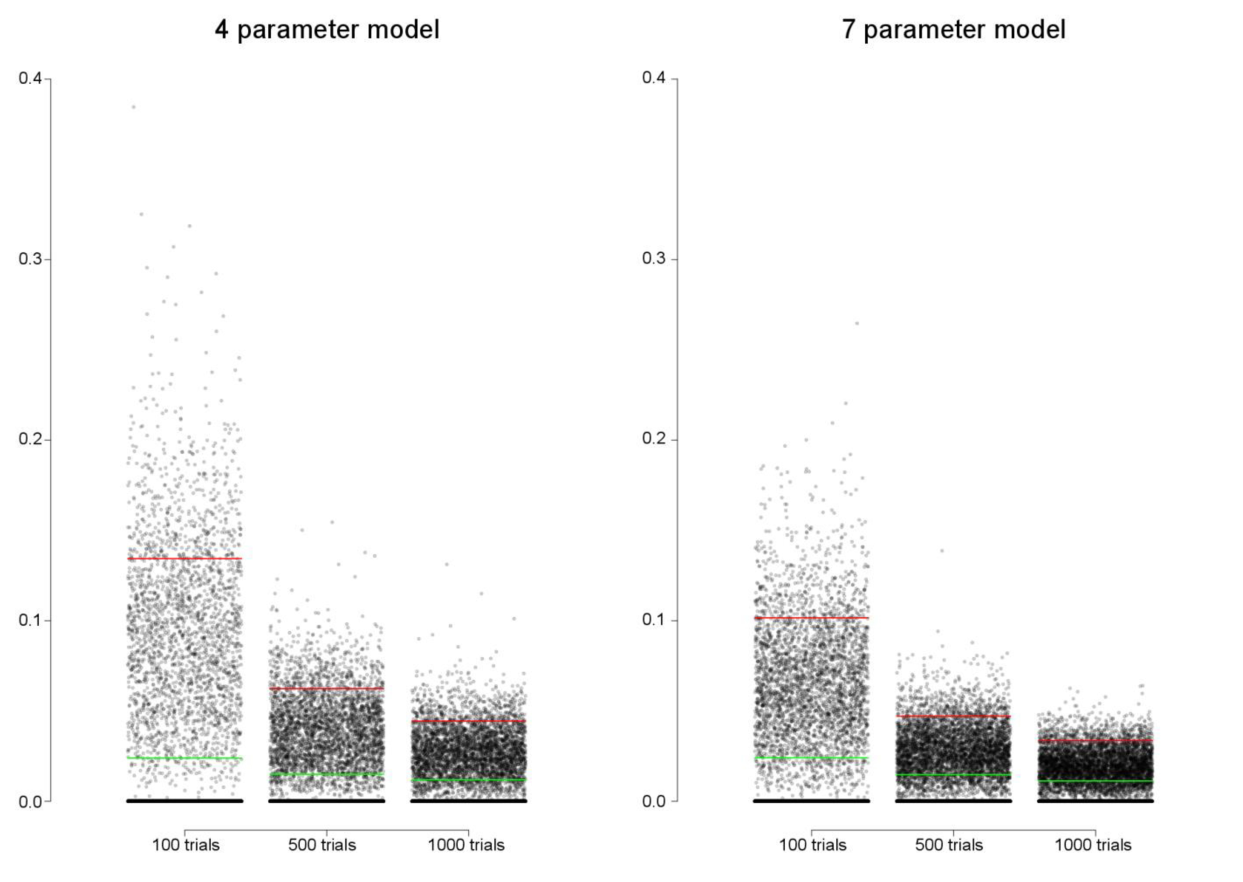

Figure 2. Depicts the full distribution of RMSEA values in the six correct models. Green bars indicate mean RMSEA values and red bars indicate the cut-off value at which only 5% of the correct models were rejected.

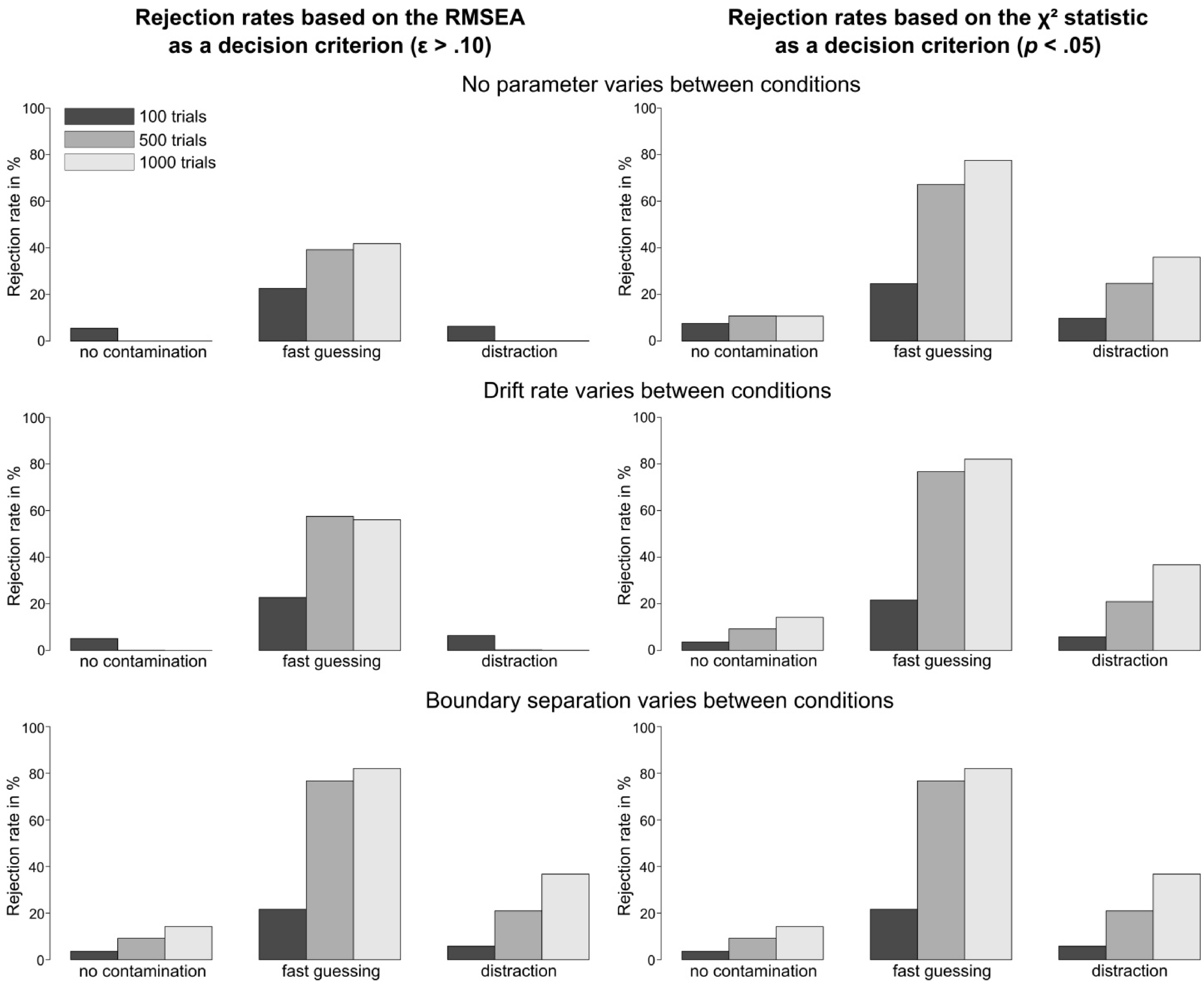

Figure 3. Rejection rates for all models with the RMSEA as a goodness-of-fit index (ε > .10) are presented on the left side and with the χ^2 statistic ($p < .05$) on the right side.

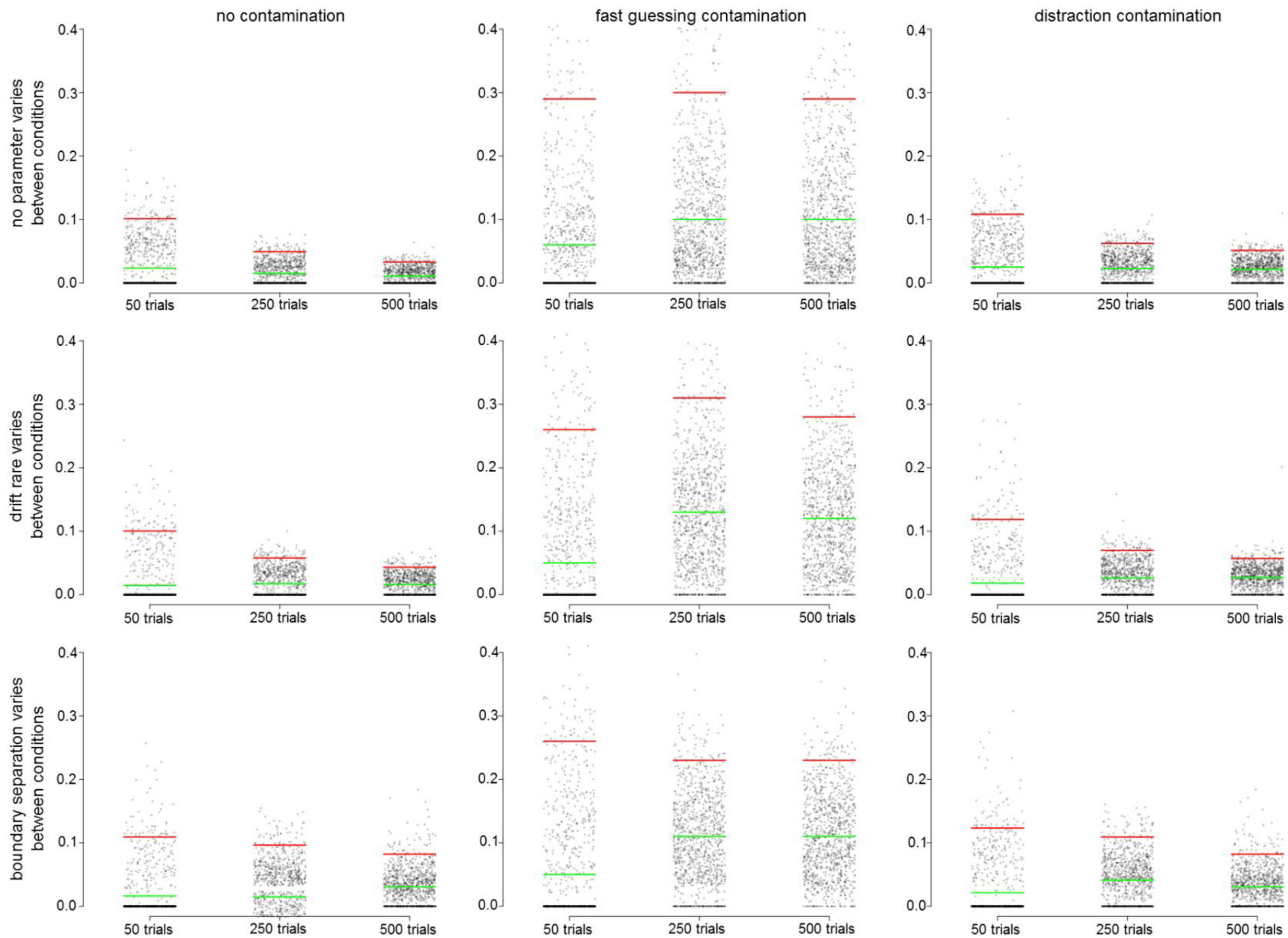

Figure 4. Depicts the full distribution of RMSEA values. Green bars indicate mean RMSEA values and red bars indicate the cut-off value at which only 5% of the correct models were rejected.

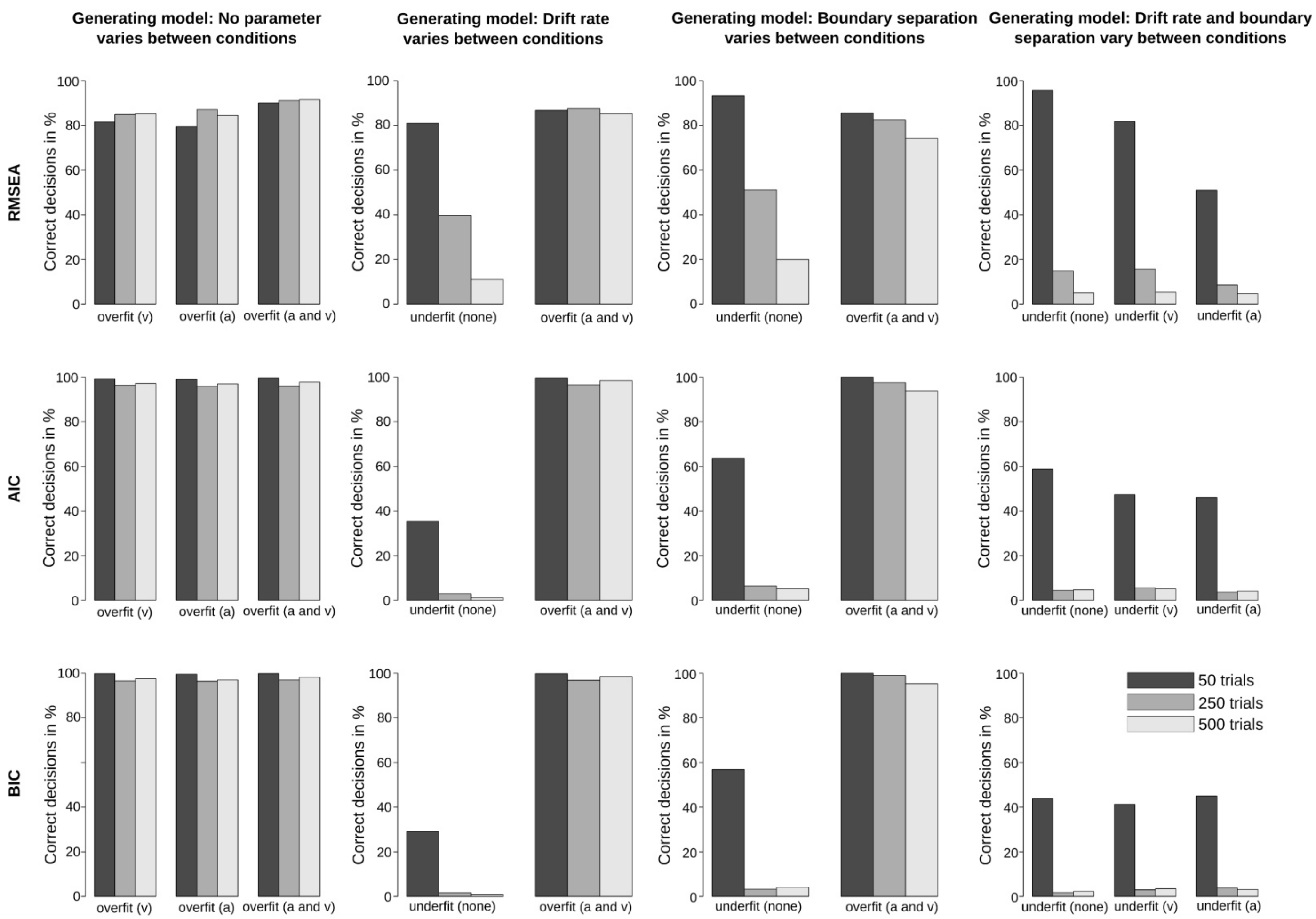

Figure 5. Correct decisions for the generating models in percent with the RMSEA (ε > .10 for 50 trials per condition, ε > .05 for 250 trials per condition, and ε > .03 for 500 trials per condition), the AIC (AIC > 10), and the BIC (BIC > 10) as model selection criteria. “Overfit” refers to over-parameterized models that are compared to the correct model, and “underfit” refers to under-parameterized models that are compared to the correct model. The parameter in parentheses indicates which parameter was allowed to vary between conditions

Appendix A3 – Manuscript 3

General intelligence is little more than the speed of higher-order processing

Anna-Lena Schubert[a], Dirk Hagemann[b], and Gidon Frischkorn[c]

University of Heidelberg

Author Note

[a]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: anna-lena.schubert@psychologie.uni-heidelberg.de

[b]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: dirk.hagemann@psychologie.uni-heidelberg.de

[c]University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, E-Mail: gidon.frischkorn@psychologie.uni-heidelberg.de

Correspondence concerning this article should be addressed to Anna-Lena Schubert, University of Heidelberg, Institute of Psychology, Hauptstrasse 47-51, D-69117 Heidelberg, Germany, Phone: +49 (0) 6221-547354, Fax: +49 (0) 6221-547325, E-mail: anna-lena.schubert@psychologie.uni-heidelberg.de

Summary

Individual differences in the speed of information processing give rise to individual differences in general intelligence. Reaction times and latencies of event-related potential have been shown to be moderately associated with intelligence. These associations have been explained either in terms of individual differences in some brain-wide property such as myelination, the speed of neural oscillations, or white-matter tract integrity, or in terms of individual differences in specific processes such as the signal-to-noise ratio in evidence accumulation, executive control, or the cholinergic system. Here we show that smarter individuals have a higher speed of higher-order information processing that explains about 80 percent of the variance in general intelligence. Our results do not support the notion that smarter individuals show advantages in some brain-wide property. Instead, they suggest that smarter individuals benefit from a more efficient transmission of information from frontal attention and working memory processes to temporal-parietal processes of memory storage.

General intelligence is little more than the speed of higher-order processing

General intelligence (*g*) is the common variance shared by different measures of cognitive ability. It is a powerful predictor for success in a variety of life outcomes, such as educational attainment, job performance[1], development of expertise[2], general health[3], longevity[4], and well-being[5]. General intelligence typically accounts for 40-50% of the variance shared by different measures of cognitive ability, and *g* factors from different cognitive test batteries are substantially related[6]. This functional invariance suggests that there may be a single common process underlying individual differences in general intelligence that affects all kinds of cognitive ability tests[7].

One likely candidate for a single neuro-cognitive property affecting a variety of cognitive abilities is the speed of information processing[8]. On a behavioral level, information-processing speed can be measured as reaction times (RTs), which show moderate, but consistent negative associations with intelligence[8,9]. Moreover, reaction times have been shown to mediate the relationship between brain-wide white matter tract integrity and general intelligence, suggesting a functional anatomical basis for fast and efficient information processing[10].

On a neurophysiological level, information-processing speed can be measured as the latency of event-related potentials (ERPs). ERPs allow decomposing the electrophysiological activity between stimulus onset and response into functionally distinct components. These ERP components are correlates of functionally distinct cognitive processes defined by their polarity, their latency, and their topography. A higher speed of information processing should be reflected in shorter ERP latencies (i.e., a shorter time interval between the onset of a stimulus and the maximum peak of the component). ERP components occurring early in the stream of information-processing reflect early stages of information processing, whereas later components reflect higher-order processing. ERPs thus provide a means to identify which cognitive processes are faster in more intelligent individuals.

Previous research has shown only weak and often inconsistent associations between ERP latencies and general intelligence[11], but several studies found a moderate negative association between ERP latencies and intelligence[12,13,14,15]. Moreover, reaction times have been shown to mediate the association between ERP latencies and intelligence, suggesting a functional neuro-cognitive basis for faster information processing that may give rise to individual differences in intelligence[16].

ERP latencies may tend to show smaller and more inconsistent associations with general intelligence than reaction times, because they may be more strongly influenced by situational factors unrelated to mental abilities. Previous research supported this view, suggesting a moderate temporal stability of reaction times[17,18] and substantial variance in the stability of ERP latencies ranging from $r = .19$ to $r = .89$[19]. Hence, the variances of ERP latencies may reflect differences in the speed of information processing both as a brain property and as brain states, resulting in an underestimation of the association between information-processing speed and general intelligence on a neurophysiological level.

The association between information-processing speed and general intelligence has been explained either in terms of individual differences in some brain-wide property such as myelination[20], the speed of neural oscillations[8], or white matter tract integrity[12], or in terms of individual differences in specific processes such as the signal-to-noise ratio in evidence accumulation[21,22], executive or attentional control[23], or the cholinergic system[24].

Instead of asking *which* neuro-cognitive processes underlie the association between general intelligence and information-processing speed, we could also ask *at which point in time* during information processing more intelligent individuals deviate from less intelligent individuals. If individual differences in some brain-wide property explained the relationship between information-processing speed and intelligence (general speed hypotheses), more intelligent individuals should show faster processing at all stages of information processing. If, however, specific cognitive or psychopharmacological processes explained this relationship (specific speed hypotheses), more intelligent individuals should show faster processing only at specific stages of information processing associated with these specific processes. We analyzed inter-individual differences in the latencies of ERP components to determine whether intelligence is associated with faster information processing at all or only at very specific stages of information processing.

Factor structure and temporal stability of RS

Because measures of general intelligence are not affected by situational (occasion-specific) factors and thus reflect a property of the person (plus measurement errors)[25], it may be presumed that occasion-specific effects in time-domain measures act as nuisance variables or "noise" when analyzing the association between these measures and general intelligence. Thus, only the temporally stable (trait-like) portion of variance in reaction times can be considered as a property of the person that may explain individual differences in general intelligence. Not separating the temporally stable portion of variance from the occasion-specific portion of variance may have led to an underestimation of the relationship between chronometric variables and general intelligence in previous research.

We used a hierarchical extension of latent state-trait (LST) theory[26] to identify the common and temporally stable trait variance of three reaction time tasks at two measurement occasions approximately eight months apart. Mean reaction speeds (RSs) for all conditions of the three experimental tasks are shown in Supplementary Table 1 separately for the two measurement occasions. 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 hierarchical method factor M_j for each of the three experimental tasks *j* as shown in the left part of Figure 1. The model provided an acceptable fit to the data, $\chi^2(133) = 265.81$, p <.001, CFI = .95, RMSEA = .09. Because the variances of the first state residual and of several method residuals were non-significant or negative, they were fixed to zero. These modifications did not impair model fit, $\chi^2(137) = 283.26$, p <.001, CFI = .94, RMSEA = .09. See Supplementary Table 3 for details of model specifications and a complete overview of estimated model parameters.

Based on this model, we calculated several LST parameters, namely the coefficient of reliability, consistency, occasion-specifity, and method-specify of the reaction speeds, which are shown in Table 1. The coefficient of reliability reflects the amount of variance accounted for by the hierarchical model, residual variances excluded, in each manifest variable. Reliabilities were high for all tasks and conditions, suggesting a great portion of systematic variance in each RS measurement and strong structural relations between variables. The coefficient of consistency reflects the amount of variance explained in each manifest variable by the shared variance of ERPs across tasks and laboratory sessions. Consistencies were greatest for the Posner letter matching task, but substantial for all tasks ranging from .53 to .76. The coefficient of occasion-specifity reflects the amount of variance explained in each manifest variable by situational influences and influences of person-situation interactions in each manifest variable. Occasion specifities were negligible, which is consistent with previous work on the stability of reaction times[17,18]. The coefficient of method-specifity reflects the amount of variance explained in each manifest variable by task- and condition-specific factors. Method specifities were moderate, ranging from .08 to .32, and highest for the SRT/CRT tasks and the Sternberg memory scanning task.

The relationship between trait RS and *g*

Previous research has shown that correlations between composite measures of mental speed and mental abilities tend to be higher than the correlations of single reaction time measures[27,28]. Therefore, we assessed the correlation between the common reaction speed trait and general intelligence. For this purpose, we added a hierarchical model of general intelligence to the LST model and allowed the reaction speed trait to correlate with general intelligence (see Figure 1, Supplementary Table 3). This model provided a good fit to the data, $\chi^2(253) = 480.32$, $p < .001$, CFI = .92, RMSEA = .09. The latent correlation between general intelligence and general behavioral information processing speed was moderate, $r = .43$, $p < .001$.

This correlation is consistent with previous studies reporting correlations ranging from $r = -.22$ to $-.45$ between reaction times in these tasks and mental abilities[9]. Given the high temporal stability and great reliability of reaction times, it is not surprising that the latent correlation did not exceed the size of these correlations notably.

Factor structure and temporal stability of ERP latencies

To analyze whether more intelligent individuals show advantages in the speed of information processing at all stages of information processing, or specifically only at earlier or later stages, we compared two structural equation models of neurophysiological processing speed. Similar to the model for reaction speed, these models were LST models[26] that allowed identifying the common and temporally stable trait variance of five ERP components across three experimental tasks and two measurement occasions. Grand-average waveforms of event-related potentials averages across measurement occasions are presented in Figure 2 separately for more and less intelligent individuals and for each of the three tasks.

The general processing speed model consisted of a hierarchical 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 specific processing speed model assumed two separate hierarchical common traits for earlier and later ERP latencies. ERP latencies, averaged across conditions, are shown in Supplementary Table 2 for each of the three experimental tasks and each of the two measurement occasions.

The specific processing speed model provided a notably better account for the data, $\chi^2(469) = 689.08$, $p < .001$, CFI = .85, RMSEA = .06, AIC = 749.08, than the general processing speed model, $\chi^2(472) = 880.95$, $p < .001$, CFI = .73, RMSEA = .09, AIC = 926.95. Therefore, we used the specific processing speed model for all further analyses and fixed all non-significant variances in this model to zero as shown in the left part of Figure 2, which did not impair model fit significantly, $\chi^2(481) = 719.22$, $p < .001$, CFI = .84, RMSEA = .06. See Supplementary Table 4 for details of model specifications and a complete overview of estimated model parameters for the specific processing speed model.

As shown in Table 2, the reliability of ERP latencies was as low as expected with very low reliabilities for the earlier latencies and somewhat higher reliabilities for the later latencies. Consistencies (.11 to .14) and method specifities (.11 to .17) contributed about equally to the variance of the earlier ERP latencies, whereas consistencies were notably greater (.42 to .63) than method specificities (.08 to .09) for the later latencies. We observed relevant occasion specifity for earlier latencies at the second measurement occasion, whereas the influence of occasion specifity on later latencies was negligible.

The relationship between trait ERP latencies and *g*

In order to compare the plausibility of general speed hypotheses with the plausibility of specific speed hypotheses, we determined whether intelligence is associated with faster information processing at all or only at very specific stages of information processing. For this purpose, we analyzed how the common traits for earlier and later latencies correlated with general intelligence. Our results concerning the psychometric properties of ERP latencies indicate that the low reliabilities and consistencies of ERP measurements may have led to an underestimation of the relationship between ERP latencies and general intelligence in previous studies[11]. Hence, the latent correlations between ERP latency traits and general intelligence should be greater than the typically observed moderate manifest correlations.

For this purpose, we again added the hierarchical model of general intelligence to the structural model of ERP latencies and allowed the general ERP latency traits to correlate with general intelligence This model (see Figure 2, Supplementary Table 4) provided a good fit to the data, $\chi^2(679) = 1048.84$, $p < .001$, CFI = .83, RMSEA = .07. The common trait for earlier latencies was positively correlated with general intelligence, $r = .33$, $p < .001$, indicating that more intelligent individual tended to show later P100 and N100 peak amplitudes. Moreover, the common trait for later latencies was negatively correlated with general intelligence, $r = -.89$, $p < .001$, indicating that more intelligent individual showed earlier P200, N200, and P300 peak amplitudes. In particular, P300 latencies showed the highest standardized loadings on the common trait for later latencies. This suggests that P300 latencies showed the greatest association of all ERP latencies with general intelligence. Taken together, ERP latency traits explained 90.10 % of the variance in general intelligence.

Discussion

The current study investigated whether more intelligent individuals have advantages in the speed of information processing at all stages of information processing, or only at specific earlier or later stages as reflected in the latency of specific ERP components. General intelligence was weakly associated with longer latencies of earlier ERP components (i.e., P100 and N100), and strongly associated with shorter latencies of later ERP components (i.e., P200, N200, and P300). This result suggests that smarter individuals do not have a general, but a very specific advantage in the speed of higher-order information processing.

Our results contradict popular theories proposing that individual differences in some brain-wide property explain the relationship between processing speed and general intelligence[8,10,20]. Instead, they suggest that more intelligent individuals process information faster specifically because of faster higher-order processing. The greatest association between ERP latencies and general intelligence was found for the P300, which is consistent with previous studies reporting an association between mental abilities and the visual or auditory P300[12-14]. According to the context-updating interpretation of the P300[29], this association may reflect a faster inhibition of extraneous processes that facilitates the transmission of information from frontal attention and working memory processes to temporal-parietal processes of memory storage[30]. This interpretation is consistent with previous research showing that individual differences in inhibition and updating are related to general intelligence[31].

The latent composite measures of neurophysiological information-processing speed explained 90 percent of the variance in general intelligence in the present study. This association exceeds the weak negative correlations between ERP latencies and mental abilities reported in previous studies notably[11] and demonstrates the benefits of latent variable modeling. Contrary to our expectations, neurophysiological processing speed was not more strongly influenced by situational factors than behavioral processing speed. Instead, ERP latencies had lower consistencies than reaction speeds, indicating that the covariance between different ERP components and tasks is relatively low and that measuring neurophysiological processing speed reliably requires multiple measurements. Hence, it is not surprising that the P300 is the only one of the measured ERP components for which associations with general intelligence have been repeatedly reported[12-14], as P300 latencies had the highest reliabilities and showed the greatest association with general intelligence of all ERP latencies in the present study.

It is intriguing that latent measures of neurophysiological information-processing speed showed greater associations with general intelligence than latent measures of behavioral processing speed. These findings contrast with reports in the related literature in which behavioral processing speed was more strongly and more consistently associated with mental abilities than neurophysiological processing speed. Our results make it clear that measures of neurophysiological processing speed contain a great amount of task- and component-specific variance, and that once this unique variance has been accounted for, neurophysiological processing speed explains a great amount of variance in general intelligence. In fact, our results suggest that ERP latencies may provide a more precise measurement of information-processing speed, whereas reaction speeds may be contaminated by other response-related processes such as motor preparation and execution.

One feature of the present study that limits the conclusions that can be drawn about the association between the speed of information processing and general intelligence is that we used only very simple reaction time tasks. These so-called elementary cognitive tasks are cognitively undemanding to minimize the unwanted influence of individual differences in strategy use and previous experience with specific elements of these tasks on reaction times[32]. Whether cognitively more demanding reaction time tasks such as working memory tasks or information-processing speed paradigms not requiring a motor response such as the inspection time task would yield comparable results is an open question.

Taken together, our results illustrate that the speed of information processing is a crucial component of general intelligence. Given that general intelligence was only associated with a higher speed in the peak latency of ERP components occurring later in the stream of information processing, and given that the latencies of these later components explained 80% of the variance in general intelligence, we conclude that general intelligence is little more than the speed of higher-order processing.

Methods

Participants

Sample size was determined based on the hypothesis of close fit (H0: $\varepsilon \leq 0.05$, H1: $\varepsilon \geq 0.08$) for the structural equation model with the fewest degrees of freedom ($df = 133$), an alpha error of $\alpha = .05$, and a power of $1 - \beta = .80$[33]. The resulting minimum sample size was $N = 109$. More participants were recruited to increase power and the stability of model estimates.

We recruited a sample of $N = 134$ participants (81 females, 53 males) between 18 and 60 years old ($M = 37.1$, $SD = 13.8$) from different educational and occupational backgrounds via local newspaper advertisement, announcements on social media platforms, and distribution of flyers in Heidelberg. Of these, $N = 122$ participants completed the second measurement occasion and $N = 114$ participants completed the third measurement occasion. We only included the $N = 122$ participants that showed up for at least the first two measurement occasions in the following analyses. This sample consisted of 72 women and 50 men with a mean age of $M = 36.7$ ($SD = 13.6$). A sample size of $N = 122$ participants corresponded to a power of $1 - \beta = .86$ for a structural equation model with 133 degrees of freedom.

All participants had normal or corrected to normal vision and no history of mental illness. At the first laboratory session, participants signed an informed consent. They received 100€ and feedback about their personal results as a reward for their participation. The study was approved by the ethics committee of the faculty of behavioral and cultural studies, Heidelberg.

Measures

Reaction time tasks

Single and choice reaction time task. We used a single and choice reaction time task with three conditions (one, two, and four alternatives) based on the computer-adapted Hick task[16]. Each trial began with the presentation of four white squares in a row on a black screen and a white fixation cross in the middle of the squares that was shown for 1000-1500 ms. Next, the fixation cross disappeared and a larger cross appeared in one of the four squares. Participants had to press the corresponding response key as fast as possible. After their response, the screen remained unchanged for 1000 ms to allow the recording of post-decisional neuronal processes. The intertrial interval (ITI) consisted of a black screen and lasted between 1000-1500 ms. During the whole task, participants' middle and index fingers rested on four keys directly underneath the squares to increase stimulus-response compatibility. All keys irrelevant to the tasks had been removed from the modified keyboard.

Each of the three conditions consisted of ten practice trials with immediate feedback followed by 200 test trials without feedback. The order of conditions was counterbalanced across participants. In the single reaction time (SRT) task, participants always knew exactly where the cross would appear. There were four blocks of 50 trials each with a counterbalanced order across participants, in which participants had to pay attention to only one of the four squares. In the two-choice reaction time (2CRT) task, participants knew in which two squares the cross could appear. There were four blocks of 50 trials each with a counterbalanced order across participants, in which participants had to pay attention only to the left/right/middle/outer two squares. In the four-choice reaction time (4CRT) task, participants were given no indication where the cross will appear.

Sternberg memory scanning task. Participants were shown memory sets consisting of digits between 0 and 9, and they had to indicate whether an immediately afterwards presented probe stimulus was part of the previously presented memory set. We administered three different experimental conditions (set size one, three, and five) in an order counterbalanced across participants. Each of the three conditions consisted of ten practice trials with immediate feedback followed by 100 test trials without feedback.

At the beginning of each trial, a white fixation cross was shown in the middle of a black screen for 1000-1500 ms. Digits were presented sequentially for 1000 ms with a blank screen shown for 400-600 ms between each digit. A black screen with a white question mark was shown for 1800-2200 ms after the last digit was presented, followed by a black screen showing the probe stimulus. Participants had to press one of two keys with their index fingers to indicate whether the digit was part of the memory set, which was the case in 50% of the trials. The position of keys was counterbalanced across participants. After their response, the screen remained unchanged for 1000 ms, followed by an ITI of 1000-1500 ms.

Posner letter matching task. Participants were shown two letters and had to decide whether they were identical. In the physical identity (PI) condition, participants were instructed to identify letters as identical only if their physical characteristics were identical (i.e., "QQ" would be identical, whereas "Qq" or "QA" would be different). In the name identity (NI) condition, they were instructed to identify letters as identical if their names were identical (i.e., both "QQ" and "Qq" would be identical, whereas "QA" would still be different). Both conditions consisted of 10 practice trials with immediate feedback and 300 test trials without feedback. At the first measurement occasion, the PI condition was administered first to all participants, whereas at the second measurement occasion the NI condition was administered first to all participants.

Each trial began with a white fixation cross shown on a black screen for 1000-1500 ms, followed by a pair of white letters presented in the middle of the screen. Participants had to press one of two keys with their index fingers to indicate whether the letters were identical, which was the case in 50% of the trials. Again, the position of keys was counterbalanced across participants. After the response, the screen remained unchanged for 1000 ms, followed by an ITI that consisted of a black screen and lasted 1000-1500 ms.

Intelligence tests

Advances Progressive Matrices (APM). We used a computer adapted version of Raven's Advanced Progressive Matrices[34] to measure participants' general intelligence with a power test. The APM has been previously shown to be the best single indicator of *g*[35]. Participants' performance was determined as the number of correctly solved items of the second set, as suggested by the test manual. Moreover, we performed an odd-even split of the test items in the second set and used the number of correctly solved items in the odd and even trials as two indicators of latent APM performance. We then transformed these raw test sores to *z*-scores for further analyses.

The number of correctly solved items in the APM was $M = 23.43$ ($SD = 6.71$), which corresponds to an IQ of $M = 98.80$ ($SD = 15.68$). The number of correctly solved items in the even trials was $M_{even} = 12.23$ ($SD = 3.51$), and in the odd trials $M_{odd} = 11.20$ ($SD = 3.52$). Data from two participants was lost due to technical reasons.

Berlin intelligence structure test (BIS). We administered the complete Berlin intelligence structure test (BIS)[36] in groups of up to four participants. The BIS is based on the bimodal Berlin intelligence structure model[37], which distinguishes between four operation-related (processing speed, memory, creativity, processing capacity) and three content-related (verbal, numerical, figural) components of general intelligence. The test consists of a total of 45 tasks with each task being a combination of one operation-related component with one content-related component of intelligence. According to the manual, participants' scores for all seven components were computed by aggregating the normalized *z*-scores of all tasks related to the respective operation and content components. We then transformed these scores to *z*-scores for further analyses. We did not compute IQ scores based on BIS results, because there is no adult normative sample with an appropriate age range available.

The mean score of the processing speed component was $M = 98.00$ ($SD = 7.10$), the mean score of the memory component was $M = 99.40$ ($SD = 6.51$), the mean score of the creativity component was $M = 98.02$ ($SD = 6.14$), the mean score of the processing capacity component was $M = 101.7$ ($SD = 7.99$), the mean score of the verbal component was $M = 102.40$ ($SD = 6.93$), the mean score of the numerical component was $M = 98.27$ ($SD = 6.79$), and the mean score of the figural component was $M = 97.69$ ($SD = 6.52$). Note that these are not IQ scores, but mean scores.

Procedure

The three measurement occasions were approximately four months apart. At the first and third measurement occasion, we administered the SRT and CRT task, the Sternberg memory scanning task, and the Posner letter matching task in the same order for all participants while an EEG was recorded. Participants were seated in a sound-attenuated, dimly lit EEG cabin. At the beginning of the third measurement occasion, we additionally recorded 12 minutes of resting state EEG, which will not be reported here. Each session took approximately 3 hours. At the second measurement occasion participants completed the BIS, a personality questionnaire not reported here, the APM, and a questionnaire about demographic data. This session took approximately 3.5 hours.

EEG recording

The EEG was recorded with 32 equidistant Ag–AgCl electrodes. We used the aFz electrode as the ground electrode. Electrodes were initially referenced to Cz and offline re-referenced to an average reference. To correct for ocular artifacts, we recorded the electrooculogram (EOG) bipolarly with two electrodes positioned above and below the left eye and two electrodes positioned at the outer canthi of the eyes. All electrode impedances were kept below 5 kΩ. The EEG signal was recorded continuously with a sampling rate of 1000 Hz (band-pass 0.1–100 Hz), and filtered offline with a low-pass filter of 16 Hz.

Data analysis

Data preprocessing

Reaction time data. For intraindividual outlier detection in RTs, we discarded any RTs faster than 100 ms or slower than 3000 ms. In a second step, we discarded any trials with incorrect responses or with logarithmized RTs exceeding ± 3 SDs of the mean of each condition. Subsequently, we calculated the mean RT for each condition and calculated inverted RTs as reaction speeds (RS). We then transformed these reaction speeds to z-scores for further analyses.

Electrophysiological data. We calculated ERPs separately for each ECT and each condition. ERPs were time-locked to the stimulus onset in the SRT, CRT and letter matching task, whereas they were time-locked to the onset of the probe in the memory scanning task. Epochs were 1200 ms long including a baseline of 200 ms before stimulus onset. We corrected ocular artifacts using a regression procedure[38]. Epochs with amplitudes exceeding ± 70 µV, with amplitude changes exceeding 100 µV within 100 ms, or with activity lower than 0.5 µV were discarded as artifacts.

We determined the P100 peak latency at occipital electrodes contralateral to the position of the cross in the SRT and CRT task, and at the occipital electrode over midline for the Sternberg memory scanning task and the Posner letter matching task because stimuli in these tasks were presented centered. We determined the N100 peak latency at the frontal electrode over midline, the P200 peak latency at fronto-central electrode over midline, and the N200 and P300 latency at the parietal electrode over midline. Peak latencies were determined separately for each condition of the three experimental tasks. Subsequently, we discarded any peak latencies exceeding ± 3 SDs of the mean peak latency of each condition. Finally, peak latencies were averaged across conditions of each experimental task and *z*-standardized for further analyses.

Statistical analysis

Prior to multivariate analyses, we calculated the Malahanobis distance to identify and subsequently exclude multivariate outliers. In the multivariate data space of reaction speeds and intelligence test scores, one participant was identified as a multivariate outlier ($D_M = 57.78$, $p < .001$) and excluded from further analyses, whereas no participant was identified as an outlier in the multivariate data space of ERP latencies and intelligence test scores. Data files are provided in the Supplementary Information file *datafiles.rar*.

Moreover, all manifest variables were inspected for univariate normal distribution, which is a prerequisite of multivariate normal distribution. Statistical tests of skewness and kurtosis indicated that the distribution of APM variables and of a few ERP latencies deviated from normal distribution ($p < .001$). Subsequent visual inspections revealed that these deviations were rather small and below threshold values of skewness = 2 and of kurtosis = 7. Because neither skewness nor kurtosis exceeded these threshold values, we followed recommendations to not use assumption-free estimates in structural equation models[39].

We used structural equation modeling to assess the associations between reaction speeds, ERP latencies, and general intelligence. All models were fitted with the full information maximum likelihood algorithm implemented in AMOS[40]. Because some variables deviated slightly from normal distribution, we repeated all analysis with the bootstrap procedure implemented in AMOS[40]. As bootstrapped results did not deviate notably from non-bootstrapped results except for small deviations in the size of standard errors, we report only the non-bootstrapped results.

Within the framework of structural equation modeling, we built latent state-trait (LST) models with hierarchical traits and hierarchical method factors to achieve a virtually error-free measurement of trait reaction speeds and ERP latencies. LST theory is an expansion of classical test theory that takes into account that any measurement is always affected by situational factors[26]. In short, LST theory proposes that the variance of an observed variable Y_{ij} can be decomposed into the variance of the latent trait T, the variance of a latent state residual SR_i, the variance of a latent method residual M_j, and the variance of a latent unsystematic error residual ε_{ij}.

First, we fitted a structural equation model with a common trait T, a state residual SR_i for each of the two measurement occasions, and a hierarchical method factor M_j for each of the three experimental tasks to the reaction speed data (for details of the model specifications see Supplementary Table 3). Subsequently, we fitted two different structural equation models to the electrophysiological data to compare the model fit of a general processing speed-model to the model fit of a specific processing speed-model. In both models, each ERP latency (P100, N100, P200, N200, P300) was modeled hierarchically as the covariance of latencies of one ERP component across the three experimental tasks at one measurement occasion (e.g., the latent P100 variable at measurement occasion *i* was defined by the covariances between the P100 latencies in the Hick, Sternberg, and Posner task at this measurement occasion). In addition, we included specific traits for each of the five ERP component latencies. Intercepts in all models were fixed to zero (for details of the model specifications see Supplementary Table 4).

The general processing speed-model consisted of a common trait *T*, a state residual SR_i for each of the two measurement occasions, a method factor M_j for each of the three experimental tasks, and five specific traits for the five ERP components. The specific processing speed-model consisted of two separate common traits, $T_{ealier\ latencies}$ and $T_{later\ latencies}$, a state residual SR_i for each of the two measurement occasions and each of the two traits, a method factor M_j for earlier and later ERP latencies of each of the three experimental tasks, and five specific traits for the five ERP components. We evaluated goodness-of-fit based on the comparative fit index (CFI)[41] and the root mean square error of approximation (RMSEA)[42] and compared model fit of the two models with the Akaike Information Criterion (AIC)[43]. The statistical significance of model parameters was assessed with the two-sided critical ratio test.

Moreover, we fitted a hierarchical measurement model of general intelligence with a common *g*-factor and two lower-order factors defined by the covariances in BIS and APM scores, respectively. We did not model general intelligence with LST theory, because previous applications of LST theory to multiple measurements of intelligence have shown that the state residuals were zero[25]. Again, we evaluated goodness-of-fit with the CFI and the RMSEA.

In a second step, we computed latent-state parameters of the behavioral and electrophysiological data based on the best-fitting LST model. For each manifest variable Y_{ij} measured with method *j* and at measurement occasion *i*, we computed coefficients of consistency, occasion-specifity, measurement-specifity, and reliability[44]. For the LST model of reaction speeds, the coefficient of trait-specifity was 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-specifity was 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-specifity was computed as $\sigma^2(M_j + C_k)/\sigma^2(Y_{ij})$ and reflects the proportion of variance that can be accounted for by a specific experimental task M_j and its conditions C_k. Taken together, these different sources of systematic variation contribute to the reliability of a manifest variable Y_{ij}, i.e. to the proportion of variance explained by the specified model. Thus, the reliability coefficient can be computed as $[\sigma^2(T) + \sigma^2(SR_i) + \sigma^2(M_j+ C_k)]/ \sigma^2(Y_{ij})$.

For the LST model of ERP latencies, the coefficient of trait-specifity was computed as $\sigma^2(T_k + T_l)/\sigma^2(Y_{ij})$ and reflects the proportion of variance of the manifest variable Y_{ij} hat can be accounted for by individual differences in the general trait T_k and by the specific trait T_l. The coefficient of occasion-specifity was computed as described above. The coefficient of method-specifity was computed as $\sigma^2(M_j)/\sigma^2(Y_{ij})$ and reflects the proportion of variance that can be accounted for by the shared variance of earlier and later latencies in a specific experimental task M_j. Finally, the coefficient of reliability was computed as $[\sigma^2(T_k + T_l) + \sigma^2(SR_i) + \sigma^2(M_j)]/ \sigma^2(Y_{ij})$.

In a third step, we combined the behavioral and the best-fitting neurophysiological structural equation model with the measurement model of general intelligence to study the structural relations between a) behavioral processing speed and general intelligence, and b) neurophysiological processing speed and general intelligence.

References

1. Schmidt, F. L. & Hunter, J. General mental ability in the world of work: occupational attainment and job performance. *Journal of personality and social psychology*, **86**, 162–173 (2004).

2. Wai, J. Experts are born, then made: Combining prospective and retrospective longitudinal data shows that cognitive ability matters. *Intelligence*, **45**, 74–80 (2004).

3. Der, G., Batty, G. D. & Deary, I. J. The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, **37**, 573–580 (2009).

4. Deary, I. Why do intelligent people live longer? *Nature*, **456**, 175–176 (2008).

5. Pesta, B. J., McDaniel, M. A. & Bertsch, S. Toward an index of well-being for the fifty U.S. states. *Intelligence*, **38**, 160–168 (2010).

6. Johnson, W., Nijenhuis, J. T. & Bouchard, T. J. Still just 1 g: Consistent results from five test batteries. *Intelligence*, **36**, 81–95 (2008).

7. Spearman, C. *The nature of 'intelligence' and the principles of cognition*. Oxford, England: Macmillan (1923).

8. Jensen, A. R. *Clocking the mind: Mental chronometry and individual differences*. Amsterdam: Elsevier (2006).

9. Sheppard, L. D. & Vernon, P. A. Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, **44**, 535–551 (2008).

10. Penke, L. *et al.* Brain white matter tract integrity as a neural foundation for general intelligence. *Mol. Psychiatry*, **17**, 1026–1030 (2012)

11. Schulter, G. & Neubauer, A. Zentralnervensystem und Persönlichkeit. In J. Henning & P. Netter (Eds.), *Biopsychologische Grundlagen der Persönlichkeit* (pp. 35-190). München: Elsevier (2005).

12. Bazana, P. G. & Stelmack, R. M. 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 (2002).

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

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

15. Troche, S. J., Indermühle, R., Leuthold, H. & Rammsayer, T. H. Intelligence and the psychological refractory period: A lateralized readiness potential study. *Intelligence*, **53**, 138–144 (2015).

16. Schubert, A.-L., Hagemann, D., Voss, A., Schankin, A. & Bergmann, K. Decomposing the Relationship between Mental Speed and Mental Abilities. *Intelligence,* **51**, 28-46 (2015).

17. Roznowski, M. & Smith, M. L. A note on some psychometric properties of Sternberg task performance: Modifications to content. *Intelligence*, **17**, 389–398 (1993).

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

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

20. Miller, E. M. Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, **17**, 803–832 (1994).

21. van Ravenzwaaij, D., Brown, S. & Wagenmakers, E. An integrated perspective on the relation between response speed and intelligence. *Cognition,* **119**, 381-393 (2011).

22. Vickers, D. & Smith, P. L. The rationale for the inspection time index. *Personality and Individual Differences,* **7**, 609-623 (1986)

23. 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38**, 525–549 (2012).

24. Stough, C., Thompson, J., Bates, T. & Nathan, P. Examining neurochemical determinants of inspection time: Development of a biological model. *Intelligence,* **29**, 511-522 (2001).

25. 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*, **39**, 323-334 (2011).

26. Steyer, R., Ferring, D. & Schmitt, M. J. States and traits in psychological assessment. *European Journal of Psychological Assessment*, **8**, 79–98 (1992).

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

28. Miller, L. T., & Vernon, P. A. Intelligence, reaction time, and working memory in 4- to 6-year-old children. *Intelligence*, **22**, 155–190 (1996).

29. Donchin, E. Surprise! … Surprise? *Psychophysiology*, **18**, 493–513 (1981).

30. Polich, J. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, **118**, 2128–2148 (2007).

31. 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*, **53**, 166–180 (2015).

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

33. MacCallum, R. C., Browne, M. W., & Sugawara, H. M. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, **1**, 130-149 (1996).

34. Raven, J. C., Court, J. H. & Raven, J. *Manual for Raven's progressive matrices and mill hill vocabulary scales. Advanced progressive matrices*. Oxford: Oxford University Press (1994).

35. Marshalek, B., Lohman, D. F. & Snow, R. E. The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, **7**, 107-127 (1983).

36. Jäger, A. O., Süß, H.-M. & Beauducel, A. *Berliner Intelligenzstruktur-Test. Form 4*. Göttingen: Hogrefe (1997).

37. Jäger, A.O. Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, **28**, 195–226 (1982).

38. Gratton, G., Coles, M. G., & Donchin, E. A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology*, **55**, 468-484 (1983).

39. West, S. G., Finch, J. F., Curran, P. J. Structural equation models with non-normal variables: Problems and remedies. *Structural equation modeling: Concepts, issues, and applications* (ed. Hoyle, R.), 56–75, Thousand Oaks, CA: Sage (1995).

40. Arbuckle, J. L. *Amos 7.0 user's guide*. Chicago: SPSS (2006).

41. Bentler, P. M. Comparative fit indexes in structural models. *Psychological Bulletin*, **107**, 238-246 (1990).

42. Browne, M. & Cudeck, R. Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Sage focus editions: Vol. 154. Testing structural equation models* (pp. 136–162). Newbury Park: Sage Publications (1993).

43. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado (1973).

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

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions D.H. and A.-L.S. designed the experiment; A.-L.S. collected the data; G.T.F. pre-processed the behavioral data; A.-L. S. pre-processed the EEG data; A.-L.S. and G.T.F. analyzed the data; A.-L.S. wrote the paper; D.H. and G.T.F. gave conceptual and technical advice. All authors discussed the results and commented on the paper.

Author Information Data used for structural equation modeling has been uploaded as Supplementary Information. Upon publication, data will be deposited with the Open Science Framework. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to anna-lena.schubert@psychologie.uni-heidelberg.de.

Table 1

Latent-state-trait theory parameters of reaction speed variables.

	Consistency		Occasion-Specifity		Method-Specifity		Reliability	
	Session		Session		Session		Session	
	1	2	1	2	1	2	1	2
SRT/CRT tasks								
SRT	.60	.54	0	.10	.30	.27	.90	.91
CRT2	.60	.54	0	.10	.30	.27	.90	.91
CRT4	.65	.59	0	.11	.24	.21	.89	.91
Sternberg task								
Set size 1	.71	.64	0	.11	.16	.15	.88	.89
Set size 3	.68	.61	0	.11	.21	.19	.89	.90
Set size 5	.58	.53	0	.09	.32	.29	.90	.91
Posner task								
Physical Identity	.76	.68	0	.12	.10	.08	.87	.89
Name Identity	.74	.66	0	.12	.13	.12	.88	.89

Note. SRT = single choice reaction time task; CRT2/4 = choice reaction time task with two/four alternatives.

Table 2

Latent-state-trait theory parameters of ERP latencies.

		Consistency		Occasion-Specifity		Method-Specifity		Reliability	
		Session		Session		Session		Session	
		1	2	1	2	1	2	1	2
SRT/CRT tasks									
	P100	.14	.11	0	.17	.14	.12	.28	.40
	N100	.14	.11	0	.17	.14	.12	.28	.40
	P200	.51	.51	0	0	.08	.08	.59	.59
	N200	.42	.48	.12	0	.08	.09	.62	.56
	P300	.57	.57	0	0	.09	.09	.66	.66
Sternberg task									
	P100	.13	.11	0	.16	.17	.14	.30	.42
	N100	.13	.11	0	.16	.17	.14	.30	.42
	P200	.55	.55	0	0	0	0	.55	.55
	N200	.46	.46	.13	0	0	0	.58	.46
	P300	.63	.63	0	0	0	0	.63	.63
Posner task									
	P100	.14	.11	0	.17	.13	.11	.27	.39
	N100	.14	.11	0	.17	.13	.11	.27	.39
	P200	.55	.55	0	0	0	0	.55	.55
	N200	.46	.46	.13	0	0	0	.58	.46
	P300	.63	.63	0	0	0	0	.63	.63

Note. SRT/CRT = single and choice reaction time task.

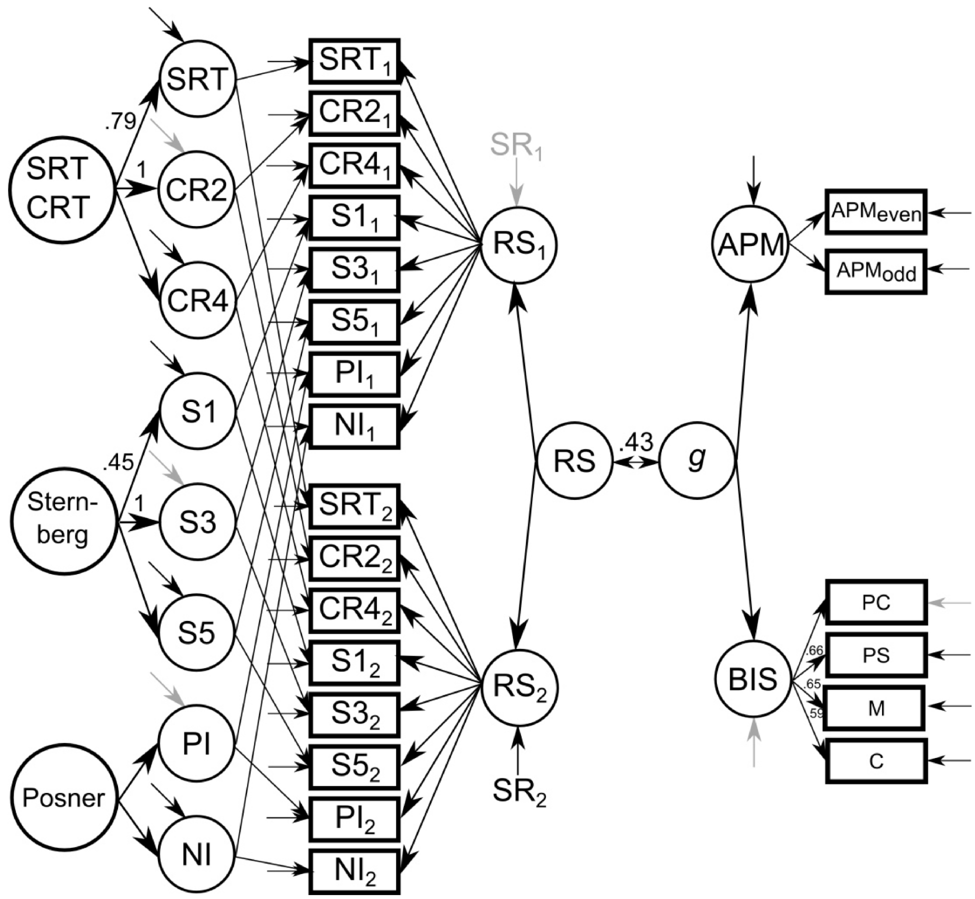

Figure 1. A structural equation model consisting of the latent-state-trait model of reaction speed and the hierarchical model of general intelligence. The LST model of reaction speed consists of a common trait *T*, a state residual SR_i for each of the two measurement occasions *i*, and a hierarchical method factor M_j for each of the three experimental tasks *j* and its conditions *k*. The hierarchical intelligence model consists of a common *g*-factor and two lower-order factors defined by the covariances in BIS and APM scores. The model provided a good fit to the data, $\chi^2(253) = 480.32$, $p < .001$, CFI = .92, RMSEA = .09. All factors loadings are fixed to one; if not, standardized regression weights are shown next to paths. Non-significant residuals ($p \geq .05$) are grayed out. SRT = single reaction time task; CR2/4 = choice reaction time task with two/four alternatives; S1 = set size one; S3 = set size three; S5 = set size give; PI = physical identity; NI = name identity; PC = processing capacity; PS = processing speed; M = memory; C = creativity. Indices *i* at states RS_i and state residuals SR_i indicate the measurement occasion. $N = 121$.

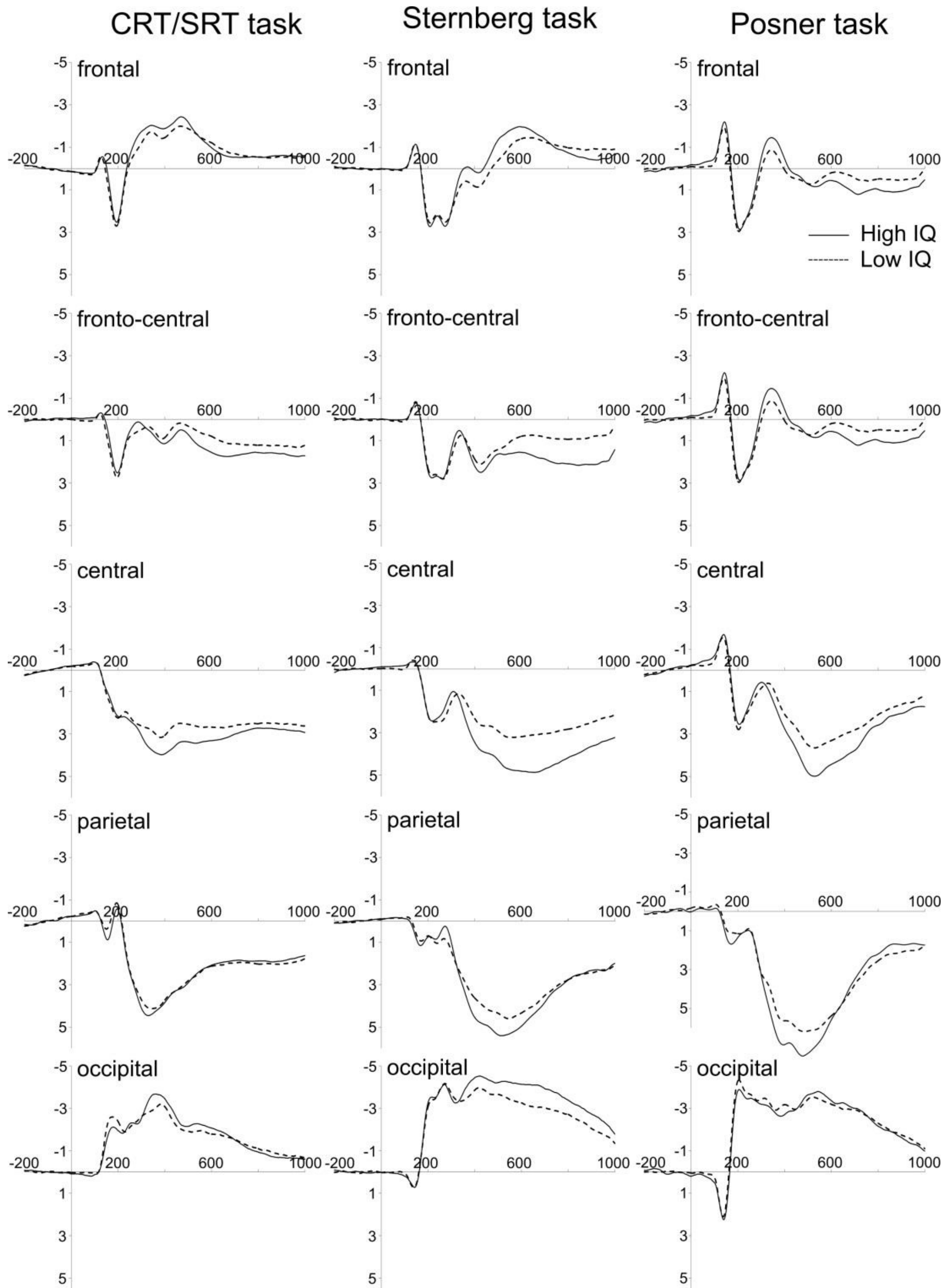

Figure 2. Grand averages of event-related potentials as measured at frontal, fronto-central, central, parietal, and occipital electrodes over midline, separately for more and less intelligent individuals and experimental tasks. ERPs were elicited by the stimulus onset and averaged across measurement occasions and conditions for each experimental task. High and low IQ groups were created based on a median split of the BIS total score. $N = 122$.

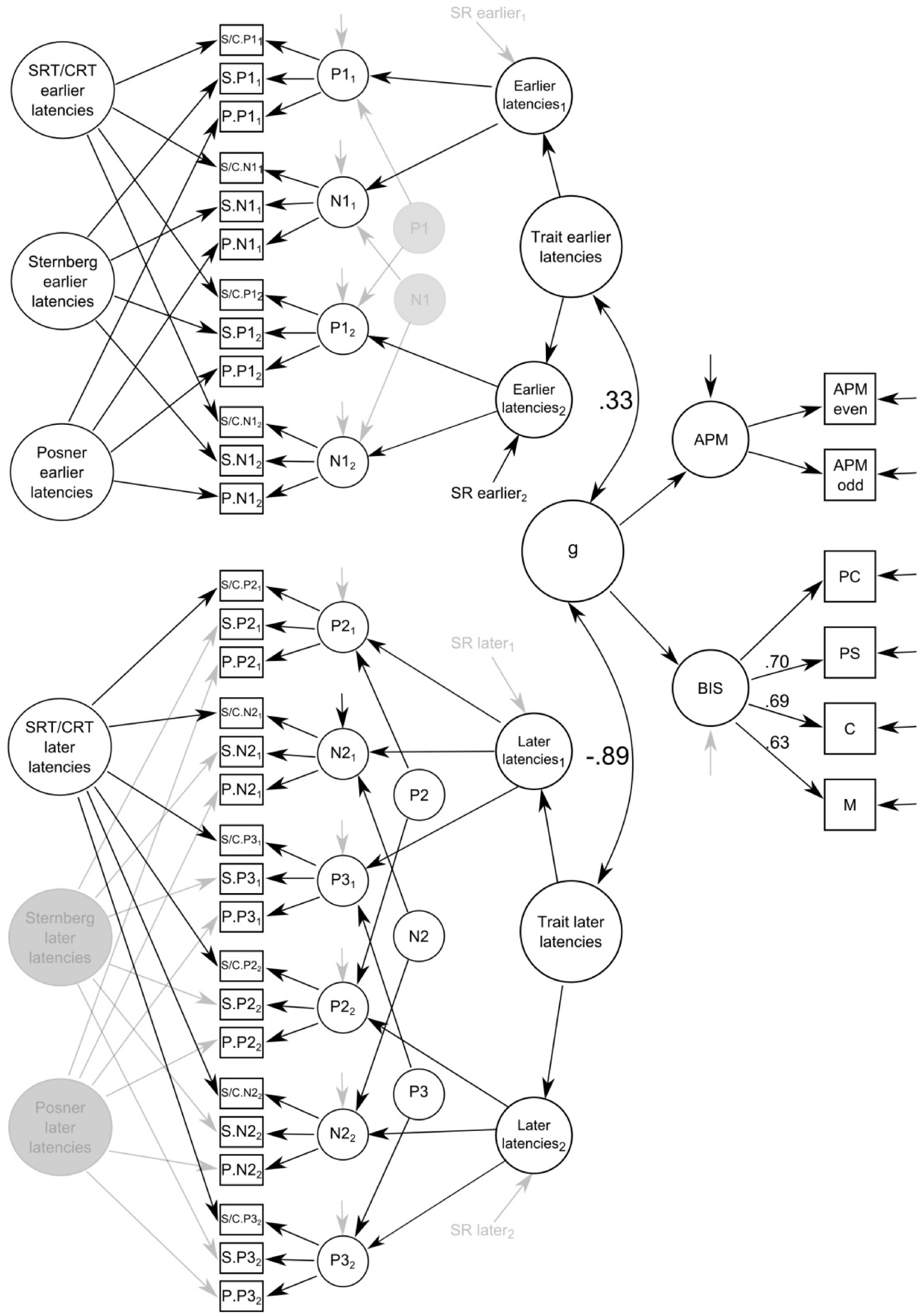

Figure 3. A structural equation model consisting of the specific processing speed-model of ERP latencies and the hierarchical model of general intelligence. The specific processing speed-model consisted of two separate hierarchical common traits, $T_{ealier\ latencies}$ and $T_{later\ latencies}$, a state residual SR_i for each of the two measurement occasions and each of the two traits, specific traits for the five ERP components (only significant specific traits are shown in the figure) and a

method factor M_j for earlier and later ERP latencies of each of the three experimental tasks. The model provided a good fit to the data, $\chi^2(679) = 1048.84$, p <.001, CFI = .83, RMSEA = .07. All factors loadings are fixed to one; if not, standardized regression weights are shown next to paths. Error residuals are not shown. Non-significant residuals ($p \geq .05$) are grayed out. SRT/CRT = single/choice reaction time task; S = Sternberg letter matching task; P = Posner letter matching task; PC = processing capacity; PS = processing speed; M = memory; C = creativity. Indices *i* at states ERP_i and state residuals SR_i indicate the measurement occasion. N = 122.

Supplementary Information

Table 1

Mean RTs (SD in parentheses) for all conditions of the three experimental tasks.

	Session 1		Session 2	
	Accuracies	RTs	Accuracies	RTs
SRT/CRT tasks				
SRT	1.00 (.01)	315.51 (53.01)	1.00 (.00)	317.20 (80.45)
CRT2	.99 (.01)	382.79 (58.02)	1.00 (.01)	381.27 (61.01)
CRT4	.99 (.01)	477.22 (82.64)	.98 (.02)	467.31 (85.7)
Sternberg task				
Set size 1	.97 (.02)	590.96 (115.67)	.98 (.02)	584.02 (135.64)
Set size 3	.97 (.02)	728.46 (167.21)	.98 (.03)	706.61 (176.81)
Set size 5	.97 (.03)	890.03 (240.74)	.95 (.09)	850.98 (223.18)
Posner task				
Physical Identity	.98 (.02)	617.79 (93.93)	.98 (.02)	605.19 (102.41)
Name Identity	.98 (.02)	699.50 (113.02)	.97 (.02)	704.38 (126.36)

Note. SRT = single choice reaction time task; CRT2/4 = choice reaction time task with two/four alternatives.

Table 2

Mean ERP latencies (SD in parentheses) averaged across conditions of each of the three experimental tasks.

	P100	N100	P200	N200	P300
			Session 1		
SRT/CRT	92.28 (23.52)	131.11 (14.88)	211.54 (32.82)	206.15 (27.71)	330.67 (44.26)
Sternberg	98.10 (27.83)	129.70 (27.02)	234.08 (34.48)	251.11 (42.05)	374.35 (74.76)
Posner	91.35 (37.39)	129.05 (25.71)	222.26 (33.74)	247.87 (36.80)	414.97 (86.45)
			Session 2		
SRT/CRT	104.21 (18.81)	128.11 (15.95)	208.44 (33.77)	210.38 (29.62)	324.40 (42.04)
Sternberg	94.82 (19.37)	132.53 (17.28)	230.35 (28.19)	248.48 (43.74)	382.39 (81.13)
Posner	124.78 (23.76)	140.76 (11.06)	218.16 (25.27)	240.02 (44.65)	377.74 (75.09)

Note. SRT/CRT = single and choice reaction time task.

Table 3

Model specifications: Structural equation model of reaction times and general intelligence.

Unstandardied regression weights							
			Estimate	S.E.	C.R.	P	Label
RT_S1	<---	mentalSpeed	1.000				
RT_S2	<---	mentalSpeed	1.000				
SRT	<---	Hick	1.129	.156	7.222	***	par_3
CRT2	<---	Hick	1.414	.152	9.313	***	par_4
CRT4	<---	Hick	1.000				
S1	<---	Sternberg	.345	.110	3.141	.002	par_5
S3	<---	Sternberg	.868	.098	8.835	***	par_6
S5	<---	Sternberg	1.000				
PI	<---	Posner	1.000				
NI	<---	Posner	1.000				
BIS	<---	g	1.000				
APM	<---	g	1.000				
SRT_RT_S1	<---	RT_S1	1.000				
CRT2_RT_S1	<---	RT_S1	1.000				
CRT4_RT_S1	<---	RT_S1	1.000				
S1_RT_S1	<---	RT_S1	1.000				
S3_RT_S1	<---	RT_S1	1.000				
S5_RT_S1	<---	RT_S1	1.000				
PI_RT_S1	<---	RT_S1	1.000				
NI_RT_S1	<---	RT_S1	1.000				
SRT_RT_S2	<---	RT_S2	1.000				
CRT2_RT_S2	<---	RT_S2	1.000				
CRT4_RT_S2	<---	RT_S2	1.000				
S1_RT_S2	<---	RT_S2	1.000				
S3_RT_S2	<---	RT_S2	1.000				
S5_RT_S2	<---	RT_S2	1.000				
PI_RT_S2	<---	RT_S2	1.000				
NI_RT_S2	<---	RT_S2	1.000				
SRT_RT_S1	<---	SRT	1.000				
SRT_RT_S2	<---	SRT	1.000				
CRT2_RT_S1	<---	CRT2	1.000				
CRT2_RT_S2	<---	CRT2	1.000				
CRT4_RT_S1	<---	CRT4	1.000				
CRT4_RT_S2	<---	CRT4	1.000				
S1_RT_S1	<---	S1	1.000				
S1_RT_S2	<---	S1	1.000				
S3_RT_S1	<---	S3	1.000				
S3_RT_S2	<---	S3	1.000				

S5_RT_S1	<---	S5	1.000				
S5_RT_S2	<---	S5	1.000				
PI_RT_S1	<---	PI	1.000				
PI_RT_S2	<---	PI	1.000				
NI_RT_S1	<---	NI	1.000				
NI_RT_S2	<---	NI	1.000				
C_Mean	<---	BIS	.506	.063	7.965	***	par_7
M_Mean	<---	BIS	.589	.063	9.274	***	par_8
PS_Mean	<---	BIS	.612	.063	9.644	***	par_9
PC_Mean	<---	BIS	1.000				
APModd	<---	APM	1.000				
APMeven	<---	APM	1.000				

Standardized regression weights

			Estimate
RT_S1	<---	mentalSpeed	1.000
RT_S2	<---	mentalSpeed	.921
SRT	<---	CRTSRT	.791
CRT2	<---	CRTSRT	1.000
CRT4	<---	CRTSRT	.828
S1	<---	Sternberg	.452
S3	<---	Sternberg	1.000
S5	<---	Sternberg	.866
PI	<---	Posner	1.000
NI	<---	Posner	.837
BIS	<---	g	1.000
APM	<---	g	.795
SRT_RT_S1	<---	RT_S1	.772
CRT2_RT_S1	<---	RT_S1	.774
CRT4_RT_S1	<---	RT_S1	.807
S1_RT_S1	<---	RT_S1	.846
S3_RT_S1	<---	RT_S1	.826
S5_RT_S1	<---	RT_S1	.768
PI_RT_S1	<---	RT_S1	.880
NI_RT_S1	<---	RT_S1	.862
SRT_RT_S2	<---	RT_S2	.797
CRT2_RT_S2	<---	RT_S2	.799
CRT4_RT_S2	<---	RT_S2	.830
S1_RT_S2	<---	RT_S2	.865
S3_RT_S2	<---	RT_S2	.847
S5_RT_S2	<---	RT_S2	.794
PI_RT_S2	<---	RT_S2	.895
NI_RT_S2	<---	RT_S2	.879
SRT_RT_S1	<---	SRT	.553
SRT_RT_S2	<---	SRT	.525
CRT2_RT_S1	<---	CRT2	.549
CRT2_RT_S2	<---	CRT2	.521

CRT4_RT_S1	<---	CRT4	.489
CRT4_RT_S2	<---	CRT4	.463
S1_RT_S1	<---	S1	.406
S1_RT_S2	<---	S1	.382
S3_RT_S1	<---	S3	.451
S3_RT_S2	<---	S3	.426
S5_RT_S1	<---	S5	.558
S5_RT_S2	<---	S5	.530
PI_RT_S1	<---	PI	.311
PI_RT_S2	<---	PI	.292
NI_RT_S1	<---	NI	.365
NI_RT_S2	<---	NI	.343
C_Mean	<---	BIS	.588
M_Mean	<---	BIS	.646
PS_Mean	<---	BIS	.661
PC_Mean	<---	BIS	1.000
APModd	<---	APM	.875
APMeven	<---	APM	.875

Covariances

			Estimate	S.E.	C.R.	P	Label
g	<-->	mentalSpeed	.228	.055	4.133	***	par_10

Correlations

			Estimate
g	<-->	mentalSpeed	.425

Variances

	Estimate	S.E.	C.R.	P	Label
mentalSpeed	.662	.093	7.102	***	par_11
CRTSRT	.166	.044	3.809	***	par_12
Sternberg	.262	.064	4.071	***	par_13
Posner	.083	.028	2.988	.003	par_14
g	.433	.056	7.746	***	par_15
SR1	.000				
R_1b	.000				
R_S3	.000				
R_PI	.000				
RBIS	.000				
SR2	.119	.020	6.056	***	par_16
R_0b	.127	.028	4.476	***	par_17
R_2b	.076	.021	3.575	***	par_18

R_S1	.121	.028	4.358	***	par_19
R_S5	.087	.029	2.992	.003	par_20
R_NI	.036	.018	1.963	.050	par_21
RAPM	.253	.047	5.374	***	par_22
e1	.110	.005	20.112	***	a
e2	.110	.005	20.112	***	a
e3	.110	.005	20.112	***	a
e4	.110	.005	20.112	***	a
e5	.110	.005	20.112	***	a
e6	.110	.005	20.112	***	a
e7	.110	.005	20.112	***	a
e8	.110	.005	20.112	***	a
e9	.110	.005	20.112	***	a
e10	.110	.005	20.112	***	a
e11	.110	.005	20.112	***	a
e12	.110	.005	20.112	***	a
e13	.110	.005	20.112	***	a
e14	.110	.005	20.112	***	a
e15	.110	.005	20.112	***	a
e16	.110	.005	20.112	***	a
e20	.210	.014	15.460	***	b
e19	.210	.014	15.460	***	b
e18	.210	.014	15.460	***	b
e17	.000				
e22	.210	.014	15.460	***	b
e21	.210	.014	15.460	***	b

Note. SRT = single choice reaction time task; CRT2/4 = choice reaction time task with two/four alternatives; S1_RT = RT in the set size one condition of the memory scanning task; S3_RT = RT in the set size three condition of the memory scanning task; S5_RT = RT in the set size five condition of the memory scanning task; PI = physical identity; NI = name identity SR = state residual; S1 = first laboratory session; S2 = second laboratory session.

*** $p < .001$

Table 4

Model specifications: Structural equation model of ERP latencies and general intelligence.

Unstandardized regression weights

			Estimate	S.E.	C.R.	P	Label
earlierERPs_S1	<---	earlierERPs	1.000				
earlierERPs_S2	<---	earlierERPs	1.000				
laterERPs_S1	<---	laterERPs	1.000				
laterERPs_S2	<---	laterERPs	1.000				
P1_S1	<---	earlierERPs_S1	1.000				
N1_S1	<---	earlierERPs_S1	1.000				
P1_S2	<---	earlierERPs_S2	1.000				
N1_S2	<---	earlierERPs_S2	1.000				
P2_S2	<---	laterERPs_S2	1.000				
N2_S2	<---	laterERPs_S2	1.000				
P3_S2	<---	laterERPs_S2	1.000				
P2_S1	<---	laterERPs_S1	1.000				
N2_S1	<---	laterERPs_S1	1.000				
P3_S1	<---	laterERPs_S1	1.000				
P2_S1	<---	P2	1.000				
P2_S2	<---	P2	1.000				
P3_S1	<---	P3	1.000				
P3_S2	<---	P3	1.000				
P1_S1	<---	P1	1.000				
P1_S2	<---	P1	1.000				
N2_S1	<---	N2	1.000				
N2_S2	<---	N2	1.000				
N1_S1	<---	N1	1.000				
N1_S2	<---	N1	1.000				
BIS	<---	g	1.000				
APM	<---	g	1.000				
SRTCRT_P1_S1	<---	P1_S1	1.000				
S_P1_S1	<---	P1_S1	1.000				
P_P1_S1	<---	P1_S1	1.000				
SRTCRT_N1_S1	<---	N1_S1	1.000				
S_N1_S1	<---	N1_S1	1.000				
P_N1_S1	<---	N1_S1	1.000				
SRTCRT_P2_S1	<---	P2_S1	1.000				
S_P2_S1	<---	P2_S1	1.000				
P_P2_S1	<---	P2_S1	1.000				
SRTCRT_N2_S1	<---	N2_S1	1.000				
S_N2_S1	<---	N2_S1	1.000				
P_N2_S1	<---	N2_S1	1.000				

SRTCRT_P3_S1	<---	P3_S1	1.000
S_P3_S1	<---	P3_S1	1.000
P_P3_S1	<---	P3_S1	1.000
SRTCRT_P1_S2	<---	P1_S2	1.000
S_P1_S2	<---	P1_S2	1.000
P_P1_S2	<---	P1_S2	1.000
SRTCRT_N1_S2	<---	N1_S2	1.000
S_N1_S2	<---	N1_S2	1.000
P_N1_S2	<---	N1_S2	1.000
SRTCRT_P2_S2	<---	P2_S2	1.000
S_P2_S2	<---	P2_S2	1.000
P_P2_S2	<---	P2_S2	1.000
SRTCRT_N2_S2	<---	N2_S2	1.000
S_N2_S2	<---	N2_S2	1.000
P_N2_S2	<---	N2_S2	1.000
SRTCRT_P3_S2	<---	P3_S2	1.000
S_P3_S2	<---	P3_S2	1.000
P_P3_S2	<---	P3_S2	1.000
SRTCRT_P1_S1	<---	SRTCRT_earlier	1.000
SRTCRT_N1_S1	<---	SRTCRT_earlier	1.000
SRTCRT_P1_S2	<---	SRTCRT_earlier	1.000
SRTCRT_N1_S2	<---	SRTCRT_earlier	1.000
S_P1_S1	<---	Sternberg_earlier	1.000
S_N1_S1	<---	Sternberg_earlier	1.000
S_P1_S2	<---	Sternberg_earlier	1.000
S_N1_S2	<---	Sternberg_earlier	1.000
P_P1_S1	<---	Posner_earlier	1.000
P_N1_S1	<---	Posner_earlier	1.000
P_P1_S2	<---	Posner_earlier	1.000
P_N1_S2	<---	Posner_earlier	1.000
SRTCRT_P2_S1	<---	SRTCRT_later	1.000
SRTCRT_N2_S1	<---	SRTCRT_later	1.000
SRTCRT_P3_S1	<---	SRTCRT_later	1.000
SRTCRT_P2_S2	<---	SRTCRT_later	1.000
SRTCRT_N2_S2	<---	SRTCRT_later	1.000
SRTCRT_P3_S2	<---	SRTCRT_later	1.000
S_P2_S1	<---	Sternberg_later	1.000
S_N2_S1	<---	Sternberg_later	1.000
S_P3_S1	<---	Sternberg_later	1.000
S_P2_S2	<---	Sternberg_later	1.000
S_N2_S2	<---	Sternberg_later	1.000
S_P3_S2	<---	Sternberg_later	1.000
P_P2_S1	<---	Posner_later	1.000
P_N2_S1	<---	Posner_later	1.000
P_P3_S1	<---	Posner_later	1.000
P_P2_S2	<---	Posner_later	1.000
P_N2_S2	<---	Posner_later	1.000
P_P3_S2	<---	Posner_later	1.000

C_Mean	<---	BIS	.584	.066	8.820	***	par_5
M_Mean	<---	BIS	.686	.068	10.150	***	par_6
PS_Mean	<---	BIS	.701	.068	10.346	***	par_7
PC_Mean	<---	BIS	1.000				
APModd	<---	APM	1.000				
APMeven	<---	APM	1.000				

Standardied regression weights

			Estimate
earlierERPs_S1	<---	earlierERPs	1.000
earlierERPs_S2	<---	earlierERPs	.669
laterERPs_S1	<---	laterERPs	1.000
laterERPs_S2	<---	laterERPs	1.000
P1_S1	<---	earlierERPs_S1	1.000
N1_S1	<---	earlierERPs_S1	1.000
P1_S2	<---	earlierERPs_S2	1.000
N1_S2	<---	earlierERPs_S2	1.000
P2_S2	<---	laterERPs_S2	.860
N2_S2	<---	laterERPs_S2	.871
P3_S2	<---	laterERPs_S2	.923
P2_S1	<---	laterERPs_S1	.860
N2_S1	<---	laterERPs_S1	.778
P3_S1	<---	laterERPs_S1	.923
P2_S1	<---	P2	.510
P2_S2	<---	P2	.510
P3_S1	<---	P3	.384
P3_S2	<---	P3	.384
N2_S1	<---	N2	.439
N2_S2	<---	N2	.491
BIS	<---	g	1.000
APM	<---	g	.774
SRTCRT_P1_S1	<---	P1_S1	.371
S_P1_S1	<---	P1_S1	.362
P_P1_S1	<---	P1_S1	.372
SRTCRT_N1_S1	<---	N1_S1	.371
S_N1_S1	<---	N1_S1	.362
P_N1_S1	<---	N1_S1	.372
SRTCRT_P2_S1	<---	P2_S1	.724
S_P2_S1	<---	P2_S1	.754
P_P2_S1	<---	P2_S1	.754
SRTCRT_N2_S1	<---	N2_S1	.757
S_N2_S1	<---	N2_S1	.786
P_N2_S1	<---	N2_S1	.786
SRTCRT_P3_S1	<---	P3_S1	.748

S_P3_S1	<---	P3_S1	.787
P_P3_S1	<---	P3_S1	.787
SRTCRT_P1_S2	<---	P1_S2	.512
S_P1_S2	<---	P1_S2	.501
P_P1_S2	<---	P1_S2	.513
SRTCRT_N1_S2	<---	N1_S2	.512
S_N1_S2	<---	N1_S2	.501
P_N1_S2	<---	N1_S2	.513
SRTCRT_P2_S2	<---	P2_S2	.724
S_P2_S2	<---	P2_S2	.754
P_P2_S2	<---	P2_S2	.754
SRTCRT_N2_S2	<---	N2_S2	.719
S_N2_S2	<---	N2_S2	.750
P_N2_S2	<---	N2_S2	.750
SRTCRT_P3_S2	<---	P3_S2	.748
S_P3_S2	<---	P3_S2	.787
P_P3_S2	<---	P3_S2	.787
SRTCRT_P1_S1	<---	SRTCRT_earlier	.362
SRTCRT_N1_S1	<---	SRTCRT_earlier	.362
SRTCRT_P1_S2	<---	SRTCRT_earlier	.334
SRTCRT_N1_S2	<---	SRTCRT_earlier	.334
S_P1_S1	<---	Sternberg_earlier	.416
S_N1_S1	<---	Sternberg_earlier	.416
S_P1_S2	<---	Sternberg_earlier	.386
S_N1_S2	<---	Sternberg_earlier	.386
P_P1_S1	<---	Posner_earlier	.355
P_N1_S1	<---	Posner_earlier	.355
P_P1_S2	<---	Posner_earlier	.328
P_N1_S2	<---	Posner_earlier	.328
SRTCRT_P2_S1	<---	SRTCRT_later	.282
SRTCRT_N2_S1	<---	SRTCRT_later	.267
SRTCRT_P3_S1	<---	SRTCRT_later	.313
SRTCRT_P2_S2	<---	SRTCRT_later	.282
SRTCRT_N2_S2	<---	SRTCRT_later	.284
SRTCRT_P3_S2	<---	SRTCRT_later	.313
C_Mean	<---	BIS	.660
M_Mean	<---	BIS	.718
PS_Mean	<---	BIS	.726
PC_Mean	<---	BIS	.937
APModd	<---	APM	.889
APMeven	<---	APM	.889

Covariances

			Estimate	S.E.	C.R.	P	Label
g	<-->	laterERPs	-.393	.057	-6.907	***	par_8

g	<-->	earlierERPs	.076	.020	3.833	***	par_9

Correlations

			Estimate
g	<-->	laterERPs	-.894
g	<-->	earlierERPs	.334

Variances

	Estimate	S.E.	C.R.	P	Label
Sternberg_later	.000				
Posner_later	.000				
P1	.000				
N1	.000				
earlierERPs	.124	.034	3.689	***	par_10
laterERPs	.462	.069	6.670	***	par_11
SRTCRT_earlier	.118	.048	2.478	.013	par_12
Sternberg_earlier	.163	.053	3.069	.002	par_13
Posner_earlier	.113	.048	2.365	.018	par_14
SRTCRT_later	.095	.026	3.682	***	par_15
P2	.162	.039	4.196	***	par_16
P3	.080	.027	2.989	.003	par_17
N2	.147	.041	3.548	***	par_18
g	.417	.059	7.063	***	par_19
SR_earlier_S1	.000				
SR_later_S1	.000				
SR_later_S2	.000				
SR_earlier_S2	.152	.044	3.433	***	par_20
R_P1S1	.000				
R_N1S1	.000				
R_P2S1	.000				
R_P3S1	.000				
R_P1S2	.000				
R_N1S2	.000				
R_P2S2	.000				
R_N2S2	.000				
R_P3S2	.000				
RBIS	.000				
R_N2S1	.154	.056	2.751	.006	par_21
RAPM	.278	.052	5.310	***	par_22
e1	.659	.031	21.497	***	a
e2	.659	.031	21.497	***	a

e3	.659	.031	21.497	***	a
e4	.659	.031	21.497	***	a
e5	.659	.031	21.497	***	a
e6	.659	.031	21.497	***	a
e7	.472	.022	21.941	***	b
e8	.472	.022	21.941	***	b
e9	.472	.022	21.941	***	b
e10	.472	.022	21.941	***	b
e11	.472	.022	21.941	***	b
e12	.472	.022	21.941	***	b
e13	.333	.021	16.119	***	c
e14	.333	.021	16.119	***	c
e15	.333	.021	16.119	***	c
e16	.659	.031	21.497	***	a
e17	.659	.031	21.497	***	a
e18	.659	.031	21.497	***	a
e19	.659	.031	21.497	***	a
e20	.659	.031	21.497	***	a
e21	.659	.031	21.497	***	a
e22	.472	.022	21.941	***	b
e23	.472	.022	21.941	***	b
e24	.472	.022	21.941	***	b
e25	.472	.022	21.941	***	b
e26	.472	.022	21.941	***	b
e27	.472	.022	21.941	***	b
e28	.333	.021	16.119	***	c
e29	.333	.021	16.119	***	c
e30	.333	.021	16.119	***	c
e34	.185	.013	14.688	***	d
e33	.185	.013	14.688	***	d
e32	.185	.013	14.688	***	d
e36	.185	.013	14.688	***	d
e35	.185	.013	14.688	***	d
e31	.058	.016	3.745	***	par_23

Note. SRTCRT = single and choice reaction time task; S = Sternberg memory scanning task; P = Posner letter matching task; SR = state residual; S1 = first laboratory session; S2 = second laboratory session.

*** $p < .001$

Erklärung gemäß § 8 Abs. (1) c) und d) der Promotionsordnung der Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften

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.

Vorname Nachname ____

Datum, Unterschrift ____