**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

Measuring Cognitive Control through Neurocognitive Process Parameters to Understand Individual Differences in Intelligence

presented by
Christoph Löffler

year of submission
2025

Dean: Prof. Dr. Guido Sprenger
Supervisors: Prof. Dr. Anna-Lena Schubert, Prof. Dr. Andreas Voß

Für meine Eltern

Vorwort

An diesem Punkt möchte ich einigen Menschen meinen Dank aussprechen, die dazu beigetragen haben, dass diese Arbeit möglich wurde, durchgeführt werden konnte und nun in Form der vorliegenden Dissertation abgeschlossen wurde.

Um ein Bild zu verwenden, das mich während der ereignisreichen Zeit meiner Promotion begleitet hat: Ich fühle mich wie Frodo aus *Der Herr der Ringe*, nachdem er den einen Ring im Schicksalsberg vernichtet hat. Der Weg war nicht immer leicht und voller Hindernisse. Doch jeder einzelne Schritt hat sich gelohnt, und ich habe dabei nicht nur fachlich, sondern auch persönlich viel gelernt.

Das alles wäre jedoch nicht möglich gewesen, wenn mich nicht viele wunderbare Menschen begleitet und unterstützt hätten.

Zunächst einmal: Vielen Dank, Anna-Lena Schubert! Ohne dich als meine Doktormutter würde es diese Arbeit nicht geben. Du hast mir die notwendigen Ressourcen, deine Zeit, dein unermessliches Wissen, deine Geduld und dein Vertrauen geschenkt, um diese Dissertation zu verwirklichen. Du bist nicht nur meine Doktormutter, sondern auch eine gute Freundin. Danke dir!

Mein besonderer Dank gilt auch zwei weiteren Personen aus meinem Forschungsteam: Dirk Hagemann und Gidon Frischkorn. Ihr beide habt eine wichtige Mentorenrolle für mich eingenommen. Dirk, dir danke ich für die Möglichkeit, viele Jahre in deiner Abteilung gearbeitet haben zu dürfen. Du hast mir die nötigen Ressourcen bereitgestellt und standest mir jederzeit mit Rat und Tat, Geduld und deinem umfassenden Wissen zur Seite. Gidon, auch dir danke ich für dein wertvolles Wissen, deine Zeit und die Ratschläge, die meine Arbeit stets vorangebracht haben. Vielen Dank euch beiden!

Solche Studien wären ohne wissenschaftliche Hilfskräfte, die viel Zeit, Mühe und persönlichen Einsatz in die Datenerhebung stecken, nicht möglich. Ein großes Dankeschön gilt daher allen, die mich auf dieser Reise begleitet und bei den vielen Projekten unterstützt haben: Johanna Hein, Florian Kaulhausen, Jan Göttmann, Larissa Kunoff, Johannes Marasek, Amelie Duru und allen weiteren Hilfskräften, die ich hier möglicherweise vergessen habe.

Ein weiterer Dank geht an meine Freundinnen, Freunde und Kolleginnen sowie Kollegen aus Heidelberg und Mainz, die mir in dieser Zeit mit Rat und Tat zur Seite gestanden haben. Besonders möchte ich hier Marianne Beschorner, Kathrin Sadus, Jan Göttmann, José Alanis und Wanja Hemmerich hervorheben. Danke euch allen – und vor allem danke dir, lieber Wanja, für deine Zeit und deine nützlichen Vorschläge beim Korrekturlesen dieser Arbeit!

Aber auch abseits der Arbeit gibt es viele Menschen, die mich über die Jahre meiner Promotion hinweg begleitet und unterstützt haben, auch wenn ich es nicht ansatzweise schaffen werde, hier allen gerecht zu werden. Danke an meine Familie und insbesondere an meine Eltern, die immer an mich geglaubt haben und mich in allem, was ich tue, unterstützen. Papa, auch wenn du diese Worte leider nicht mehr lesen wirst und das Endprodukt meiner Arbeit niemals sehen kannst, weiß ich, dass du immer überzeugt davon warst, dass ich es schaffen werde. Du und Mama, ihr habt einen großen Teil zu dieser Arbeit beigetragen. Ohne euch beide, liebe Mama und lieber Papa, wäre ich nicht der Mensch, der ich heute bin, und diese Arbeit hätte es nie gegeben. Danke!

Zudem möchte ich zwei weitere Personen hervorheben: Liebe Leonie, danke, dass du mich und meine Arbeit so viele Jahre lang begleitet, mitgetragen und unterstützt hast. Danke auch an deine Familie. Liebe Katja, dir danke ich für deine unendliche Geduld, dein offenes Ohr und deine alltägliche Unterstützung, ohne die ich diese Arbeit nicht in dieser Form hätte abschließen können. Danke euch beiden!

Auch wenn ich jetzt nicht alle wichtigen Menschen in meinem Leben in der ihnen gebührenden Fülle erwähnen kann, gilt mein Dank dennoch euch allen, die mich in dieser Zeit unterstützt haben. Von ganzem Herzen: Danke!

Table of contents

List of the scientific publications included in this publication-based dissertation

Manuscript I.

Löffler, C., Frischkorn, G. T., Rummel, J., Hagemann, D., & Schubert, A. L. (2021). Do attentional lapses account for the worst performance rule? *Journal of Intelligence*, *10*(1), 2. https://doi.org/10.3390/jintelligence10010002

Manuscript II

Löffler, C., Frischkorn, G. T., Hagemann, D., Sadus, K., & Schubert, A. L. (2024). The common factor of executive functions measures nothing but speed of information uptake. *Psychological Research, 88*(6), 1092–1114. https://doi.org/10.1007/s00426-023-01924-7.

Manuscript III

Löffler, C., Sadus, K., Frischkorn, G. T., Hagemann, D., & Schubert, A. L. (2024). The factor structure of executive functions measured with electrophysiological correlates: An event-related potential analysis. *PsyArXiv Preprints*. https://doi.org/10.31234/osf.io/kfqt4

This manuscript was submitted to *Journal of Experimental Psychology: Learning, Memory, and Cognition* on July 9, 2024, and is currently under peer review.

Theoretical background and introduction

Human intelligence is defined as the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, [and] to overcome obstacles by taking thought" (Neisser et al., 1996, p. 77). Consequently, two key derivations emerge from this definition.

First, the broad advantages associated with intelligence lead more intelligent individuals to experience positive outcomes throughout their lives. Empirically, this is evident in better academic performances (e.g., Lozano-Blasco et al., 2022), improved job performance (e.g., Schmidt & Hunter, 2004), better general health (e.g., Der et al., 2009), increased well-being (e.g., Pesta et al., 2010), and an extended lifespan (e.g., Deary, 2008). These and similar findings highlight the importance of intelligence research for a broader range of fields within psychology, beyond cognitive psychological research.

Second, intelligence is not a single ability, but it is complex and encompasses various domains, such as mathematical-, figural-, and verbal abilities. These abilities are positively correlated, whereby the positive correlations can be described by a hierarchical structure with one common factor, *g* (Spearman, 1904). This *g* or *general intelligence* factor describes the shared variance between different tasks measuring the different components of intelligence as mentioned in some theories (e.g., Carroll, 1993; Horn & Cattell, 1966; Jäger et al., 1997; Spearman, 1904; Vernon, 1950). However, the specific cognitive and neurophysiological processes underlying performance in intelligence tasks remain unclear. In the current cognitive psychological and electrophysiological research, three cognitive processes are mainly discussed as important in explaining individual differences in intelligence: (1) information processing speed, (2) working memory, and (3) cognitive control. Thereby, these three processes may independently contribute to individual differences in intelligence or interact in their contribution to *g*.

Information processing speed and intelligence

One frequently discussed predictor of interindividual differences in intelligence is *information processing speed*. Jensen (2006) regarded information processing speed as a key factor in explaining these differences. Similarly, Carroll (1993), emphasized the fundamental role of information processing speed in determining individual differences in intelligence, in his seminal publication. Carroll conducted a meta-analysis of data from 460 datasets, detailing its factor structure. Based on this analysis, he developed the *Three-Stratum Theory*, also known as the *Cattell-Horn-Carroll* model, a hierarchical model of intelligence with the *g*-factor at its apex. He identified two speed-related factors underlying the *g*-factor: processing speed and decision speed. In our work, we refer to the decision speed factor when examining information processing speed as an underlying process of intelligence, whereas other studies have focused on the processing speed factor (e.g., Draheim et al., 2021).

Typically, in cognitive psychological research, individuals' information processing speed is measured by using reaction times (RTs), inspection times, or parameters from computational mathematical models (e.g., the drift rate parameter v from the drift-diffusion model; Ratcliff, 1978) as behavioral indicators in tasks where participants are required to make decisions. However, real-life decision-making comprises many processes preceding the final decision. This makes it challenging to measure information processing speed accurately. In complex decision-making tasks, various processes—beyond information processing—can influence RTs until a decision is made. Therefore, using simple decision tasks minimizes the impact of unrelated processes. Different strategies should not influence participants' performance.

In psychological research, information processing speed is typically assessed with RTs in elementary cognitive tasks (ECTs). These tasks are experimentally conceptualized to control for task complexity and reduce the influence of non-information processing speed-related processes on the dependent variable (Jensen, 2006). Typical ECTs are the Hick paradigm (Hick, 1952), where participants are required to indicate the position of an appearing stimulus out of a set of eight possible positions; the Posner letter matching task (Posner & Mitchell, 1967), where participants had to state whether two presented letters are physically or namely identical or not; and the Sternberg short term memory scanning task (Sternberg, 1969), where participants had to decide whether a digit was part of a sequence of numbers that had been presented earlier.

Neubauer and Knorr (1998) found substantial negative correlations up to $r = -.70$ between paper and pencil versions of these three tasks and several measures of intelligence, which make the three tasks a valuable battery of ECTs assessing information processing speed (Neubauer & Knorr, 1998). Empirically, the robustness of the relationship between information processing speed and intelligence is well established, as shown in the substantial body of research spanning many different papers, reviews, and meta-analyses employing behavioral scores (e.g., Doebler & Scheffler, 2016; Frischkorn et al., 2019; Neubauer & Knorr, 1997, 1998; Schmiedek et al., 2007; Schubert et al., 2015; Sheppard & Vernon, 2008). These findings suggest that more intelligent individuals benefit from faster information processing.

Furthermore, the robustness in the correlations between information processing speed and intelligence is not only evident at the level of behavioral scores but also when information processing speed is assessed using electrophysiological measures. (Schubert et al., 2017; Schubert, Löffler, Hagemann, et al., 2023). The authors quantified information processing speed with the latencies of the event-related potential (ERP) as process parameters reflecting neural information processing speed measured with the electroencephalograph (EEG). More precisely, they used the latencies from the P3, P2, and N2 ERP components measured in a battery of ECTs as indicators of neural information processing speed. Both papers found strong correlations between a latent intelligence factor and a latent neural information processing factor, ranging from $-.89 \leq r \leq -.49$, suggesting that more intelligent individuals show faster neural responses in the corresponding ERP components (Schubert et al., 2017; Schubert, Löffler, Hagemann, et al., 2023). Together, information processing speed accounts for up to 80% of the variance in *g* (Schubert et al., 2017), making it a key factor in understanding individual differences in intelligence.

However, not all studies have found empirical evidence supporting information processing speed as a key factor in explaining individual differences in intelligence. In a latent variable analysis, Conway et al. (2002) demonstrated that working memory-related processes, not information processing speed, were correlated with intelligence. Furthermore, Schubert et al. (2018) conducted an experimental study and manipulated participants' information processing speed by pharmacological intervention with nicotine patches. They observed that, as a consequence of this intervention, the participants became faster in their responses and in their neural information processing speed, as measured by ERP latencies. However, there was no evidence for improved performance on intelligence test scores (Schubert et al., 2018). The studies by Conway et al. (2002) and Schubert et al. (2018) provided empirical evidence suggesting either no relationship or a non-causal relationship between information processing speed and intelligence.

Working memory capacity and intelligence

Beyond its crucial role in information processing speed, working memory is also widely considered to be a fundamental cognitive process underlying individual differences in intelligence. Working memory refers to a memory system that enables individuals to access and use information necessary for completing an ongoing task (Wilhelm et al., 2013). More specifically, Oberauer et al. (2008) conceptualized working memory as a cognitive system wherein multiple chunks of information are simultaneously represented and accessible. Various cognitive operations can be performed on these chunks, including combining them to create new associations.

The capacity of working memory, referred to as working memory capacity (WMC), describes the number of chunks that can be held and accessed simultaneously (Oberauer et al., 2008). This capacity limit restricts the creation of novel associations and relationships between the information in working memory. Such processes are considered as critical operations required in intelligence tests. Thus, WMC is a strong predictor of individual differences in intelligence.

In addition to working memory, other memory systems exist, such as short-term and long-term memory. What differentiates working memory is its ability to actively manipulate and integrate its contents with other information. In principle, working memory functions as a workspace where information is processed, modified, and adjusted. Moreover, additional information from short-term and long-term memory systems can be integrated into working memory at any time.

Since the last decade of the twentieth century, many researchers have examined the relationships between intelligence and WMC. One of the first seminal studies by Kyllonen and Christal (1990) revealed very strong correlations between these two constructs. They assessed different tests measuring intelligence as well as WMC in four samples of US military recruits and found very strong latent correlations ranging from, $.80 \leq r \leq .88$, suggesting that more intelligent individuals possess higher storage potentials in their working memory (Kyllonen & Christal, 1990). In subsequent years, many studies replicated these positive latent relationships between intelligence and WMC in independent samples (e.g., Ackerman et al., 2002; Buehner et al., 2005; Colom et al., 2008; Conway et al., 2002; Engle et al., 1999; Frischkorn et al., 2019; Gignac, 2014; Oberauer et al., 2005, 2008; Unsworth et al., 2014), despite fluctuations in the absolute values of the correlation coefficients. Furthermore, two meta-analyses found strong latent correlations between WMC and intelligence, with $r = .71$ (Kane et al., 2005) and $r = .85$ (Oberauer et al., 2005).

However, until two years ago, it remained unclear whether WMC is the underlying mechanism driving individual differences in intelligence or, conversely, if the direction of causality is reversed, as the majority of existing findings and interpretations were based on purely correlational analyses. Two recently published studies by Hagemann et al. (2023) and Schubert, Löffler, Sadus, et al. (2023) experimentally increased participants' working memory load. The results demonstrated a significant reduction in intelligence test performance compared to an active control group, which did not experience any additional working memory load due to the experimental manipulation. The experimental designs of both studies were grounded in the theoretical assumptions of the *multi-component* model of working memory (Baddeley & Hitch, 1974). In its earliest iteration, Baddeley and Hitch's (1974) theory emphasizes that the working memory system consists of three components: (1) the visuospatial sketchpad, (2) the phonological loop, and (3) the central executive. The visuospatial sketchpad and the phonological loop are considered short-term memory systems for visual and spatial material, respectively (Baddeley & Hitch, 1974)[1]. The central executive, however, is the core component of the multi-component model of working memory. It integrates information from the visuospatial sketchpad and the phonological loop, maintains information, and directs goal-directed attention. The central executive becomes

[1] In a later revision of the model, Baddeley (2000) added the episodic buffer as a further component, responsible for integrating visual and auditive information with contents from long-term memory.

engaged in information processing when the task at hand exceeds the processing capacities of both short-term memory systems.

In the studies by Hagemann et al. (2023) and Schubert, Löffler, Sadus, et al. (2023), participants completed the Raven matrices (Raven & Raven, 2003) while both the experimental and control groups were required to spell out numbers aloud in one-second intervals. The experimental group was tasked with generating random numbers within a specified range, while the control group repeated a fixed sequence of numbers. In line with Baddeley and Hitch's (1974) multi-component model of working memory, the random number generation in a one-second rhythm is a cognitively demanding task, requiring the central executive's additional processing capacities (Baddeley, 1990). This increases the load on working memory and reduces participants' capacity for the primary reasoning task (Gilhooly et al., 1993; Klauer et al., 1997). In contrast, according to Baddeley and Hitch's (1974) model, repeating a simple sequence of given numbers is considered an easier task processed within the phonological loop buffer system. As a result, it does not increase the working memory load or affect performance on the primary reasoning task (Gilhooly et al., 1993). The studies by Hagemann et al. (2023) and Schubert, Löffler, Sadus, et al. (2023) experimentally demonstrated that working memory is a fundamental process underlying intelligence, and that individual differences in WMC causally predict differences in intelligence. Over the past few decades, cognitive psychological research has amassed a substantial body of empirical evidence showing high correlations between WMC and intelligence (e.g., Ackerman et al., 2002; Buehner et al., 2005; Colom et al., 2008; Conway et al., 2002; Engle et al., 1999; Frischkorn et al., 2019; Gignac, 2014; Kyllonen & Christal, 1990; Oberauer et al., 2005, 2008; Unsworth et al., 2014).

In their latent variable analyses, Kyllonen and Christal (1990) demonstrated that WMC shares up to 81% of its variance with general intelligence. Moreover, intelligence also shares substantial variance with neural information processing speed—with estimates ranging from 61% (Schubert, Löffler, Hagemann, et al., 2023) to 79% (Schubert et al., 2017). Both WMC and information processing speed explain more than 50% of the variance in general intelligence, suggesting that the contributions of these two processes to individual differences in intelligence are not entirely independent.

Do cognitive control processes serve as potential candidates to bridge the gap between WMC and information processing speed in their contribution to intelligence?

In light of these empirical findings, the question arises whether a single process mediates the relationships between information processing speed and intelligence, as well as the relationship between WMC and intelligence. Several theoretical frameworks propose that cognitive control processes may serve as key mechanisms that integrate information processing speed, WMC, and intelligence. Cognitive control is an umbrella term (Diamond, 2013) that encompasses top-down regulated processes that guide an individual's attention and filter out irrelevant or distracting information during goal-directed information processing. These processes are also commonly termed as executive functions (EF; e.g., Miyake et al., 2000), executive attention (e.g., Engle, 2002; Engle & Kane, 2004; Kane et al., 2008), or attention control (e.g., Unsworth & Spillers, 2010) depending on the specific theoretical framework or focus of research[2]. Furthermore, cognitive control processes can be further subdivided into specific abilities such as inhibition, shifting, and updating, with these abilities collectively referred to as EFs (Friedman et al., 2008; Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012; Rey-Mermet et al., 2018). Inhibition refers to the ability to focus attention on the current task and its goals while ignoring irrelevant or distracting information. Shifting is the ability to switch between different tasks or mindsets, and updating

[2] Depending on the theoretical context, these terms are used synonymously for cognitive control processes in this manuscript.

involves monitoring current memory content and storing new information (Miyake et al., 2000; von Bastian et al., 2020).

Cognitive control, the most important processing component of working memory

Working memory differs from short-term memory in its ability to manage concurrent interfering information while maintaining stored information chunks in short-term memory. In everyday situations, remembering a small number of items, such as a shopping list of five items to buy at the supermarket, is typically not demanding, when no additional tasks are involved. This example would be considered a purely short-term memory task. However, the introduction of additional processing steps increases cognitive demand, making the task more challenging.

For example, if the task is to go to the supermarket and find the cheapest product for each of the five items, with the items arranged in a different order on the shelves than how you originally memorized them, this task no longer purely relies on short-term memory. The introduction of additional tasks, such as comparing prices and adjusting to the different shelf arrangements of the items, makes it more challenging to maintain the five items in memory. In this context, the distinction between short-term memory and working memory becomes clear: short-term memory refers to the simple retention of information for a brief period, without any additional cognitive manipulation. In contrast, working memory involves both the storage and simultaneous processing of information. The cognitive demands of comparing prices and navigating through the store engage working memory, as it must manage and manipulate information about the items while performing the additional tasks.

WMC is typically measured using complex-span tasks, such as the reading-span task (Daneman & Carpenter, 1980) or the operation-span task (Turner & Engle, 1989). These tasks are designed to engage participants in two concurrent processes: a processing component and a storage component. For instance, in Turner and Engle's (1989) operation span task, participants must determine whether the result of a mathematical equation involving simple multiplication and summation is correct (processing component). Following each equation, one word is presented, which the participant must remember (storage component). In the original version, the set size of words to be remembered ranges from two to five. This task challenges working memory by requiring participants to actively maintain the presented words in mind while simultaneously verifying the correctness of the equations.

The theoretical foundation for complex-span tasks lies in the working memory model by Baddeley and Hitch (1974). In their seminal work, they defined working memory as a system for temporarily storing a limited amount of information while performing ongoing tasks. In their multi-component model of working memory, the central executive is considered the main component. The central executive serves as the cognitive control mechanism of the working memory system, regulating and integrating information from the three subcomponents: (1) the visuospatial sketchpad, (2) the episodic buffer, and (3) the phonological loop, ensuring that relevant information is briefly stored and accessible for the ongoing task goals (Baddeley, 2010; Baddeley & Hitch, 1974). The central executive enables participants to evaluate the equations in a complex-span task while ensuring that the memorized words are stored and remain retrievable.

Building on the theoretical framework of Baddeley and Hitch (1974), subsequent concepts and theories have incorporated cognitive control processes as central components in explaining the capacity limits of the working memory system. One such model is the *time-based resource-sharing* model of working memory by Barrouillet et al. (2004). The core idea of the time-based resource-sharing model is that contents in working memory are maintained by a continuous refreshing process. If a stored item is not refreshed within a certain timeframe, its activation falls below a decay threshold, resulting in forgetting. The efficiency of refreshing depends on two factors: (1) the available free time and (2) the amount of content stored in working memory. The more free time there is for refreshing, the less likely it is that these contents will be forgotten. Conversely, if too many items are stored in working memory or if new content must be encoded too quickly at an accelerated rate, the time available to refresh each stored item decreases, leading to forgetting and limiting WMC. In the time-based resource-sharing model, attentional and shifting processes play a key role. The focus of attention activates working memory contents during retrieval, necessitating a shift of attentional resources between processing and retrieval tasks. The greater the activation an item receives during retrieval through focused attention, and the more efficiently attention can shift between these tasks, the more effectively items can be refreshed and maintained in working memory (Barrouillet et al., 2004).

For example, in the operation-span task, participants are required to alternate their attention between solving equations and remembering words. According to the time-based resource-sharing model, after solving each equation, participants have a brief period to shift their attention back to the stored words to refresh their activation levels. If the number of sequences (equations and words) is too high, or the sequences are presented too quickly (less free time), the time available for refreshing the words is reduced (Barrouillet et al., 2004). Consequently, the activation of the already stored words may fall below the decay threshold, leading to forgetting. However, the centrality of switching processes in WMC has been questioned, as Oberauer et al. (2003) found only small correlations ($r = .30$) between these cognitive control processes and WMC.

The most prominent theoretical framework highlighting cognitive control processes as the core component for explaining individual differences in WMC is the *executive attention* model, proposed by Randall Engle’s research group. Kane and Engle (2002) described executive attention as the ability to actively maintain memory representations, such as action plans, task goals, or task-relevant stimuli, even in the presence of interference. They emphasized that this active maintenance of information is particularly crucial in high-interference situations, where the risk of retrieving incorrect information or generating incorrect responses is greater (Kane & Engle, 2002). Executive attention helps maintain focus on task goals while effectively blocking out irrelevant or distracting information during task performance (Engle, 2002; Engle & Kane, 2004; Kane et al., 2008; Kane & Engle, 2002).

Executive attention is often seen as a broader concept than inhibition, as defined by Miyake et al. (2000). While inhibition specifically refers to the suppression of prepotent responses, executive attention includes the protection of working memory and attentional focus from various internal and external sources of distraction. Nevertheless, both share similarities, as many studies measure executive attention using classical inhibition tasks, such as the Stroop task (e.g., Kane & Engle, 2003), go/no-go task (Redick et al., 2011), or the antisaccade task (e.g., Kane et al., 2001).

To illustrate the role of executive attention within the working memory system, let us revisit the operation-span task. Executive attention enables participants to maintain focus on the primary task goals—solving equations accurately and remembering words. It supports the active maintenance of the memorized words in working memory by shielding them from the interference caused by the equation-solving process. Additionally, participants with higher executive attention abilities process the equations more efficiently. This dual advantage allows them to store more words correctly while evaluating the equations more accurately.

Following the executive attention framework, both storage processes (short-term memory) and executive attention abilities underlie individual differences in WMC (Engle et al., 1999). Numerous empirical findings support the assumptions outlined in the executive attention account. For instance, extreme group comparisons revealed that participants with lower performance in complex-span tasks also performed worse on various executive attention tasks, indicating that lower WMC is empirically associated with lower executive attention abilities (Conway et al., 2001; Kane et al., 2001; see also Redick et al., 2011). Additionally, studies using latent variable analyses with structural equation modeling have found that WMC consists of two components. By controlling a WMC factor for individual differences in storage capabilities, a residual factor representing executive attention abilities emerged (e.g., Conway et al., 2002; Engle et al., 1999; Unsworth et al., 2010; Unsworth & Spillers, 2010). Also, Unsworth and Spillers (2010) demonstrated that both storage and cognitive control processes are equally significant predictors of WMC, leading to their *dual-component model* of working memory.

Although cognitive control processes (e.g., executive attention) appear to be significant predictors of individual differences in WMC, Colom et al. (2008) demonstrated through latent variable analysis that neither information processing speed nor cognitive control processes, but rather the storage component of short-term memory, predicted individual performances in working memory tasks. Other researchers have also questioned the centrality of cognitive control processes in explaining WMC differences. For example, Wilhelm et al. (2013) argued that the ability to form elementary bindings between several chunks of information is the key factor underlying individual differences in WMC. They suggested that the stronger an individual’s ability to form such bindings, the better these chunks of information can be retrieved, ultimately determining an individual’s WMC.

While some theories of working memory emphasize short-term memory processes as a fundamental explanation for differences in working memory, the majority of theories attribute a central role to cognitive control processes in maintenance and processing of information in working memory. Although these theories emphasize different aspects of cognitive control, they consistently consider these processes to be the fundamental basis of working memory and distinguish them from simple storage components such as short-term memory.

Cognitive control and intelligence

Cognitive control processes play a crucial role as predictor variables for individual differences in intelligence with previous research demonstrating positive correlations between cognitive control processes and intelligence (e.g., Burgoyne et al., 2023; Conway et al., 2002; Draheim et al., 2021, 2023; Engle et al., 1999; Friedman et al., 2006, 2008, 2011; Friedman & Miyake, 2017; Kane et al., 2004).

As previously mentioned, Engle et al. (1999) examined the role of cognitive control processes (executive attention) in both working memory and intelligence. In their executive attention framework, they argued that working memory comprises both storage and cognitive control processes. Through hierarchical factor analysis, Engle et al. (1999) separated storage processes from the working memory factor, leaving residual variance attributed to cognitive control processes. These cognitive control processes revealed a strong correlation with intelligence (r = .49), with the storage factor showing a weaker correlation to intelligence (r = .29). The authors concluded that cognitive control processes give rise to the relationship between intelligence and WMC. Subsequent research by Conway et al. (2002)[3] replicated these findings, supporting Engle et al.'s (1999) view that cognitive control is a key factor in explaining individual differences in WMC and serves as a robust predictor of intelligence.

Why is cognitive control necessary for accurate performance on working memory tasks and intelligence tests? In simpler tasks involving automatized or basic storage processes, like learning a short-term memory sequence, there are no significant demands on cognitive

[3] It is important to note that Conway et al. (2002) found no significant correlation between the storage factor and intelligence.

control processes from the central executive (Baddeley & Hitch, 1974). In contrast, working memory tasks like the operation-span task require control processes from the central executive to maintain focus on goal-relevant information, even in the presence of distractors or concurrent tasks (Baddeley & Hitch, 1974; Turner & Engle, 1989). Similarly, intelligence tests can be seen as complex tasks that require participants to process various types of information simultaneously while maintaining focus on goal-relevant information despite distractions and interference (Conway et al., 2002; Engle et al., 1999). According to Engle et al. (1999) the relationship between WMC and intelligence is explained by cognitive control processes. Executive attention is crucial in both working memory tasks and intelligence tests, as it helps shield relevant information from distractions and keeps it in an active, retrievable state (Engle et al., 1999; Mashburn et al., 2023).

Instead of a single underlying mechanism of cognitive control processes, Shipstead et al. (2016) proposed a theory explaining the relationship between WMC and intelligence through two independent processes: *maintenance* and *disengagement*. These processes are top-down regulated by cognitive control and are essential for both working memory tasks and intelligence tests. Performance in these tasks depends on participants' ability to maintain relevant information and disengage from no longer relevant information. Cognitive control aims to balance these processes, as excessive maintenance can lead to information overload, while excessive disengagement results in the rapid loss of important information (Shipstead et al., 2016). These processes, however, are not equally involved in every task. Working memory tasks require a higher degree of maintenance, whereas intelligence test tasks demand more disengagement (Mashburn et al., 2023; Shipstead et al., 2016).

In addition to theoretical accounts proposing cognitive control processes as underlying mechanisms for both working memory and intelligence, Kovacs and Conway (2016) introduced the process-overlap theory (POT), which posits that cognitive control processes –

EFs – are the central component explaining individual differences in intelligence. According to POT, these EFs are domain-general, meaning they are not specific to any one task but are needed across all tasks, regardless of the specific abilities being measured (e.g., verbal abilities, mathematical abilities, reading comprehension). The authors argue that these EFs or cognitive control processes act as a bottleneck for the performance in each ability domain. Consequently, domain-general executive abilities contribute to the positive manifold observed in the *g*-factor of intelligence, making EFs the determining factor for individual differences in higher-order cognitive abilities, such as WMC and intelligence (Kovacs & Conway, 2016).

In addition to these theoretical accounts, further empirical findings by Friedman, Miyake, and colleagues revealed significant correlations between cognitive control processes and intelligence (see also Kane et al., 2004). While Friedman et al. (2006) found positive correlations only between updating and intelligence, Friedman et al. (2008) identified correlations between intelligence, shifting, and updating, as well as a common factor of EFs. These theoretical frameworks, along with empirical evidence, provide a deeper understanding of the critical role that cognitive control processes play as underlying mechanisms for individual differences in higher-order cognitive processes such as WMC and intelligence. However, several later studies failed to find an association between EFs and intelligence (e.g., Frischkorn et al., 2019; Rey-Mermet et al., 2019). The implications will be discussed below.

Cognitive control and information processing speed

The relationships between cognitive control processes and information processing speed remains a topic of ongoing research, with various studies offering contradictory evidence regarding their predictive power for individual differences in intelligence (e.g., Conway et al., 2002; Frischkorn et al., 2019). Research on cognitive control processes and information processing speed originates from different streams of cognitive psychological research, and the notion that they may share a common basis is relatively new. Mashburn et al. (2023) provided an excellent summary of the history of both constructs and their contribution to individual differences in WMC and intelligence. Frischkorn et al. (2019) suggested that differences in information processing speed might contribute to differences in cognitive control processes or vice versa. Furthermore, they proposed that cognitive control processes could bridge the gap between the substantial contributions of WMC and information processing speed to individual differences in intelligence, as several studies have shown that each of these two factors explains more than 50% of the variance in intelligence (Kyllonen & Christal, 1990; Oberauer et al., 2005; Schubert et al., 2017; Schubert, Löffler, Hagemann, et al., 2023).

To date, only a limited number of studies have empirically tested this hypothesis. One notable example is Conway et al. (2002), who examined the relationships between WMC, short-term memory, information processing speed, and their unique contributions to intelligence at the latent level. To isolate cognitive control-specific variance, they controlled a WMC factor for individual differences in short-term memory based on the executive attention framework of WMC by Engle al. (1999). Their findings revealed that cognitive control processes were strong predictors of individual differences in intelligence, whereas neither the unique components of short-term memory nor information processing speed significantly predicted intelligence (Conway et al., 2002).

These findings contrast with Frischkorn et al. (2019), who used latent-level analysis to separate cognitive control processes from information processing speed. By employing various cognitive control tasks and ECTs, they aimed to examine their unique contributions to higher-order cognitive abilities, such as WMC and intelligence. They found that isolating cognitive control-specific variances was challenging, with only small amounts of variance attributed to cognitive control. At the latent level, only information processing speed significantly predicted WMC and intelligence (Frischkorn et al., 2019).

Importantly, the differing results between these two studies may be due to their different conceptualizations of information processing speed. Conway et al. (2002) focused on perceptual speed, while Frischkorn et al. (2019) used tasks that measured reaction and decision speed. According to the Cattell-Horn-Carroll model (Carroll, 1993)—a hierarchical factor model that describes the underlying abilities of intelligence—these two types of tasks represent distinct factors of information processing speed, each making independent contributions to explaining differences in intelligence. Nonetheless, the differing results between Conway et al. (2002) and Frischkorn et al. (2019) highlight the mixed evidence regarding the relationship between cognitive control and information processing speed.

One potential explanation for these inconsistent findings could be that cognitive control processes may not be adequately reflected in individuals' overall RTs, as often measured in tasks (e.g., mean or median RTs) but instead are particularly evident in their slower or slowest RTs. According to Kane et al. (2008), executive attention processes help maintain task goals and shield relevant information from distractions. Attentional lapses may occur when there is a momentary decline in executive attention during a task, resulting in either a quick but incorrect response or slow yet correct decisions if the individual successfully refocuses their executive attention (Kane et al., 2008). Individuals with higher executive attentional resources experience fewer attentional lapses and, when such lapses occur, they can redirect their attention back to task-relevant information more quickly than those with lower executive attention abilities.

Thus, focusing on slower RTs, rather than on mean RTs may provide more valuable insights into individual differences in cognitive control processes, as such differences are likely reflected in the variability of slower responses (Kane et al., 2008; McVay & Kane, 2012; Unsworth et al., 2010). This focus on slower RTs may help clarify the relationship between cognitive control and information processing speed, as not all measures of information processing speed, but particularly the variability in slower responses, seem to capture processes specific to cognitive control.

Do differences in executive attentional lapses account for the worst performance rule (Manuscript 1)

As discussed before, it is well established that mean RTs are negatively correlated with higher-order cognitive abilities, such as WMC and intelligence (e.g., Doebler & Scheffler, 2016; Frischkorn et al., 2019; Jensen, 2006; Sheppard & Vernon, 2008). Larson and Alderton (1990) were the first to describe the *Worst Performance Rule* (WPR). They divided their participants' intra-individual RT distributions into 16 bins and correlated the mean RTs of each bin with intelligence. Contrary to intuitive expectations, they observed that it was not the fastest but the slowest bins that showed stronger correlations with intelligence. The negative correlations between RTs and intelligence increased from r = -.20 in the first bin to r = -.37 in the last bin. As a result, they termed this phenomenon the WPR, highlighting that participants' worst performances, rather than their best, were most predictive of intelligence (Larson & Alderton, 1990). This finding has been replicated in several studies, showing consistent correlations between RTs and intelligence (e.g., Frischkorn et al., 2016; Kranzler, 1992; Larson & Alderton, 1990; Rammsayer & Troche, 2016) as well as with WMC (e.g., McVay & Kane, 2012; Schmiedek et al., 2007; Unsworth et al., 2010; Welhaf et al., 2020). Moreover, a recent meta-analysis by Schubert (2019) provided robust evidence supporting the reliability of the WPR.

Why are the slower parts of the RT distribution most predictive of intelligence? Larson and Alderton (1990) found a monotonous increase in RT variability across the 16 bins in their seminal paper. While one might argue that this increase in variability across the RT distribution is due to outliers or unsystematic errors, which could lower the reliability of slower RTs, this explanation seems implausible. As Coyle (2003) and Mashburn et al. (2023) pointed out, unsystematic noise would not be expected to result in higher correlations with intelligence. It is more likely that the WPR phenomenon occurs because differences in one or more cognitive processes become more pronounced in slower RTs, whereas they are less evident in faster RTs.

Based on the phenomenon of the WPR, the idea that individual differences in WMC reflect differences in executive attention (Engle et al., 1999; Engle & Kane, 2004; Kane et al., 2008), Unsworth et al. (2010) developed the attentional lapses account of the WPR. This theory suggests that differences in executive attention lead to variations in the frequency and intensity of attentional lapses. These attentional lapses, in turn, result in slower reactions, meaning that differences in executive attention primarily affect the variability of slower RTs. Since differences in executive attentional processes are more pronounced in slower RTs, these parts of the intraindividual RT distribution are also more predictive of differences in WMC and intelligence. Empirical findings from the 1960s support the idea of the attentional lapses account. Baumeister and Kellas (1968) found that cognitively impaired individuals exhibited greater variability in the slower RTs than individuals with normal cognitive abilities, while, in contrast, they observed no differences in the variability of faster RTs between the two groups. These findings may suggest that cognitively impaired participants experienced more frequent attentional lapses.

McVay and Kane (2012) empirically investigated the attentional lapses' account of the WPR by using mind-wandering episodes as direct indicators of attentional lapses during a reaction time task. Mind-wandering can be considered an individual's attentional shift from an ongoing task toward internal information, such as thoughts, plans, or memories (e.g., Smallwood & Schooler, 2006). McVay and Kane (2012) operationalized participants' experience of mind-wandering episodes using *task-unrelated thoughts* (TUTs), measured with *online thought probes*. These thought probes were administered at irregular intervals, asking participants what they had been thinking about just before the probe (Smallwood & Schooler, 2006). Possible answers could be, for example, *on task* or *not on task*, with the latter indicating an attentional lapse (= TUT), in line with the idea of the attentional lapses account of the WPR (McVay & Kane, 2012; Unsworth et al., 2010). McVay and Kane (2012) correlated participants' TUT rates with the parameters of the ex-Gaussian distribution, a mathematical description of the RT distributions with three freely estimated parameters that describe the individual reaction time distribution. The authors found that the individual TUT rate correlated most strongly with the τ parameter, which characterizes the slow end of the RT distribution. Additionally, individual differences in executive attention partially explained the negative relationship between τ and WMC (McVay & Kane, 2012). These findings indicate that executive attentional lapses are primarily included in the slowest RTs and partially explain the relationship between individual differences in (slower) information processing speed and higher-order cognitive abilities.

Following the attentional lapses account of the WPR, we aimed to empirically test whether inter-individual differences in executive attentional lapses could explain the monotonous increase in WPR correlations. We recruited a sample of $N = 85$ participants (29 male, 56 female), aged between 18 and 60 years ($M_{age} = 30.21$; $SD_{age} = 12.33$). Participants completed a shifting task adopted from Sudevan and Taylor (1987) with 640 trials in total. Depending on the color of the presented stimulus, participants had to decide in each trial whether the presented number between 1 and 9 was greater or smaller than five or whether the number was odd or even. Participants' intelligence was measured using the short version of the *Berlin Intelligence Structure Test* (Jäger et al., 1997).

Since attentional lapses are not a unidimensional construct limited to TUTs but rather multifaceted (Robison et al., 2020), we employed a multiverse approach to assess participants' tendencies for attentional lapses using several covariates derived from prior research on spontaneous mind wandering. We used online thought probes (Smallwood & Schooler, 2006) during the shifting task to capture current attentional lapses. To assess participants' general tendency to experience attentional lapses (trait mind-wandering), we used the Mind-Wandering Questionnaire by Mrazek et al. (2013) and the Spontaneous Mind-Wandering Scale by Carriere et al. (2013). Reaction time variability was measured using the Metronome Response Task (Seli et al., 2013), as individual variability in this task is related to the general tendency for spontaneous mind-wandering (Seli et al., 2013, 2014). Additionally, while participants completed the shifting task, their neural activity was recorded using the EEG. This allowed us to use electrophysiological correlates of mind-wandering as additional covariates for attentional lapses. Specifically, we used ERPs[4] such as P3 and P1 components (e.g., Baird et al., 2014; Kam et al., 2011; Kam & Handy, 2013; Smallwood et al., 2008) as well as pre-stimulus theta and post-stimulus alpha power (Arnau et al., 2020) to assess spontaneous mind-wandering. With these neurocognitive markers, it could be possible to gain deeper insights into the multiverse structure of attention errors that go beyond behavioral measurement scores.

To statistically evaluate both the significance of the increase of the correlations in the WPR and the influence of attentional lapses covariates on this phenomenon, we employed a multilevel modeling approach as recommended by Frischkorn et al. (2016). This method

[4] Further information about EEG and ERP measurements is provided in the description of Manuscript 3.

allowed us to address the limitations of prior approaches that used rank correlations to test the WPR's significance or Fisher's Z-test to compare correlations. In our study, we examined the WPR and the influence of covariates at both the covariance and correlation levels. While previous WPR analyses typically divided the reaction time distribution into bins or percentiles (e.g., Diascro & Brody, 1993; Fernandez et al., 2014; Frischkorn et al., 2016; Kranzler, 1992; Larson & Alderton, 1990; Rammsayer & Troche, 2016), we opted for a trial-by-trial analysis. This approach enabled us to capture the full progression of the WPR and assess the influence of covariates across the entire RT distribution.

In our sample of N = 85, we found a significant worst performance effect at the covariance level between RTs and intelligence. However, this significant effect did not appear on the level of correlations. The slope of the correlational trend from the fastest to the slowest trials showed an increase in the negative correlations of Δr = -.08. Nevertheless, the worst performance effect was only small (η^2part = .01). We concluded that the sample size of N = 85 was too small for this small worst performance effect to reach significance, as correlations typically show a high degree of estimation uncertainty, which usually stabilizes in sample sizes of N > 250 (Schönbrodt & Perugini, 2013).

At the covariance level, online thought probes (TUTs), the metronome response task, and pre-stimulus theta power showed an effect on the WPR. Interestingly, even after controlling for the single influence of each covariate, significant worst performance effects remained. However, when the combined influences of these three covariates on the WPR were examined simultaneously, the WPR effect disappeared at the covariance level, suggesting that different measures of attentional lapses independently contribute to the variance in RTs.

Even though we did not find a significant WPR at the correlation level, online thought probes (TUTs) and the metronome response task significantly influenced the course of the correlations. However, these results are difficult to interpret since no significant monotonous increase in negative correlations was found from the fastest to the slowest RTs.

The different covariates of attentional lapses showed little correlation with each other. Significant correlations were only observed between online thought probes with the questionnaire scale for spontaneous mind-wandering and the P1 amplitude with the metronome response task and with the P3 amplitude. This lack of convergent validity of the mind-wandering covariates suggested that it is challenging to measure attentional lapses as a unidimensional construct.

To address the limitations of the small sample size and the absence of the worst performance effect at the correlation level, we replicated our results in a second independent sample. For this purpose, we used an age-homogeneous dataset of college students from Kane et al. (2016) with a total of $N = 545$ participants (see also Welhaf et al., 2020). In this dataset, four RT tasks were included, during which the participants were repeatedly asked about current attentional lapses using online thought probes. Instead of intelligence data, the dataset contained participants' WMC, operationalized using various complex-span tasks (Kane et al., 2016). Importantly, substituting WMC for intelligence data is justified, as the two constructs are highly correlated and can serve as alternative measures of one another (e.g., Ackerman et al., 2002; Kyllonen & Christal, 1990; Oberauer et al., 2005). Furthermore, several studies found a worst-performance pattern in the correlations between RTs and WMC (McVay & Kane, 2012; Schmiedek et al., 2007; Welhaf et al., 2020), supporting the use of WMC as a proxy for intelligence.

In these analyses, we found a significant worst-performance effect in this independent dataset at both covariance and correlation levels. Controlling for attentional lapses significantly reduced this WPR effect. However, the influence of the covariates on the worst performance patterns was relatively small, and significant WPRs remained in each task. This suggests that while attentional lapses do contribute to the WPR, they do not fully account for the observed worst-performance patterns, indicating that other factors may also play a role in this phenomenon.

In summary, our findings suggested that cognitive control processes can partially account for the monotonic increase in the negative relationship between intelligence or WMC and information processing speed from participants' fastest to slowest responses, known as the WPR. However, the influence of the attentional lapse covariates on the worst performance effects was small and could not fully explain the WPR. Some covariates showed no effect on the WPR and revealed only small or absent correlations with the other attentional lapses covariates. This indicates that attentional lapses are not an unidimensional construct but have various facets (Robison et al., 2020), which is also reflected in their unique contributions to the WPR. Beyond the multidimensional nature of attentional lapses, the low correlations between the covariates could also indicate that different mind-wandering measures lacked sufficient validity. Future research should explore the validity of cognitive control measures in greater depth to clarify their role in intelligence differences.

The factor structure of executive functions and their relations to working memory and intelligence

In our first manuscript, we provided evidence that cognitive control processes, particularly attentional lapses, can partially explain the relationship between information processing speed and higher-order cognitive abilities. From these findings, we conclude that cognitive control processes play a role in accounting for individual differences in higher-order cognitive abilities. However, it also became clear that the valid measurement of cognitive control processes is challenging.

In our next study, we aimed to investigate cognitive control processes in greater detail, specifically focusing on their contribution to individual differences in WMC and intelligence. To achieve this, we examined the three EFs proposed by Miyake et al. (2000): inhibition, updating, and shifting. Previous research revealed evidence for the contribution of these EFs to individual differences in higher-order cognitive abilities (e.g., Friedman et al., 2006, 2008). Beyond empirical evidence, EFs also play a central role in theoretical models of intelligence. For instance, the POT by Kovacs and Conway (2016) posits that EFs are key processes that explain individual differences in intelligence. Kovacs and Conway (2016) suggested that EF processes are domain-general, meaning they are required for all cognitive tasks, regardless of the specific ability being measured. This domain-general nature makes EFs a bottleneck for performance across tasks, which in turn contributes to the positive correlations among different cognitive tasks, known as the *positive manifold*. The positive manifold refers to the observation that individuals' performance on one cognitive task tends to be positively correlated with their performance on other cognitive tasks, even when the tasks measure seemingly unrelated abilities. This pattern of positive correlations gives rise to the *g*-factor. The POT suggests that the domain-general nature of EFs can help explain the emergence of the positive manifold and the *g*-factor: since EFs are required for all cognitive tasks, individual differences in EF abilities can lead to consistent performance differences across a wide range of tasks. These consistent differences are reflected in the positive correlations observed in the positive manifold.

However, recent research has shown that measuring EFs is not as straightforward as it appears at first glance. For instance, scores obtained from EF tasks frequently revealed insufficient reliability and validity when used to assess interindividual differences (see von Bastian et al., 2020). Frischkorn et al. (2019) used a structural equation modeling approach to determine the overlap between information processing speed tasks and EF tasks (such as inhibition, updating, and shifting). They found that the variance specific to each of the EFs was quite small, while a large proportion of variance was shared with a domain-general factor of EFs, which in turn overlapped significantly with information processing speed tasks. This finding raises concerns about the validity of EF tasks for specifically measuring distinct EF processes.

In light of these findings, understanding how EF tasks measure these cognitive control processes is critical. For example, the Arrow-Flanker task (Eriksen & Eriksen, 1974) is commonly used to measure inhibition. In this task, participants must decide whether a central arrow points left or right, while flanking arrows on each side may either point in the same direction (congruent condition) or in the opposite direction (incongruent condition). The incongruent condition introduces interference, requiring participants to inhibit the distracting influence of the flanking arrows. While both conditions engage similar cognitive processes, the incongruent condition requires a greater degree of participants' inhibition abilities.

To isolate the specific contribution of inhibition, a difference score can be calculated by subtracting the mean RT of the congruent condition from that of the incongruent condition. Underlying this logic is the classic assumption of additive processes (Donders, 1869), suggesting that task performance is the sum of multiple contributing cognitive processes. The difference score is assumed to represent individual inhibition performance. Many classical EF tasks (measuring inhibition, updating, and shifting) are designed under the assumption of additive processes.

However, the assumption of additive processes may oversimplify the complex interplay of cognitive functions required for EF tasks as difference scores from EF tasks often lack sufficient reliability estimates. In their review, von Bastian et al. (2020) showed that the difference scores of 406 inhibition tasks had an average reliability of only .63. Schubert et al. (2022) discussed the reasons for this in detail, arguing that the EF-specific variance in difference scores is quite small. Additionally, difference scores incorporate error variance from both the congruent and incongruent conditions, resulting in a high proportion of error variance relative to true-score variance, which ultimately leads to low reliability estimates. In recent years, several researchers have advised caution when using difference scores to measure EF abilities (e.g., Ackerman & Hambrick, 2020; Draheim et al., 2019, 2023; Hedge et al., 2018; Mashburn et al., 2023; Miller & Ulrich, 2013; von Bastian et al., 2020; Weigard et al., 2021).

Another issue in measuring EFs is the use of different types of scores, which directly affects their relationships with other variables. Typical inhibition and shifting tasks use RT-based difference scores, while updating processes are generally measured using accuracy-based scores of average performance either across one condition or throughout the entire task (e.g., Frischkorn et al., 2019; Miyake et al., 2000; von Bastian et al., 2020). However, the methods used to measure executive abilities are inconsistent across studies. For instance, Wongupparaj et al. (2015) measured inhibition using only the incongruent condition of inhibition tasks (e.g., the Stroop task), whereas inhibition is more commonly measured using difference scores (von Bastian et al., 2020). These average performance scores are not process-pure, as they capture not only individuals' inhibition abilities but also domain-general processes such as information processing speed and perception speed.

Moreover, across studies, there is a lack of consistency in the use of RT and accuracy-based scores to measure specific EFs, with studies often using a mix of both types of scores (von Bastian et al., 2020). This is problematic because RT-based scores only consider correct responses, while accuracy-based scores reflect the ratio of correct to incorrect responses. The use of different score types within tasks and across studies can lead to measuring different processes and abilities, making cross-study comparisons challenging. Additionally, it is well established that RT- and accuracy-based scores within the same task tend to show only weak correlations (Hedge et al., 2018), indicating that they capture different underlying processes.

The psychometric challenges associated with difference scores—such as low reliability estimates—and the inconsistent use of RT and accuracy-based scores lead to varying interpretations and relationship patterns with other variables. To address these shortcomings, our goal in the following manuscripts was to measure EF abilities more consistently, using uniform measurement variables across tasks in each manuscript (Manuscript 2 = behavioral level, Manuscript 3 = electrophysiological level). This was achieved by collecting three typical tasks for each of the three EFs in a sample of $N = 148$ individuals. In addition to EF tasks, participants completed three ECTs, an intelligence test, and a battery of working memory tasks. Participants attended three sessions: The initial two sessions focused on the 12 decision-making tasks (EF tasks and ECTs), during which EEG data was recorded. In the third diagnostic session, participants completed the intelligence and working memory tests.

The factor structure of executive functions based on behavioral score (manuscript 2)

In our second manuscript, we addressed the two main issues in measuring executive functions—unreliable variables and the conflation of RT and accuracy-based scores—by applying a cognitive mathematical modeling approach combined with latent variance decomposition using structural equation models (SEMs).

First, to model the decision-making process in EF tasks, we used the Drift-Diffusion Model (Ratcliff, 1978), a mathematical model that describes underlying cognitive processes in binary decision tasks. The Drift-Diffusion Model allows us to quantify various cognitive processes involved in decision-making, such as participants' response biases, motor response processes, or speed-accuracy tradeoffs (see Voss et al., 2013, for detailed information). We specifically utilized the drift rate parameter (v) of the model to quantify individual executive abilities. Typically, this parameter reflects the speed at which a participant accumulates evidence to reach a decision (Ratcliff, 1978). However, while v is a useful metric in an EF task, it does not exclusively represent EF abilities. For instance, in the incongruent condition of the Arrow-Flanker task, v includes both domain-general information processing and domain-specific executive processes.

Given that v captures both domain-general and domain-specific processes, recent research has systematically explored the extent to which v reflects these different processes. Lerche et al. (2020) demonstrated that v reflects domain-general processes, such as general speed of information uptake, that contribute to decision-making in binary response tasks. At the same time, v also captures domain-specific processes, such as verbal, numerical, and figural abilities. Therefore, v should be understood as reflecting both domain-general processes, such as information processing speed, and domain-specific processes, such as EF abilities. Notably, a key advantage of the v parameter is that its estimation incorporates both intra-individual RTs for correct and incorrect responses, using the full informational content of all responses by integrating both RTs and accuracies.

Second, to avoid unreliable difference scores, we focused exclusively on conditions with higher processing demands in EF tasks, such as incongruent conditions in inhibition tasks, shifting conditions in shifting tasks, and updating conditions in updating tasks. However, since the individual v parameters of these conditions included both domain-general and EF-specific processes, we employed SEMs to separate these variances at the latent level by controlling the latent factors for domain-general information processing speed. The resulting latent EF-specific factors should be highly reliable because, based on the core idea of SEMs, latent factors capture only the covariance of the manifest variables, which is considered error-free and can, therefore, be regarded as true-score variance (e.g., Bowen & Guo, 2012).

In our first analysis step, we attempted to replicate the three-factor model by Miyake et al. (2000). However, based on RT difference scores and accuracy scores, we were not able to replicate the three-factor model of EFs and could not find a latent factor for inhibition. These results reinforced the need for an alternative approach using the *v* parameter and latent variance decomposition to measure EF abilities.

After we used individuals' *v* parameters as manifest variables for EF tasks, we found a latent factor for each of the three EFs, with high intercorrelations ranging from, $.75 \leq r \leq .89$. Based on this positive manifold, we modeled a hierarchical common factor above the first-order latent factors. Subsequently, only small, nonsignificant residual variances remained in the three first-level factors, with the common factor explaining almost all the variance in the EFs. Based on these results, we adopted a single-factor model, which described the data equally well (ΔAIC = 7.60). Importantly, since we did not use difference scores, this common factor contained EF-specific and domain-general variances. To separate these variances, we created a general processing factor containing the shared variance of the three ECTs, which included only low portions of EF-specific processing demands. Using a latent regression analysis, we controlled the common EF factor for general processes to obtain EF-specific variance as a residual in the hierarchical factor. However, the general processing factor fully explained the common factor, leaving no EF-specific residual variance. Additionally, the general processing factor was substantially correlated with intelligence (r = .46) and working memory (r = .46).

Based on these behavioral scores, our results showed that the shared variance of EF tasks included only a small portion of EF-specific processes. Instead, EF tasks primarily reflected two components: (1) individual differences in domain-general speed of evidence accumulation, which is correlated with higher-order cognitive abilities, and (2) task-specific variances, which are not correlated with one another. Notably, in addition to the covariance of information processing speed (speed of evidence accumulation), we also observed substantial portions of uncorrelated variance for each task, which could be interpreted as task-specific and are reflected by the error terms of the manifest variables.

After addressing the issues of inconsistent scores and reliability using cognitive mathematical model parameters and a latent modeling approach, it became evident that, based on behavioral scores, EF tasks revealed substantial validity problems in capturing interindividual differences in EF abilities.

The factor structure of executive functions based on electrophysiological scores (manuscript 3)

When we attempted to measure individual differences in EF abilities at the behavioral level using cognitive mathematical modeling and structure equation modeling approaches, we were not able to isolate variance specific to EF processes beyond general information processing speed in the tasks. In the third manuscript, we used electrophysiological process parameters from EEG to quantify EF-specific processes. As before, we avoided using difference scores and focused our analyses only on trials with higher EF processing demands from the nine EF tasks. Again, we used SEMs to separate the variance specific to EF processes at the latent level from domain-general processes. In contrast to the second manuscript, we estimated the SEMs using a Bayesian estimation procedure, as this approach generally provides more accurate parameter estimates in smaller sample sizes compared to models estimated using a frequentist approach (McNeish, 2016).

Typical cumulative measures, like RTs, are rough estimators of specific psychological processes, as they represent the sum of various processes. In contrast, EEG offers high temporal resolution, allowing the analysis of psychological processes with millisecond-level precision. In our third manuscript, we quantified individuals' performances in EF tasks using the mean amplitudes of ERP components. ERPs reflect the time-locked electrophysiological activity at specific scalp locations at particular moments following an event during a task (see Luck, 2014, for detailed information). ERP components are identified based on their deflection and sequential timing characteristics, and specific components are linked to certain cognitive processes. For example, the first positive deflection at occipital channels is known as the visual P1, which is associated with early visual processing (Luck, 2014). Generally, it is well-known that ERP components are sensitive to experimental task manipulations (Gaillard, 1988). Furthermore, ERP parameters are suitable for capturing interindividual differences in psychological processes (Sadus et al., 2023; Schubert et al., 2017; Schubert, Löffler, Hagemann, et al., 2023), although reliability issues have occasionally been noted (Cassidy et al., 2012; Nebe et al., 2023; Schubert, Löffler, Hagemann, et al., 2023).

To parameterize interindividual differences in EFs, we used participants' mean amplitudes of the fronto-central N2 and parieto-central P3 components, both of which are linked to cognitive control processes and EFs. The N2 reflects cognitive control under increased EF processing demands across various tasks (Folstein & Van Petten, 2008; Luck, 2014), with increased amplitudes consistently observed in incongruent trials of typical inhibition tasks (e.g., Bartholow et al., 2005; Heil et al., 2000; Yeung et al., 2004), such as the Arrow-Flanker task. Additionally, the N2 component reflects attentional allocation and working memory updating, with increased amplitudes in trials with higher shifting demands in shifting tasks (e.g., Gajewski et al., 2010) and decreased amplitudes in updating tasks, as seen in typical N-back tasks (Gevins et al., 1996; Salmi et al., 2019).

The P3 component is considered an indicator for updating contextual information (Donchin, 1981; Donchin & Coles, 1988). It reflects cognitive task demands and the allocation of attentional processes (Polich, 2007) and is closely linked to the updating of working memory (Luck, 1998, 2014; Polich, 2007; Polich & Kok, 1995; Vogel et al., 1998; Vogel & Luck, 2002). Furthermore, the P3 component is also associated with cognitive processes, such as stimulus classification (Duncan-Johnson, 1981; Luck, 2014) and response selection (Verleger, 2020). According to Polich (2007), the P3 amplitude can specifically be seen as a parameter associated with EFs and cognitive control. In shifting and inhibition tasks, the P3 component typically shows increased amplitudes in trials with higher EF processing demands (e.g., Gajewski & Falkenstein, 2011; Pratt et al., 2011). Previous findings also showed that in updating tasks, the P3 component exhibited decreased amplitudes in trials with increased updating demands (Dong et al., 2015; Scharinger et al., 2015; Watter et al., 2001). This body of research supported the selection of the N2 and P3 components as appropriate parameters for measuring individual differences in EF tasks and EF abilities.

Our results revealed three EF factors, each for both the N2 and P3 components, with moderate (r = .37) to high correlations (r = .80) within the components and moderate (r = -.40) to no correlations (r = -.03) between the components. Each set of factors exhibited a hierarchical structure, with a strong correlation between the common second-level factors (r = -.51). For most first-order latent factors, significant residual variances were observed, suggesting potential EF-specific processes. However, the P3 updating factor was fully explained by its common P3 factor, and no residual EF-specific variance remained. After controlling for general processes, represented by covariance across ECTs, the common N2 factor retained significant residual variance, while the common P3 factor was fully explained by domain-general characteristics of the P3 component.

To examine whether the observed residual variances were specific to EF processes or merely task-specific, we extended the model. Specifically, we introduced a hierarchical model with the same structure as before; however, this time, the model separately accounted for both conditions included in EF task: conditions with higher EF processing demands and those with lower EF processing demands. Additionally, we introduced higher-order factors over each of the associated residual factors (e.g., one hierarchical factor for the residuals of the N2 component in the shifting and non-shifting conditions). Our analysis revealed that all residual variances could be explained by their respective common factors, with no remaining variance that could be attributed to EF processes. Instead, the residual factors represented task-specific processes that were shared across conditions with both higher and lower EF processing demands. These findings suggest that the residual variances observed in the previous model did not reflect variances specific to EF processes introduced by the experimental manipulations in the conditions with higher EF processing demands.

Similar to the findings in Manuscript 2, this study was unable to identify variance in the EF tasks that was specifically attributable to executive processes at the level of electrophysiological parameters. While we observed significant experimental differences in the N2 and P3 components between the conditions of the EF tasks, these tasks were ultimately shown to be invalid for capturing interindividual differences in executive processes.

Discussion

Taken together, we demonstrated that cognitive control processes underlie interindividual differences in intelligence and can mediate the relationship between information processing speed and intelligence (Manuscript 1). However, it also became evident that substantial challenges are involved in validly measuring individual differences in cognitive control processes (Manuscripts 1, 2, and 3). Based on our findings, the question arises whether cognitive control processes are necessary predictors in examining processes underlying intelligence. If so, what approaches could be helpful to measure such cognitive control processes validly?

The complex interplay between cognitive control, information processing speed, and intelligence?

Our analyses showed that general information processing abilities explained a large portion of interindividual differences in both intelligence and WMC (Manuscripts 2 and 3). These findings are consistent with previous studies demonstrating robust relationships between intelligence and information processing speed (Doebler & Scheffler, 2016; Frischkorn et al., 2019; Neubauer & Knorr, 1997, 1998; Schmiedek et al., 2007; Schubert et al., 2015, 2017; Schubert, Löffler, Hagemann, et al., 2023; Sheppard & Vernon, 2008). Furthermore, we found that our battery of classical cognitive control tasks (EF tasks) showed insufficient validity measuring individual differences in cognitive control processes. These findings are in line with previous research that has also reported issues in measuring individual differences in cognitive control (Frischkorn et al., 2019; Hedge et al., 2018; Hull et al., 2008; Karr et al., 2018; Klauer et al., 2010; Krumm et al., 2009; Rey-Mermet et al., 2018, 2019; Rouder & Haaf, 2019; Stahl et al., 2014; von Bastian et al., 2020).

This raises the question: Is the strong predictive power of information processing speed sufficient for explaining individual differences in higher-order cognitive abilities? If this were the case, it would, in turn, render future studies on cognitive control processes as predictors of intelligence obsolete.

Although our study (Manuscript 1) attempted to bridge the gap between information processing speed and cognitive control, the precise contributions of cognitive control to individual differences in (slower) information processing speed and intelligence remain unclear. This ambiguity is largely due to the persistent challenges in measuring cognitive control processes (see also Manuscripts 2 and 3). Before concluding that cognitive control is irrelevant to differences in higher-order cognitive abilities, several open research questions need to be addressed. For example, traditional methods for measuring cognitive control—such as classical EF tasks and difference scores—remain challenging and often yield inconclusive results (cf. von Bastian et al., 2020). Therefore, new approaches in cognitive psychological research are necessary to measure interindividual differences in cognitive control more validly.

Development of new tasks that capture interindividual differences in cognitive control processes more validly

One potential approach to improving the measurement of cognitive control processes involves developing new tasks, modifying existing tasks, and using new measurement scores. In recent years, various advancements have emerged in cognitive psychological research. Draheim et al. (2021) developed a battery of new and modified tasks designed to measure cognitive control processes, particularly inhibition. In the toolbox, task parameters are dynamically adjusted through a staircase procedure based on participants' current performance (e.g., stimulus presentation time or maximum allowed response time), making the task either more complex or simpler. For instance, if an individual performs well, subsequent trials reduce the stimulus presentation duration or the allowed response time, thereby increasing difficulty. This adaptive procedure generates an individually calibrated value for stimulus presentation or allowed response time, which is then used as a dependent variable representing cognitive control performance. Empirically, the authors found a coherent cognitive control factor across tasks, which correlated with higher-order cognitive abilities independently of information processing speed (Burgoyne et al., 2023; Draheim et al., 2021). Additionally, the new tasks showed high reliability estimates (all estimates ≥ .86) and are quick and easy to administer (Burgoyne et al., 2023).

Another approach to improving the measurement of cognitive control processes was proposed by Kucina et al. (2022), who argued against the abandonment of difference scores in cognitive control research, as previously suggested by other researchers (e.g., Draheim et al., 2021; Hedge et al., 2018; Mashburn et al., 2023; Schubert et al., 2022). Instead, Kucina et al. (2022) introduced modified tasks to enhance the reliability of these scores. To achieve this, they implemented several improvements: First, they increased the salience and similarity of the stimuli's irrelevant attributes, making them more difficult to ignore. Additionally, the authors occasionally required a second response based on these irrelevant attributes. By incorporating a gamification approach, they increased task complexity, cognitive load, and participant engagement. Furthermore, they combined various tasks (e.g., Stroop, Flanker, and Simon tasks) to create more robust conflict effects through the combination (Kucina et al., 2022). These adjustments yielded acceptable to good reliability estimates for difference scores in fewer than 100 trials, making these tasks highly practical, as classical EF tasks typically require over 1000 trials to achieve acceptable reliability estimates for difference scores (Lee et al., 2023). Interestingly, Lee et al. (2023) also found that their difference scores showed moderate correlations with each other, indicating a certain degree of validity for classical EF tasks, despite the impractical number of trials required for their measurement.

The field of cognitive psychology is seeing promising new methods in the form of novel or modified tasks and scoring techniques that improve the measurement of individual differences in cognitive control processes. However, it is crucial to ensure that these new tasks validly capture cognitive control. To achieve this, they must meet two key validity criteria: (1) from an experimental psychological perspective, it must be demonstrated that the specific experimental manipulations within the tasks directly influence the dependent variables (Borsboom et al., 2004), and (2) the extracted scores, intended to reflect individual cognitive control abilities, must exhibit substantial correlations across tasks, thereby meeting the criterion of construct validity. Additionally, these scores should demonstrate divergent validity by showing only low correlations with other processes, such as information processing speed. Until now, the new cognitive control tasks by Draheim et al. (2021) and Kucina et al. (2022) have not been sufficiently examined to determine whether they meet both validity criteria. Future research should focus on validating these cognitive control tasks, which are both innovative and practical in their application.

Specific mathematical models for the valid parameterization of cognitive control processes

Beyond developing new tasks, the application and further refinement of specific mathematical models present a promising approach to measure cognitive control processes precisely. Models such as the dual-stage two-phase model by Hübner et al. (2010) or the conflict drift-diffusion model (White et al., 2011) have already demonstrated their effectiveness in describing individual inhibition performance and resolving conflicts caused by flanker stimuli using specific parameters in tasks like the Arrow-Flanker Task (see also Ulrich et al., 2015).

In one of our studies, we validated Hübner et al.'s (2010) dual-stage two-phase model using electrophysiological correlates specific to individual inhibition abilities (Schubert et al., 2022). We found substantial correlations between the model parameters and electrophysiological measures, which could be interpreted in a theoretically meaningful way. Furthermore, these inhibition parameters explained 37% of the variance in higher-order cognitive abilities (Schubert et al., 2022), demonstrating that it is possible to validly measure individual differences in cognitive control processes using specific mathematical models like the dual-stage two-phase model.

Further empirical evidence underscores the usefulness of advancing such models. In a recently published study, Robinson and Steyvers (2023) used a parameter substitution approach to investigate whether the parameters from the Arrow Flanker Task could predict performance in a Shifting Task and vice versa. Their results showed that cognitive control parameters from one task were not task-specific but domain-general, allowing them to predict performance across other tasks (Robinson & Steyvers, 2023). This approach significantly contributes to understanding cognitive control by offering a more unified and precise method for linking performance across tasks, highlighting that new developments in mathematical modeling can improve our understanding of the factorial structure of cognitive control processes.

Mathematical models have proven extremely useful in studying cognitive control processes. They offer the potential to accurately and validly capture individual performance. Future research should emphasize the development and validation of models to measure and investigate cognitive control processes through specific parameters.

More precise theoretical description and validation of cognitive control processes

Besides developing new tasks and mathematical models, addressing the theoretical issues and providing a more detailed description of the underlying cognitive control processes could also prove beneficial. In recent years, the definition of cognitive control processes, particularly EFs, has been criticized for its vagueness (Kucina et al., 2022; Verbruggen et al., 2014). Verbruggen et al. (2014) pointed out that executive control and inhibition are often described only in terms of group differences, leaving the underlying mechanisms unclear and turning process descriptions into a “black box.” A clearer understanding of these mechanisms could contribute to the development of more effective measurement methods and improve the accuracy of future studies.

Empirical investigations have also highlighted the need for a more differentiated description of the specific abilities involved in cognitive control. For example, Rey-Mermet et al. (2018) found that inhibition could be separated into two distinct factors across different tasks: one representing the inhibition of prepotent responses and the other reflecting the inhibition of distractor interference. However, Gärtner and Strobel (2021) could not replicate this two-factor solution, suggesting that further clarification is needed in this area.

Moreover, a recent meta-analysis by Unsworth et al. (2024) which reanalyzed data from 90 independent datasets comprising over 23,000 participants, identified a coherent cognitive control factor that was strongly correlated with other cognitive abilities such as intelligence, reading ability, WMC, information processing speed, and long-term memory. Further analyses revealed a hierarchical structure of cognitive control, with three factors of cognitive control at the first level: (1) inhibiting attentional resources (restraining attention), (2) focusing attentional resources on relevant information (constraining attention), and (3) maintaining attentional resources over time (sustaining attention). These findings suggest that cognitive control might be considered a cognitive ability composed of several sub-components. The findings of Unsworth et al. (2024) and Rey-Mermet et al. (2018) indicate that a more detailed description and definition of cognitive control processes would be beneficial.

In addition to better understanding the factorial structure of cognitive control, robust validation of these factors in cognitive psychology is essential for interpreting empirical findings more accurately. Early research on cognitive control revealed validation issues. Using latent variance decomposition, both Engle et al. (1999) and Conway et al. (2002) controlled WMC for short-term memory processes and information processing speed and still found substantial correlations between residual WMC and intelligence. Based on their theoretical framework, which posits that working memory comprises both short-term memory and cognitive control (Engle et al., 1999), they interpreted the residual variance of WMC as cognitive control processes. However, this interpretation is questionable, as working memory is a complex cognitive mechanism (e.g., Oberauer & Lewandowsky, 2019) involving more than two processes. It remains unclear how much of this residual variance in Engle et al. (1999) and Conway et al. (2002) reflects cognitive control processes and how much could be attributed to other processes, such as binding. The insufficient examination of the validity leads to uncertainty regarding the actual contribution of cognitive control processes to differences in WMC and their predictive power for intelligence.

Overall, both empirical and theoretical advancements would benefit from a more precise and detailed description of cognitive control processes. This would help resolve inconsistencies in previous research and foster the development of improved measurement methods.

Conclusion

Cognitive control processes could help to bridge the gap between WMC and information processing speed in explaining interindividual differences in intelligence, as suggested by Frischkorn et al. (2019). In our first Manuscript, we have shown that cognitive control processes (executive attention) contribute to differences in WMC and intelligence. Specifically, differences in cognitive control processes mediated the relationship between intelligence and information processing speed, as individual differences in these cognitive control processes partially explain the WPR. However, in each of the three manuscripts (especially in Manuscripts 2 and 3), we found challenges in validly measuring cognitive control processes or could not identify a coherent structure in these processes, limiting their explanatory contribution to the relationship between intelligence and information processing speed. Moving forward, using new approaches, such as mathematical models to parameterize individual cognitive control processes (e.g., Hübner et al., 2010; Robinson & Steyvers, 2023;

Ulrich et al., 2015; White et al., 2011), developing new cognitive control tasks (e.g., Draheim et al., 2021; Kucina et al., 2022), and refining theoretical definitions and descriptions of the underlying processes of cognitive control (e.g., Rey-Mermet et al., 2018; Unsworth et al., 2024), future research may provide stronger evidence on the contribution of cognitive control processes to individual differences in intelligence.

References

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

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, *80*, 101440. https://doi.org/10.1016/j.intell.2020.101440

Arnau, S., Löffler, C., Rummel, J., Hagemann, D., Wascher, E., & Schubert, A.-L. (2020). Inter-trial alpha power indicates mind wandering. *Psychophysiology*, *57*(6), e13581. https://doi.org/10.1111/psyp.13581

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baddeley, A. D. (1990). *Human memory: Theory and practice* (S. xi, 515). Allyn & Bacon.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Hrsg.), *Psychology of Learning and Motivation* (Bd. 8, S. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baird, B., Smallwood, J., Lutz, A., & Schooler, J. W. (2014). The Decoupled Mind: Mind-wandering Disrupts Cortical Phase-locking to Perceptual Events. *Journal of Cognitive Neuroscience*, *26*(11), 2596–2607. https://doi.org/10.1162/jocn_a_00656

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

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, *42*(1), 33–42. https://doi.org/10.1111/j.1469-8986.2005.00258.x

Baumeister, A. A., & Kellas, G. (1968). Reaction Time and Mental Retardation. In N. R. Ellis (Hrsg.), *International Review of Research in Mental Retardation* (Bd. 3, S. 163–193). Academic Press. https://doi.org/10.1016/S0074-7750(08)60011-7

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Bowen, N. K., & Guo, S. (2012). *Structural equation modeling*. Oxford University Press.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory≠attention. *Intelligence*, *33*(3), 251–272. https://doi.org/10.1016/j.intell.2005.01.002

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, *152*(8), 2369–2402. https://doi.org/10.1037/xge0001408

Carriere, J. S. A., Seli, P., & Smilek, D. (2013). Wandering in both mind and body: Individual differences in mind wandering and inattention predict fidgeting. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(1), 19–31. https://doi.org/10.1037/a0031438

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. 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*(5), 659–664. https://doi.org/10.1111/j.1469-8986.2011.01349.x

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. https://doi.org/10.1016/j.intell.2008.01.002

Conway, A. R. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331–335. https://doi.org/10.3758/BF03196169

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163–183. https://doi.org/10.1016/S0160-2896(01)00096-4

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

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Deary, I. (2008). Why do intelligent people live longer? *Nature*, *456*(7219), 175–176. https://doi.org/10.1038/456175a

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*(6), 573–580. https://doi.org/10.1016/j.intell.2008.12.002

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diascro, M. N., & Brody, N. (1993). Serial versus parallel processing in visual search tasks and IQ. *Personality and Individual Differences*, *14*(1), 243–245. https://doi.org/10.1016/0191-8869(93)90195-9

Doebler, P., & Scheffler, B. (2016). The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learning and Individual Differences*, *52*, 157–166. https://doi.org/10.1016/j.lindif.2015.02.009

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

Donders, F. C. (1869). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Dong, S., Reder, L. M., Yao, Y., Liu, Y., & Chen, F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Research*, *1616*, 146–156. https://doi.org/10.1016/j.brainres.2015.05.003

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508–535. https://doi.org/10.1037/bul0000192

Draheim, C., Tshukara, J. S., & Engle, R. W. (2023). Replication and extension of the toolbox approach to measuring attention control. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02140-2

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242–275. https://doi.org/10.1037/xge0000783

Duncan-Johnson, C. C. (1981). Young Psychophysiologist Award Address, 1980. *Psychophysiology*, *18*(3), 207–215. https://doi.org/10.1111/j.1469-8986.1981.tb03020.x

Engle, R. W. (2002). Working Memory Capacity as Executive Attention. *Current Directions in Psychological Science*, *11*(1), 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In *The psychology of learning and motivation: Advances in research and theory, Vol. 44* (S. 145–199). Elsevier Science.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fernandez, S., Fagot, D., Dirk, J., & de Ribaupierre, A. (2014). Generalization of the worst performance rule across the lifespan. *Intelligence*, *42*, 31–43. https://doi.org/10.1016/j.intell.2013.10.001

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*(5), 1410–1430. https://doi.org/10.1037/a0023750

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of experimental psychology. General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Frischkorn, G., Schubert, A.-L., & Hagemann, D. (2019). Processing speed, working memory, and executive functions: Independent or inter-related predictors of general intelligence. *Intelligence*, *75*, 95–110. https://doi.org/10.1016/j.intell.2019.05.003

Frischkorn, G., Schubert, A.-L., Neubauer, A., & Hagemann, D. (2016). The worst performance rule as moderation: New methods for worst performance analysis. *Journal of Intelligence*, *4*(3), 9. https://doi.org/10.3390/jintelligence4030009

Gaillard, A. W. K. (1988). Problems and paradigms in ERP research. *Biological Psychology*, *26*(1), 91–109. https://doi.org/10.1016/0301-0511(88)90015-4

Gajewski, P. D., & Falkenstein, M. (2011). Diversity of the P3 in the task-switching paradigm. *Brain Research*, *1411*, 87–97. https://doi.org/10.1016/j.brainres.2011.07.010

Gajewski, P. D., Kleinsorge, T., & Falkenstein, M. (2010). Electrophysiological correlates of residual switch costs. *Cortex*, *46*(9), 1138–1148. https://doi.org/10.1016/j.cortex.2009.07.014

Gärtner, A., & Strobel, A. (2021). Individual Differences in Inhibitory Control: A latent Variable Analysis. *Journal of Cognition*, *4*(1), 1. https://doi.org/10.5334/joc.150

Gevins, A., Smith, M. E., Le, J., Leong, H., Bennett, J., Martin, N., McEvoy, L., Du, R., & Whitfield, S. (1996). High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroencephalography and Clinical Neurophysiology*, *98*(4), 327–348. https://doi.org/10.1016/0013-4694(96)00288-X

Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence*, *47*, 122–133. https://doi.org/10.1016/j.intell.2014.09.004

Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, *21*(1), 115–124. https://doi.org/10.3758/BF03211170

Hagemann, D., Ihmels, M., Bast, N., Neubauer, A. B., Schankin, A., & Schubert, A.-L. (2023). Fluid Intelligence Is (Much) More than Working Memory Capacity: An Experimental Analysis. *Journal of Intelligence*, *11*(4), 4. https://doi.org/10.3390/jintelligence11040070

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-Task: Inhibitory Executive Processes? *Journal of Psychophysiology*, *14*(4), 218–225. https://doi.org/10.1027//0269-8803.14.4.218

Hick, W. E. (1952). On the Rate of Gain of Information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10.1080/17470215208416600

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*(3), 759–784. https://doi.org/10.1037/a0019471

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology*, *22*(4), 508–522. https://doi.org/10.1037/0894-4105.22.4.508

Jäger, A. O., Süß, H., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe. https://madoc.bib.uni-mannheim.de/14578/

Jensen, A. R. (2006). Chapter 11—Theory of the Correlation Between Response Time and Intelligence. In A. R. Jensen (Hrsg.), *Clocking the Mind* (S. 199–213). Elsevier Science Ltd. https://doi.org/10.1016/B978-008044939-5/50012-4

Kam, J. W. Y., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., & Handy, T. C. (2011). Slow fluctuations in attentional control of sensory cortex. *Journal of Cognitive Neuroscience*, *23*(2), 460–470. https://doi.org/10.1162/jocn.2010.21443

Kam, J. W. Y., & Handy, T. C. (2013). The neurocognitive consequences of the wandering mind: A mechanistic account of sensory-motor decoupling. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00725

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology. General*, *130*(2), 169–183. https://doi.org/10.1037//0096-3445.130.2.169

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Hrsg.), *Variation in Working Memory* (S. 21–48). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0002

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637–671. https://doi.org/10.3758/BF03196323

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70. https://doi.org/10.1037/0096-3445.132.1.47

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of experimental psychology. General*, *145*(8), 1017–1048. https://doi.org/10.1037/xge0000184

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology*, *63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Klauer, K. C., Stegmaier, R., & Meiser, T. (1997). Working Memory Involvement in Propositional and Spatial Reasoning: Thinking & Reasoning. *Thinking & Reasoning*, *3*(1), 9–47. https://doi.org/10.1080/135467897394419

Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kranzler, J. H. (1992). A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences*, *13*(3), 255–261. https://doi.org/10.1016/0191-8869(92)90099-B

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*(4), 347–364. https://doi.org/10.1016/j.intell.2009.02.003

Kucina, T., Wells, L., Lewis, I., Salas, K. de, Kohl, A., Palmer, M., Sauer, J. D., Matzke, D., Aidman, E., & Heathcote, A. (2022). *A Solution to The Reliability Paradox for Decision-Conflict Tasks*. PsyArXiv. https://doi.org/10.31234/osf.io/bc6nk

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

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

Lee, H. J., Smith, D. M., Hauenstein, C., Dworetsky, A., Kraus, B. T., Dorn, M., Nee, D., & Gratton, C. (2023). *Precise Individual Measures of Inhibitory Control*. OSF. https://doi.org/10.31234/osf.io/rj2bu

Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General*, *149*(12), 2207–2249. https://doi.org/10.1037/xge0000774

Lozano-Blasco, R., Quílez-Robres, A., Usán, P., Salavera, C., & Casanovas-López, R. (2022). Types of Intelligence and Academic Performance: A Systematic Review and Meta-Analysis. *Journal of Intelligence*, *10*(4), 123. https://doi.org/10.3390/jintelligence10040123

Luck, S. J. (1998). Neurophysiology of selective attention. In *Attention* (S. 257–295). Psychology Press/Erlbaum (UK) Taylor & Francis.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press.

Mashburn, C. A., Barnett, M. K., & Engle, R. W. (2023). Processing speed and executive attention as causes of intelligence. *Psychological Review*. https://doi.org/10.1037/rev0000439

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. https://doi.org/10.1080/10705511.2016.1186549

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*(3), 525–549. https://doi.org/10.1037/a0025896

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review*, *20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., & Schooler, J. W. (2013). Young and restless: Validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00560

Nebe, S., Reutter, M., Baker, D. H., Bölte, J., Domes, G., Gamer, M., Gärtner, A., Gießing, C., Gurr, C., Hilger, K., Jawinski, P., Kulke, L., Lischke, A., Markett, S., Meier, M., Merz, C. J., Popov, T., Puhlmann, L. M., Quintana, D. S., … Feld, G. B. (2023).

Enhancing precision in human neuroscience. *eLife*, *12*, e85980. https://doi.org/10.7554/eLife.85980

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. E., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*.

Neubauer, A. C., & Knorr, E. (1997). Elementary cognitive processes in choice reaction time tasks and their correlations with intelligence. *Personality and Individual Differences*, *23*(5), 715–728. https://doi.org/10.1016/S0191-8869(97)00108-6

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. https://doi.org/10.1016/S0160-2896(99)80058-0

Oberauer, K., & Lewandowsky, S. (2019). Simple measurement models for complex working-memory tasks. *Psychological Review*, *126*(6), 880–932. https://doi.org/10.1037/rev0000159

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.-M. (2005). Working Memory and Intelligence--Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 61–65. https://doi.org/10.1037/0033-2909.131.1.61

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2008). Individual Differences in Working Memory Capacity and Reasoning Ability. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Hrsg.), *Variation in Working Memory* (S. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0003

Oberauer, K., Süß, H.-M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*(2), 167–193. https://doi.org/10.1016/S0160-2896(02)00115-0

Pesta, B. J., McDaniel, M. A., & Bertsch, S. (2010). Toward an index of well-being for the fifty U.S. states. *Intelligence*, *38*(1), 160–168. https://doi.org/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*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, *41*(2), 103–146. https://doi.org/10.1016/0301-0511(95)05130-9

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, *74*(5), 392–409. https://doi.org/10.1037/h0024913

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of Working Memory Load on Visual Selective Attention: Behavioral and Electrophysiological Evidence. *Frontiers in Human Neuroscience*, *5*. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00057

Rammsayer, T., & Troche, S. (2016). Validity of the worst performance rule as a function of task complexity and psychometric g: On the crucial role of g saturation. *Journal of Intelligence*, *4*(1), 5. https://doi.org/10.3390/jintelligence4010005

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

Raven, J., & Raven, J. (2003). Raven Progressive Matrices. In *Handbook of nonverbal assessment* (S. 223–237). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-1-4615-0153-4_11

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: Selective effects of updating, maintenance, and inhibition.

Journal of Experimental Psychology. Learning, Memory, and Cognition, *37*(2), 308–324. https://doi.org/10.1037/a0022216

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501–526. https://doi.org/10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *148*(8), 1335–1372. https://doi.org/10.1037/xge0000593

Robinson, M. M., & Steyvers, M. (2023). Linking computational models of two core tasks of cognitive control. *Psychological Review*, *130*(1), 71–101. https://doi.org/10.1037/rev0000395

Robison, M. K., Miller, A. L., & Unsworth, N. (2020). A multi-faceted approach to understanding individual differences in mind-wandering. *Cognition*, *198*, 104078. https://doi.org/10.1016/j.cognition.2019.104078

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Sadus, K., Schubert, A.-L., Löffler, C., & Hagemann, D. (2023). An explorative multiverse study for extracting differences in P3 latencies between young and old adults. *Psychophysiology*, *n/a*(n/a), e14459. https://doi.org/10.1111/psyp.14459

Salmi, J., Vilà-Balló, A., Soveri, A., Rostan, C., Rodríguez-Fornells, A., Lehtonen, M., & Laine, M. (2019). Working memory updating training modulates a cascade of event-related potentials depending on task load. *Neurobiology of Learning and Memory*, *166*, 107085. https://doi.org/10.1016/j.nlm.2019.107085

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. *Psychophysiology*, *52*(10), 1293–1304. https://doi.org/10.1111/psyp.12500

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*(1), 162–173. https://doi.org/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*(3), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

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

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

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

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

Schubert, A.-L., Löffler, C., & Hagemann, D. (2022). A neurocognitive psychometrics account of individual differences in attentional control. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001184

Schubert, A.-L., Löffler, C., Hagemann, D., & Sadus, K. (2023). How robust is the relationship between neural processing speed and cognitive abilities? *Psychophysiology*, *60*(2), e14165. https://doi.org/10.1111/psyp.14165

Schubert, A.-L., Löffler, C., Sadus, K., Göttmann, J., Hein, J., Schröer, P., Teuber, A., & Hagemann, D. (2023). Working memory load affects intelligence test performance by reducing the strength of relational item bindings and impairing the filtering of irrelevant information. *Cognition*, *236*, 105438. https://doi.org/10.1016/j.cognition.2023.105438

Seli, P., Carriere, J. S. A., Thomson, D. R., Cheyne, J. A., Martens, K. A. E., & Smilek, D. (2014). Restless mind, restless body. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 660–668. https://doi.org/10.1037/a0035260

Seli, P., Cheyne, J. A., & Smilek, D. (2013). Wandering minds and wavering rhythms: Linking mind wandering and behavioral variability. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 1–5. https://doi.org/10.1037/a0030954

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

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799. https://doi.org/10.1177/1745691616650647

Smallwood, J., Beach, E., Schooler, J. W., & Handy, T. C. (2008). Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. *Journal of Cognitive Neuroscience*, *20*(3), 458–469. https://doi.org/10.1162/jocn.2008.20037

Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*(6), 946–958. https://doi.org/10.1037/0033-2909.132.6.946

Spearman, C. (1904). „General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201–292. https://doi.org/10.2307/1412107

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*(2), 850–886. https://doi.org/10.1037/a0033981

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–457.

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 89–103. https://doi.org/10.1037/0096-1523.13.1.89

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174. https://doi.org/10.1016/j.cogpsych.2015.02.005

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26. https://doi.org/10.1016/j.cogpsych.2014.01.003

Unsworth, N., Miller, A. L., & Strayer, D. L. (2024). Individual differences in attention control: A meta-analysis and re-analysis of latent variable studies. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-024-02516-1

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*(1), 111–122. https://doi.org/10.1016/j.intell.2009.08.002

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Verbruggen, F., McLaren, I. P. L., & Chambers, C. D. (2014). Banishing the Control Homunculi in Studies of Action Control and Behavior Change. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(5), 497–524. https://doi.org/10.1177/1745691614526414

Verleger, R. (2020). Effects of relevance and response frequency on P3b amplitudes: Review of findings and comparison of hypotheses about the process reflected by P3b. *Psychophysiology*, *57*(7), e13542. https://doi.org/10.1111/psyp.13542

Vernon, P. E. (1950). *The structure of human abilities* (S. 160). Methuen.

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743. https://doi.org/10.3758/BF03196329

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. https://doi.org/10.1037/0096-1523.24.6.1656

von Bastian, C. C., Blais, C., Brewer, G., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., … Wiemers, E. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

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

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003. https://doi.org/10.1111/1469-8986.3860998

Weigard, A., Clark, D. A., & Sripada, C. (2021). Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition*, *215*, 104818. https://doi.org/10.1016/j.cognition.2021.104818

Welhaf, M. S., Smeekens, B. A., Meier, M. E., Silvia, P. J., Kwapil, T. R., & Kane, M. J. (2020). The worst performance rule, or the not-best performance rule? Latent-variable analyses of working memory capacity, mind-wandering propensity, and reaction time. *Journal of Intelligence*, *8*(2). https://doi.org/10.3390/jintelligence8020025

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, *63*(4), 210–238. https://doi.org/10.1016/j.cogpsych.2011.08.001

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

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

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

List of abbreviations

AIC	Akaike information criterion
ECT	Elementary cognitive task
EEG	Electroencephalogram
EF	Executive function
ERP	Event-related potential
g	General intelligence
POT	Process-overlap theory
RT	Reaction time
TUT	Task-unrelated thoughts
SEM	Structure equation model
WMC	Working memory capacity

Appendix Manuscript I

Journal of *Intelligence*

MDPI

Article

Do Attentional Lapses Account for the Worst Performance Rule?

Christoph Löffler [1,2], Gidon T. Frischkorn [3], Jan Rummel [1], Dirk Hagemann [1] and Anna-Lena Schubert [2,*]

1 Institute of Psychology, Heidelberg University, Hauptstr. 47-51, 66917 Heidelberg, Germany; christoph.loeffler@psychologie.uni-heidelberg.de (C.L.); jan.rummel@psychologie.uni-heidelberg.de (J.R.); dirk.hagemann@psychologie.uni-heidelberg.de (D.H.)
2 Department of Psychology, University of Mainz, Wallstr. 3, 55122 Mainz, Germany
3 Department of Psychology, University of Zurich, Binzmühlestr. 14, 8050 Zurich, Switzerland; gidon.frischkorn@psychologie.uzh.ch
* Correspondence: anna-lena.schubert@uni-mainz.de

Abstract: The worst performance rule (WPR) describes the phenomenon that individuals' slowest responses in a task are often more predictive of their intelligence than their fastest or average responses. To explain this phenomenon, it was previously suggested that occasional lapses of attention during task completion might be associated with particularly slow reaction times. Because less intelligent individuals should experience lapses of attention more frequently, reaction time distribution should be more heavily skewed for them than for more intelligent people. Consequently, the correlation between intelligence and reaction times should increase from the lowest to the highest quantile of the response time distribution. This attentional lapses account has some intuitive appeal, but has not yet been tested empirically. Using a hierarchical modeling approach, we investigated whether the WPR pattern would disappear when including different behavioral, self-report, and neural measurements of attentional lapses as predictors. In a sample of N = 85, we found that attentional lapses accounted for the WPR, but effect sizes of single covariates were mostly small to very small. We replicated these results in a reanalysis of a much larger previously published data set. Our findings render empirical support to the attentional lapses account of the WPR.

Keywords: worst performance rule; attentional lapses; attentional lapses account; intelligence; multilevel analysis; task-unrelated thoughts

Citation: Löffler, Christoph, Gidon T. Frischkorn, Jan Rummel, Dirk Hagemann, and Anna-Lena Schubert. 2022. Do Attentional Lapses Account for the Worst Performance Rule? *Journal of Intelligence* 10: 2. https://doi.org/10.3390/jintelligence10010002

Received: 8 September 2021
Accepted: 21 December 2021
Published: 24 December 2021

1. Introduction

Reaction times (RTs) in elementary cognitive tasks typically correlate moderately with general intelligence (Doebler and Scheffler 2016; Sheppard and Vernon 2008). Moreover, if intra-individual RT distributions are divided into bins from the fastest to the slowest RTs, the negative relations between mean RT within each bin and intelligence increase from the fastest to the slowest parts of the distribution (Baumeister and Kellas 1968; Coyle 2003; Larson and Alderton 1990; Schubert 2019). Larson and Alderton (1990) named this phenomenon the worst performance rule (WPR). The WPR suggests that inter-individual differences in slower RTs explain more of the variance in individuals' cognitive abilities than faster RTs (see Figure 1 for an illustration of the typical WPR pattern). As pointed out by Larson and Alderton (1990), a better understanding of this phenomenon is desirable as it may be informative of the cognitive mechanisms underlying inter-individual differences in intelligence.

The WPR has been observed in several studies (Diascro and Brody 1993; Fernandez et al. 2014; Frischkorn et al. 2016; Kranzler 1992; Leite 2009; McVay and Kane 2012; Rammsayer and Troche 2016; Schmiedek et al. 2007; Schmitz et al. 2018; Schmitz and Wilhelm 2016; Unsworth et al. 2010), although there are a few studies that did not find evidence for a WPR (Dutilh et al. 2017; Ratcliff et al. 2010; Salthouse 1993, 1998; Saville et al. 2016). A recent meta-analysis addressed the question of the strength, consistency, and generalizability of WPR across 23 datasets (from 19 different studies and 3767 participants) and found evidence in favor of the WPR (Schubert 2019).

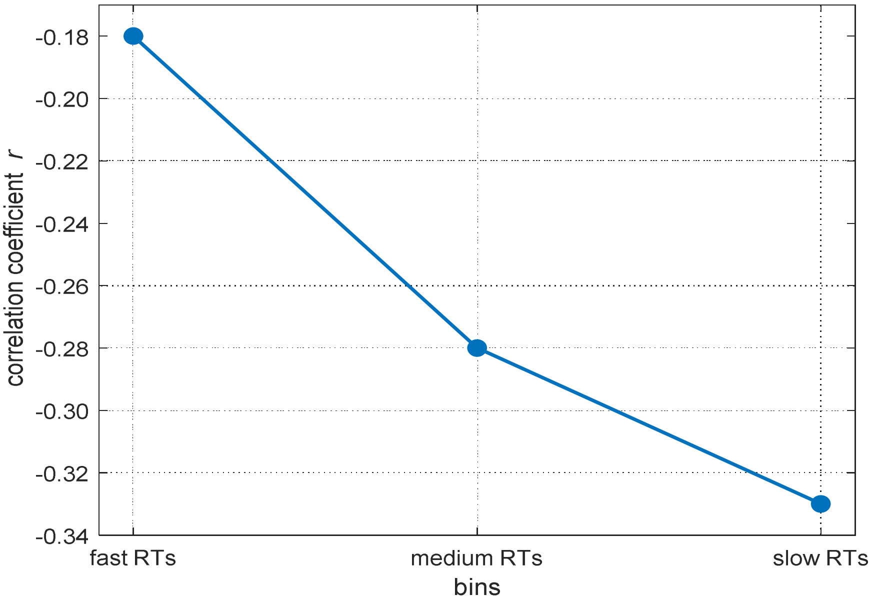


Figure 1. An example for the increasing magnitude in correlations between RT and mental abilities from fast to slow RT-bins. Data are based on the meta-analysis from Schubert (2019).

Identifying the underlying mechanisms of the WPR may help to identify the elementary processes producing inter-individual differences in intelligence, because whichever process is measured particularly with the slowest RTs may also contribute to differences in mental abilities. Different candidate accounts for explaining the occurrence of the WPR have been proposed. Several authors suggested an attentional lapses account of the WPR which states that the WPR occurs due to lapses of attention to which less intelligent people are particularly prone (Jensen 1992; Larson and Alderton 1990; Unsworth et al. 2010). On a neural level, this could be reflected by less intelligent individuals showing a higher frequency of neural transmission errors (Coyle 2001; Miller 1994) or spending more processing time on neural subthreshold and refractory periods, resulting in errors or delays during information processing (Jensen 1992). As the attentional lapses account is currently the most prominent account for explaining the WPR, we put this account at critical test in the present study.

1.1. The Attentional Lapses Account of the WPR and Its Examination

According to the executive attention theory of working memory (Kane et al. 2008), individual differences in executive attention predict differences in working memory capacity (WMC) and higher cognitive abilities such as fluid intelligence. While performing any type of (demanding) cognitive tasks, external distractors (such as a loud noise) and internal distractors (such as thoughts about the last or next vacation) may interfere with task completion by impairing task processing and goal maintenance. Accordingly, individuals who are able to shield their current thoughts against such task-irrelevant external or internal distractors should show better task performance. Kane et al. (2008) suggested that certain people are better at blocking out task-irrelevant information and maintaining current task goals than others, in particular those people with high executive attention (see also Kane et al. 2004). Individuals with lower executive attentional control, however, should perform worse in cognitive ability tests, because they are not able to keep their attention adequately focused on a task.

The consequence of such executive attention failures is that people who are not able to focus their attention on the task at hand experience attentional lapses while performing a task. Empirically, this will result in slower correct responses or in fast response failures (Unsworth et al. 2010). From an individual differences perspective, one would therefore expect that individuals with a higher propensity for attentional lapses occasionally show very slow but correct responses or a higher error rate. In fact, previous research has shown that self-reported attentional lapses were moderately associated with individual differences in the right tail of the RT distribution, that is, individuals who reported higher rates of attentional lapses showed more positively skewed RT distributions and hence more frequent slow responses (McVay and Kane 2012). In addition, self-reported attentional lapses predicted error rates in simple experimental tasks such as the sustained attention to response task (McVay and Kane 2009; Smallwood and Schooler 2006).

If individual differences in attentional lapses are related to differences in cognitive abilities such as fluid intelligence and if attentional lapses lead to slow responses, it is in consequence not surprising that slower responses are more strongly related to intelligence than fast responses (i.e., the phenomenon of the WPR). In contrast to faster responses, slower RTs reflect attentional lapses as an additional process, which results in the typical pattern of the WPR. Additional analyses by McVay and Kane (2012), in which they demonstrated that individual differences in self-reported attentional lapses partly mediated the association between slowest RTs and WMC, provided first evidence supporting this hypothesis.

1.2. Multiverse Manifestation and Measurement of Attentional Lapses

Attentional lapses are a multi-faceted construct (Robison et al. 2020) and that is the reason why the measurement of attentional lapses is not straightforward. There are different possibilities to operationalize participants' attentional states (McVay and Kane 2012; Unsworth et al. 2010). Most of the measurements—which we used in this study—were adapted from mind wandering research and possess face validity to the construct of attentional lapses. Possible manifestations of attentional lapses can be found in participants' self-reported attentional states, their response behavior, or psychophysiological measures.

Many studies measured attentional lapses as participants' self-reported mental states (Smallwood and Schooler 2015). During an ongoing task, participants are typically asked whether their thoughts are on- or off-task. In consequence, if their thoughts are not on the ongoing task, they are experiencing task-unrelated-thoughts (TUTs; Smallwood and Schooler 2006), which are considered as attentional drifts or attentional lapses (McVay and Kane 2010; Watkins 2008). Individuals tend to show a larger variability in those RTs in which they report TUTs, but they do not consistently show shifts of mean RTs (Leszczynski et al. 2017; McVay and Kane 2009, 2012; Seli et al. 2013, 2014; Thomson et al. 2014). These results suggest that attentional lapses may lead to an increase in the variability of RTs due to occasional failures in an experimental task.

Another method to measure the subjective frequency of attentional lapses are questionnaires that measure participants' attentional states during everyday life experiences and their personal tendencies for attentional lapses in everyday situations. Individuals who reported a higher tendency for attentional lapses also tended to report a higher frequency of TUTs during experimental tasks (Mrazek et al. 2013; Schubert et al. 2020). This suggests that both measurements assess—at least to some degree—the same underlying construct.

As a more objective alternative, it has been proposed to assess attentional states with behavioral measures such as the metronome response task (MRT; Seli et al. 2013). This task measures attentional lapses based on intraindividual variability in participants' tapping response to a continuous metronome beat. It has been suggested that individuals' tapping variance may reflect their attentional states (Seli et al. 2013). Seli et al. (2013, 2014) showed that self-reported attentional lapses are related to increases in tapping variability on the metronome beat in this task.

Beyond behavioral and self-report measures, former research identified several electrophysiological correlates of attentional lapses. The P3 is a component of the event-related potential (ERP) that occurs about 300 ms after stimulus onset at parietal electrodes and is associated with a wide range of higher-order cognitive processes such as stimulus evaluation and memory updating (Polich 2007; Verleger 2020). It has been repeatedly associated with self-reported attentional lapses: Several studies found reduced P3 amplitudes during trials in which participants reported not having been focused on the task (Kam and Handy 2013; Smallwood et al. 2008). The same pattern of results was reported by Barron et al. (2011), who found a reduced P3 amplitude in participants who experienced more attentional lapses in comparison to more focused participants. In addition, attentional lapses have been shown to affect sensory processing, as smaller visual P1 amplitudes have been observed for trials in which participants reported attentional lapses (Baird et al. 2014; Kam et al. 2011; see also Kam and Handy 2013). The P1 is a component of the event-related potential that occurs about 100 ms after stimulus onset at occipital electrodes. These findings suggest that attentional lapses affect the neurocognitive processing of information and that they are accompanied by a reduction of amplitudes of ERP components associated with stimulus perception and evaluation.

Furthermore, several studies reported that attentional lapses were associated with changes in the time-frequency domain, in particular with increases in inter-stimulus alpha power and increases in stimulus-locked theta power. Alpha activity is known to reflect an internally oriented mental state (Hanslmayr et al. 2011) and has, for example, been shown to increase during episodes of mental imaging (Cooper et al. 2003) and to be suppressed during sensory stimulation (Berger 1929; Thut et al. 2006). Episodes during which attention is not fully oriented towards the actual task have therefore been associated with greater alpha power (Baldwin et al. 2017; Compton et al. 2019; O'Connell et al. 2009). Arnau et al. (2020) further disentangled the time-course of this association and found alpha power to increase overall, but particularly at lateral parietal and occipital electrodes during the inter-trial-interval before participants reported TUTs. This internal focus of attention was redirected to the primary experimental task once an imperative stimulus (e.g., the fixation cross) was presented.

Theta power, especially event-related frontal-midline theta power, is associated with executive control and regulation processes (Cavanagh et al. 2012; Cavanagh and Frank 2014). Previous research has suggested that theta power may decrease when attentional lapses occur and may be subsequently upregulated as a compensatory mechanism once attentional drifts are noticed (Arnau et al. 2020; Atchley et al. 2017; Braboszcz and Delorme 2011). This redirection of attention towards the primary task may be initiated by either meta-awareness regarding one's attentional state (Braboszcz and Delorme 2011; Smallwood et al. 2007) or by external cues such as the presentation of the fixation cross or the next experimental trial (Arnau et al. 2020).

To achieve a most comprehensive analysis in the present study, we combined these heterogeneous approaches and applied a multiverse strategy for capturing participants' attentional states with different operationalizations in a multimethod approach. Therefore, we combined the listed self-report measurements with the listed behavioral and psychophysiological measures.

1.3. Identifying Occurrences of the WPR

In the present study, we analyzed the WPR before and after controlling for individual differences in attentional lapses by applying a recently proposed multilevel approach to the WPR (Frischkorn et al. 2016). Most WPR studies reported only the correlations of the mean or median RTs in the performance bands with intelligence, which is merely a description of the WPR rather than an inferential statistical examination of the phenomenon. If studies tested increasing correlations over RT bands for significance, they used rank-correlations (e.g., Kranzler 1992; Larson and Alderton 1990) or comparisons of correlation coefficients from dependent samples with Fisher's Z-values (e.g., Rammsayer and Troche 2016). Both statistical methods have certain weaknesses.

One weakness of rank-correlations is that they only quantify the extent of monotonicity in the increase of negative covariances or correlations between RTs and intelligence over the different bins. If this increase is quite monotonic, a rank-correlation close to one will be found no matter how large the increase is. By using the rank-correlation as a method to test the WPR, it is not possible to quantify the slope of the increase of correlations over bins of the RT distribution, which is needed to quantify the size of the WPR. The second weakness of rank-correlations is that they ignore the estimation uncertainty of correlations if these correlations are first estimated and then subsequently entered as observed variables into rank-correlations. This sequential approach results in a possible overestimation of the significance of the WPR (Skrondal and Laake 2001). Moreover, tests assessing the difference between dependent correlations suffer from low statistical power, possibly underestimating the WPR. For these reasons, we used the recently proposed multilevel account to test the WPR more adequately, i.e., in a single estimation step and with higher statistical power (Frischkorn et al. 2016).

There are two possible ways to measure the worst performance pattern by using either unstandardized (covariances) or standardized (correlations) coefficients in the multi-level models. On the one hand, covariances reflect the unstandardized relation between two variables, which means that an increase in magnitude of covariances can have two reasons: They can either reflect an actual increase of the relation between both variables or they can reflect increases in inter-individual variances in at least one of the two variables. On the other hand, increasing correlations represent increases in the relationships between two variables, because correlations are controlled for inter-individual variances. To understand attentional lapses' influences on the RT variances and their effects on the relation between RT and intelligence, we used both unstandardized and standardized coefficients in the present analyses. In order to obtain a higher resolution of the course of the WPR and the influence of attentional lapses on the WPR, we analyzed the RT distribution on trial-by-trial basis with multilevel models and did not apply a binning procedure as, e.g., Frischkorn et al. (2016) did.

The aim of the present study was to assess if individual differences in the frequency of attentional lapses could account for the WPR. Due to the multiverse nature of attentional lapses, we used behavioral, self-report, and electrophysiological methods to assess individual differences in the frequency of attentional lapses. In addition, we used the previously proposed multilevel account of the WPR (Frischkorn et al. 2016) to quantify and test any moderating effect of attentional lapses on the strength of the worst performance effect. Based on the attentional lapses account, we assumed that individual differences in attentional lapses explain—at least partially—the emergence of the WPR. Hence, we expected the slope of the WPR to be significantly reduced if we controlled RTs for individual differences in attentional lapses.

2. Study 1

2.1. Materials and Methods

The study was approved by the ethics committee of the faculty of behavioral and cultural studies of Heidelberg University. At the beginning of an experimental session, participants signed an informed consent.

2.1.1. Participants

We recruited a sample of $N = 100$ general population participants via the local newspaper, via our own university homepage, via a pool of potentially interested participants in psychological studies, and by distributing flyers in Heidelberg. All volunteers were admitted if they were between 18 and 60 years old and had no history of mental illnesses. Two participants were removed because they did not complete the experiment. In consequence of the outlier analysis (see below), 13 more participants were removed from further analyses. The remaining sample ($N = 85$) consists of 29 males and 56 females. Participants' mean age was 30.21 years ($SD = 12.33$). All participants either stated that German was their mother tongue or that they spoke German on a level comparable to native speakers. The educational degrees were distributed in the following way: As highest educational level, 49 participants had a high school diploma (German Abitur), 30 had a university degree, and six had an educational degree lower than a high school diploma. All participants had normal or corrected to normal vision. They received 30 € and personal feedback as compensation for their participation.

2.1.2. Materials

Berlin Intelligence Structure Test (BIS)

To measure participants' intelligence, we used the short version of the Berlin Intelligence Structure Test (BIS-4, booklet 2: Jäger et al. 1997). The short version of the BIS is a particularly suitable instrument for measuring cognitive abilities in a relatively short time (about 50–60 min). Moreover, the short version of the BIS is a heterogeneous test battery for different abilities and includes 15 different tasks. Based on the theory by Jäger (1984), the test consists of four operation-related (processing speed, memory, creativity, processing capacity) and three context-related (verbal, numerical, figural) components of intelligence. Furthermore, the test allows the calculation of a general intelligence (g) score. We used the sum scores across all subtests as an independent variable.

Five participants had already completed the same test within the last year at our department. Because there may be a training effect between the two measurement occasions within one year (Scharfen et al. 2018), we used their BIS-scores from the previous study for all further analyses. The mean test score of the whole sample (N = 85) was 1498.29 (SD = 80.02) which corresponds to a converted mean IQ score of 94.58 (SD = 16.12). Cronbach's α showed a good internal consistency for the test scores (α = .79).

Choice RT Task: Switching Task

We measured RTs in a switching task, which was based on a task used by Sudevan and Taylor (1987). An unpublished reanalysis of a previous study in which we used this task (Frischkorn et al. 2019) suggested that it yields a significant worst performance effect.

While participants were working on this task, they had to decide whether a presented digit was smaller or larger than five or whether it was an odd or an even number. This task is constructed based on a 2 $\times$ 2 design and consists of four different experimental conditions. Which rule currently applied depended on the color in which the stimuli were presented (red = less/more condition, green = odd/even condition). The digit of a single trial could be either presented in the color of the former trial (=repeat condition) or in the other color (=shifting condition). The stimulus set included the digits between one and nine, excluding five.

The task was programmed in MATLAB (The MathWorks Inc., Natick, MA, USA) with the open source software package Psychtoolbox version 3.0.13 (Kleiner et al. 2007). We implemented restrictions that the same digits could never appear twice in a row as well as the same color could never appear more than three times in a row. Participants were instructed to answer as correctly and as fast as possible. On the keyboard, they had to press "L" to indicate that a digit was either larger than five or even and they had to press "D" to indicate that a digit was either smaller than five or odd.

All stimuli were presented in the middle of the screen on a black background (Figure 2). At the beginning of each trial, a gray fixation cross was shown for 512–768 ms. Following the fixation cross, a blank screen was presented as inter stimulus interval for 1024–1278 ms. Subsequently the digit followed and disappeared 1024–1278 ms after the participants responded. The stimulus disappeared after three seconds if the participants did not respond. At the end of each trial a blank screen was presented again as an inter-trial interval of 1000–1500 ms.

Participants completed 40 practice trials (ten trials task pure less/more, ten trials task pure odd/even, and 20 trials including task shifting) during which they received feedback. After that, they worked on the experimental trials, which consisted of ten blocks with 64 trials each. Participants took self-paced breaks between the blocks.

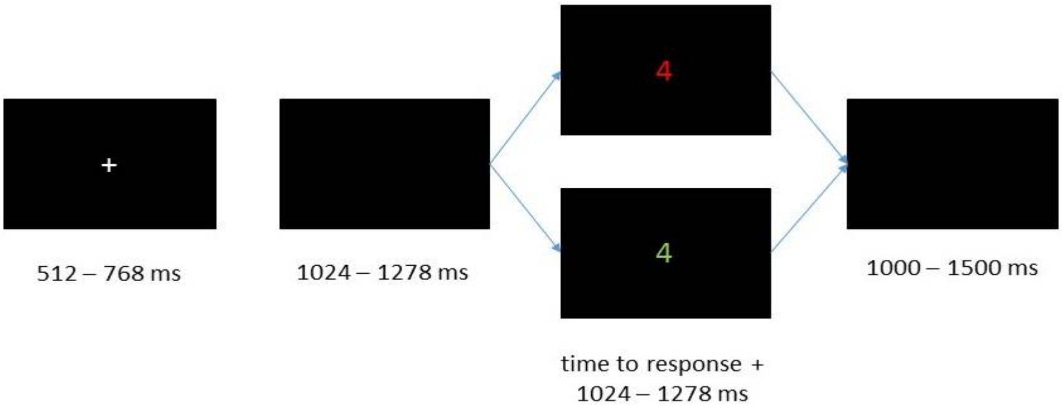


Figure 2. Representation of the sequence of one trial.

Online Thought-Probing Procedure

We administered an online thought-probing procedure by monitoring TUTs with a binary either/or question (see Weinstein 2018). This method is a subjective self-report in which the participants are intermittently asked what their current state of mind is (on task/off task) while they are working on a task. This report is one of the most frequently used methods for capturing online mind wandering at the moment of occurrence (Weinstein 2018).

Participants were randomly asked about TUTs between every fifth and tenth trial. The question was: “Where have you been with your thoughts right now?” Participants could either answer “on task” or “not on task” by pressing the right or left arrow key on the keyboard. On average, participants were probed 91.62 times (SD = 2.16) for TUTs while they worked completed 640 trials of the experimental task. On-task-reports were coded as 0 and off task reports were coded as 1 in our data.

Questionnaire of Spontaneous Mind Wandering (Q-SMW)

We used a nine-items measure of spontaneous mind wandering to assess trait mind wandering. For this we combined five items of the Mind Wandering Questionnaire (MWQ; Mrazek et al. 2013) and four items of a scale measuring spontaneous mind wandering (Carriere et al. 2013) into one questionnaire. Participants could answer these questions on a seven-point Likert scale from “almost never” (coded as 1) to “almost always” (coded as 7). Cronbach’s α showed a good internal consistency (α = .81). Because the original items were in English, they were translated into German by two people and translated back into English by another person. We present two items as examples to show their original wording and their context: “I have difficulty maintaining focus on simple or repetitive work” (Mrazek et al. 2013); “I find my thoughts wandering spontaneously” (Carriere et al. 2013).

Metronome Response Task (MRT)

We used the MRT as a more objective behavioral assessment of attentional lapses. This task was developed by Seli et al. (2013) as a new method measuring mind wandering that does not rely on self-reports. In the MRT, participants had to answer to the rhythmic beat of a metronome. A larger variability in responses (measured as the standard deviation of discrepancy) is supposed to indicate a higher frequency of attentional lapses, as lapses in executive control are thought to increase behavioral variability.

Participants heard a rhythmic metronome beat every 1600 ms for 400 times while they were looking at a black screen. They were instructed to press the spacebar on the keyboard simultaneously to the sound/rhythmic beat. We calculated the standard deviation of participants’ response discrepancy from the metronome beat after discarding the first five trials as a measure of attentional lapses.

Electrophysiological Correlates of Attentional Lapses

The EEG was recorded during the switching task. Based on previous findings, we chose mean amplitudes of lateral occipital P1 (time window: 100–140 ms after stimulus onset), central parietal P3 (time window: 300–630 ms after stimulus onset), pre-fixation cross parieto-occipital alpha power (from 1000 to 200 ms before the onset of the imperative fixation cross) from central and dorsolateral electrodes, and post fixation cross fronto-central theta power (from 0 to 500 ms after the onset of the imperative fixation) as electrophysiological covariates representing attentional lapses.

2.1.3. Procedure

After participants signed an informed consent, they completed the intelligence test under the supervision of the experimenter. This took between 50 and 60 min. After that, electrodes were administered to the scalp and participants were seated in a sound-attenuated, dimly lit cabin. Subsequently, participants worked on the switching task, working memory tasks (not included in the present manuscript), and the MRT in the same order. At the end of the session, participants completed the Q-SMW as well as a questionnaire for the assessment of demographic data. The whole procedure lasted about 3.5 h.

2.1.4. EEG Recording

While participants worked on the switching task the EEG was recorded with 32 equidistant Ag/AgCl electrodes (32Ch-EasyCap, EASYCAP, Herrsching, Germany) and amplified by a BrainAmp DC amplifier (Brain Products, Gilching, Germany). For more information on electrode positions, see Figure S1 in the Supplementary Materials. We used the aFz electrode as the ground electrode. All electrodes were initially referenced to Cz and offline re-referenced to an average reference. For the whole time we kept impedances of all electrodes below 5 kΩ. The EEG signal was recorded continuously with a sampling rate of 1024 Hz (high-pass 0.1 Hz).

2.1.5. Data Analyses

For data preparation and analyses we used the statistics software R—version 4.0.0 (R Core Team 2021). The following packages were used in R: For data processing and easier data management the package "tidyverse"(Wickham et al. 2019), for estimating Cronbach's α the package "psych" (Revelle 2020), for estimating multilevel models the package "lme4" (Bates et al. 2015) and the "optimx" algorithm (Nash and Varadhan 2011), for estimating the degrees of freedom in the multilevel models the package "lmerTest" (Kuznetsova et al. 2017), and for estimating the effect-sizes the package "effectsize" (Ben-Shachar et al. 2020). For preprocessing and quantification of EEG measures, we used EEGLAB (Delorme and Makeig 2004) and ERPLAB (Lopez-Calderon and Luck 2014) open source toolboxes on MATLAB 2018a (The MathWorks Inc., Natick, MA, USA).

Analysis of Behavioral and Self-Report Data

Responses faster than 150 ms and incorrect responses were discarded. Furthermore, the two trials following an online thought probe were excluded from the dataset, because thought probes may interrupt the ongoing task (Steindorf and Rummel 2020). Next, we conducted an intraindividual outlier analysis of the remaining trials and discarded all trials with RTs that deviated more than 3 *SD*s from the mean of the intraindividual logarithmic RT distribution. We conducted a careful outlier analysis, because outlier trials should not have any influence on the occurrence of the WPR (Coyle 2003).

In addition, participants with extremely low (sum score $\leq$ 1316) or high (sum score $\geq$ 1747) BIS performance were removed from further analyses. These cut-off values correspond to z-values <-3 and >3, which would be considered as clear outliers. This led to the exclusion of five datasets from further analyses. Moreover, we removed one additional participant because they had a mean RT that deviated more than 3 *SD*s from the sample mean.

To get the full information of the whole RT distribution, we decided not to summarize individual RTs in several bins, but to use information of every trial within each participant. Therefore, after the outlier analyses, we sorted all remaining trials in ascending order according to their RTs. All participants with at least 400 correct responses were included to ensure a sufficient and comparable number of trials across participants on the one hand and to minimize the number of participants with fewer trials who had to be excluded from the analyses on the other hand. This led to a final sample of 85 participants. We used the middle 400 trials of each participant's RT distribution and removed the remaining trials symmetrically from both ends of each intraindividual RT distribution. Single trial RTs served as the dependent variable in the following analyses. However, in the slowest 15 percent of the trials, the increases in the magnitude of the covariances accelerated whereas the negative relations became smaller (see Figure 3 and also General Discussion). As this course does not correspond to the definition of the WPR, which assumes a monotonic increase of correlations, we analyzed only the fastest 85 percent of the trials (340 trials). Further, we centered the data to the middle trial of each participant's RT distribution and rescaled the trial numbers in the range from -2 to 2. The central trial with the rescaled value 0 is equivalent to the trial with the number 170 and the trials with the values -2 and 2 correspond to the fastest trial 1 and the slowest trial 340. This is important for the interpretation of the *b*-weights in the multilevel models, both for the main effects and the interaction terms.

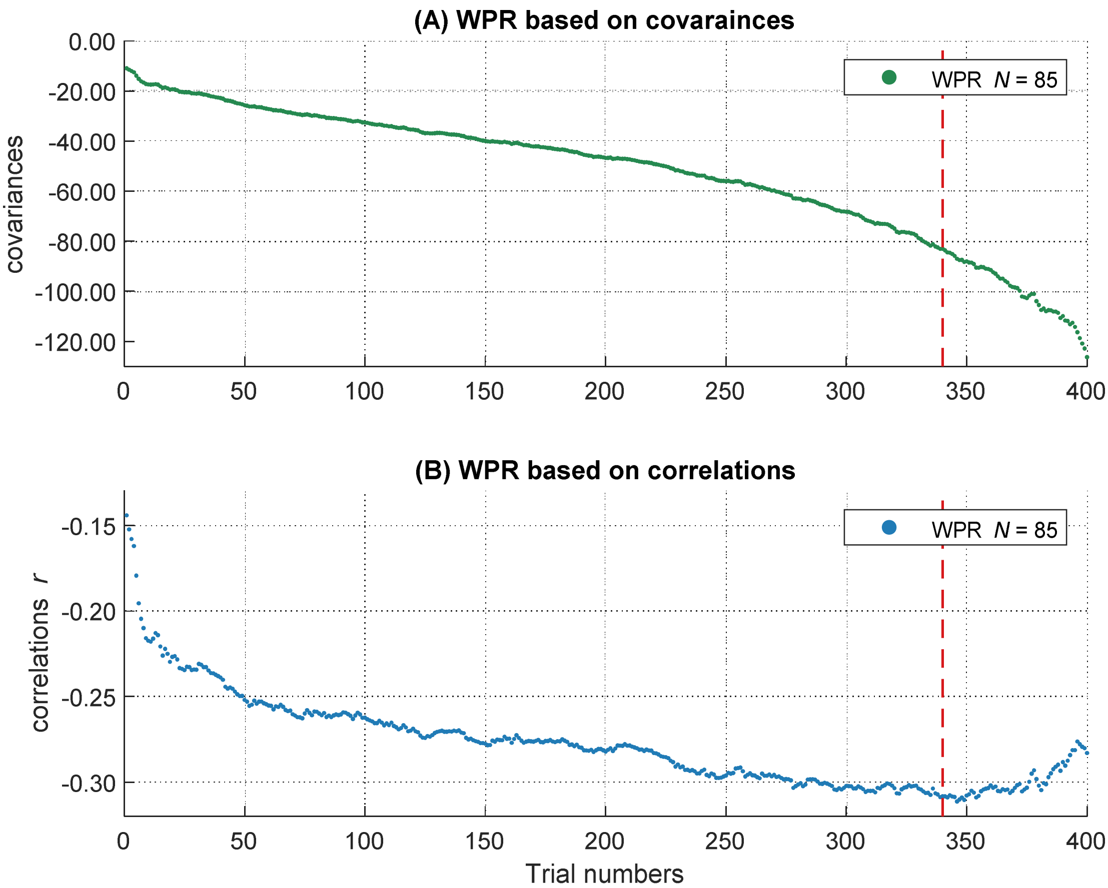


Figure 3. The increasing magnitude of negative correlations and covariances over RT distribution. The course of the covariances over 400 trials is shown above (**A**), the course of correlations over 400 trials is shown below (**B**). The dashed line represents the 85 percent threshold. Only the left part of the red dashed line was analyzed in the following multi-level analyses.

Preprocessing of Electrophysiological Data for Event-Related Potentials (ERPs)

Only correct trials were included. EEG data were filtered with an offline band-pass filter of 0.1–30 Hz. Bad channels were identified based on probability, kurtosis, and spectrum of the channel data. Data were down sampled to 512 Hz. Then, the stream of EEG data was divided into epochs of 1200 ms including the baseline window of 200 ms before stimulus onset. We conducted an independent component analysis (ICA) to identify and remove ocular artifacts and generic discontinuities based on visual inspection and the ADJUST algorithm (Mognon et al. 2011).

To ensure that experimental conditions of the switching task were evenly distributed within each participant, we identified each participant's experimental condition with the lowest number of trials and randomly drew the same number of trials from each of the other three experimental conditions. For example, when a participant had only 60 experimental trials in the odd/even-repeat condition, 60 trials each from the other three experimental conditions were randomly drawn to balance task demands. Subsequently, we calculated the ERP for each participant by averaging across trials and experimental conditions.

One participant's EEG data set was lost for technical reasons, resulting in a final sample of 84 persons for electrophysiological analyses.

Preprocessing and Time-Frequency Decomposition of Electrophysiological Data

For the time frequency analyses, most of the preprocessing steps were equal to the ERP preparation. However, data were segmented into longer epochs of 4000 ms, starting 2000 ms before the onset of the fixation cross. Also, identical to the sample composition for ERP analyses, for time-frequency analyses the total sample size consisted of N = 84 participants.

Time frequency decomposition was performed with complex Morlet wavelets with frequencies ranging from 1 to 20 Hz in 20 linearly spaced steps. To specify the width of the Gaussian distribution, the number of n cycles was set to 4. This was chosen to provide a good trade-off between temporal and frequency resolution. Decibel-normalized alpha power was calculated for each participant in the time window from 1000 to 200 ms before the onset of the fixation cross as the mean power of the frequency bands between 8 to 12 Hz recorded at parieto-occipital electrode sites. This time window was chosen to examine variations in alpha power in an attentionally undemanding phase (within the inter-trial interval) before an imperative stimulus appears, which catches participants' attentional focus back to the task at hand. To measure an internally directed attentional focus before the fixation cross was presented, the baseline window for inter-trial alpha power was set between 700 ms and 1000 ms after fixation cross onset. This allowed us to contrast alpha power of an attentionally undemanding phase to an attentionally focused phase. Decibel-normalized theta power was calculated for each participant in the time window from 0 to 500 ms after fixation cross onset as the mean power of the frequency bands between 4 to 7 Hz at fronto-central electrodes sites to examine differences in theta power after an imperative stimulus appeared and attentional resources had to be allocated. Theta power was averaged across frequencies and fronto-central electrode sites. The baseline window for task-evoked theta power was set between 1000 ms and 200 ms before the fixation cross was presented to assess attention-allocation following the presentation of the imperative stimulus. We selected the time-windows for both time-frequency domains based on findings of Arnau et al. (2020) who analyzed data from a subsample of Study 1.

Analyses of the Worst Performance Rule

In this study the WPR was examined with multilevel models based on the recommendations by Frischkorn et al. (2016). We were interested to test differences in covariances and correlations. Therefore, we followed the recommendations by Frischkorn et al. (2016) and used unstandardized as well as standardized coefficients for multilevel analyses to examine the increase of the magnitude in covariances and correlations between RT and intelligence across the RT distribution.

To get the full information of the whole RT distribution, we applied trial-by-trial analyses. To evaluate differences in the relations of intelligence and RT between faster and slower responses, we used the ascending number of the sorted trials to predict increases in RTs from the fastest to slowest trials. We included individual differences in intelligence as a between-subject predictor. A significant interaction in the multilevel model between trial number and intelligence would indicate that the relationship between RTs and intelligence changed across the RT distribution. In particular, the WPR implies a stronger negative relationship between RTs and intelligence in slower compared to faster trials. This was our baseline model.

To evaluate the effects of attentional lapses on response behavior in an ongoing task and their moderating implications on the WPR, we controlled for different combinations of attentional lapses indicators (behavioral, self-report, and electrophysiological measures). Therefore, we regressed the RTs for each corresponding sorted trial on these indicators. Afterwards we used the residuals of this regression as a new dependent variable. We then employed a stepwise procedure to test if controlling for attentional lapses reduced or removed the WPR. First, we tested if we still found a significant WPR after controlling for individual differences in attentional lapses. For this purpose, we again applied our baseline model, but instead of raw RTs, we used the residualized RTs as the new dependent variable. A non-significant WPR interaction between trial number and intelligence indicated a possible reduction of the slope of the WPR by attentional lapses. Because the difference between a significant and a non-significant result is not necessarily significant (Gelman and Stern 2006), we conducted further multilevel analyses to confirm this decrease statistically. For this purpose, we modified the multilevel models and included a dummy-coded within-subjects level-2 factor "control". This factor indicated whether participants' RTs were controlled for individual differences in attentional lapses (control = 1) or not (control = 0). If the interaction of trial number and intelligence changed as a function of this control factor—that is, if the three-way interaction between intelligence, trial number, and the control factor was significant—this would indicate that the size of the WPR changed after controlling for attentional lapses. We then used model comparisons based on the Akaike information criterion (AIC; Akaike 1998) to formally check if the introduction of this three-way-interaction (between the level-1 factor trial number, the level-2 factor control, and the between-subjects factor intelligence) improved substantially the model fit. Differences > 10 in AIC would indicate substantial differences in model fits (Burnham and Anderson 2002). For all analyses, we report degrees of freedom rounded to the nearest integer in case of non-integer numbers.

2.2. Results

The preprocessed data supporting the findings of Study 1 and the code for the statistical analysis used in this manuscript are available via the Open Science Framework (https://osf.io/5pafg/, accessed on 23 December 2021). Access to raw data of Study 1 will be granted upon request.

2.2.1. Descriptive Results

For descriptive statistics of all variables see Table 1. All variables showed acceptable to very good reliabilities, estimated with Spearman-Brown corrected odd-even correlations or Cronbach's alpha. Sample sizes differed slightly between the behavioral and the electrophysiological covariates, because EEG data from one participant were lost due to a technical error. For the correlations between all variables see Table 2. The closer the trial numbers were to each other, the higher their RTs were related.

Table 1. Descriptive statistics of all variables.

	Mean	*SD*	Reliability	*N*
ACC	96	2	—	85
RT	836.69	154.06	.99	85
Intelligence	1498.29	80.02	.79	85
IQ	94.58	16.12	.79	85
TUT	26.07	19.24	.96	85
Q-SMW over all	37.64	8.88	.81	85
Q-SMW/item	5.38	1.29	—	85
MRT	73.49	29.45	.99	85
P1 amplitude	0.94	1.34	.96	84
P3 amplitude	3.91	2.97	.99	84
Alpha power	1.20	0.94	.92	84
Theta power	0.00	0.84	.72	84

Note: ACC = percent of correct responded trials, RT = reaction time in ms (340 trials of each subject were included), Intelligence = sum score of all scales of the Berliner Intelligence Structure Test, IQ = the intelligence sum score transformed to an IQ score, TUTs = percentage of task-unrelated-thoughts, Q-SMW = mean score in the questionnaire measuring spontaneous mind wandering, MRT = response variability in ms in the metronome response task, P1 = mean amplitude of the occipital P1 in microvolts, P3 = mean amplitude of the centro-parietal P3 in microvolts, Alpha = mean parieto-occipital alpha power in decibel before an imperative stimulus was presented, Theta = mean fronto-central theta power in decibel after an imperative stimulus was presented, reliability: either estimated with the Spearman-Brown corrected correlation coefficients based on an odd-even split (RT, TUTs, MRT, P1 amplitude, P3 amplitude) or with Cronbach's α (Intelligence test score, Q-SMW, Alpha power). Theta power reliability was estimated by the correlation between the two corresponding electrodes.

Table 2. Correlations between all variables.

	1	2	3	4	5	6	7	8	9
1. Mean RT									
2. *SD* RT	.86 ***								
3. Intelligence	−.29 **	−.30 **							
4. TUT	−.12	−.27 *	.15						
5. Q-SMW	−.11	−.04	.09	.30 **					
6. MRT	.31 **	.32 **	−.27 *	−.03	−.11				
7. P1 amplitude	−.11	−.06	.03	−.02	.06	−.22 *			
8. P3 amplitude	.03	.03	−.05	.01	−.07	−.02	.27 *		
9. Alpha power	−.18	−.16	.03	−.11	−.13	.06	.06	.02	
10. Theta power	−.18	−.19	.18	.09	.09	.03	−.09	−.16	−.05

Note: Mean RT = mean reaction times (340 trials of each subject were included), *SD* RT = standard deviation of reaction times (340 trials of each subject were included), TUT = mean rate of task-unrelated thoughts, Q-SMW = mean score in the questionnaire for spontaneous mind wandering, MRT = response variability in the metronome response task, P1 amplitude = mean amplitude of occipital P1, P3 amplitude = mean amplitude of centro-parietal P3, Alpha power = mean pre-fixation cross alpha power, Theta power = mean post fixations cross theta power, * $p < .05$, ** $p < .01$, *** $p < .001$.

2.2.2. Descriptive Analyses of Covariance and Correlation Patterns over the RT Distribution

On a descriptive level, we found increases of the magnitude in covariation from the fastest, *cov trial.1* = −10.93, to the slowest trials, *cov trial.340* = −83.01, as well as increases in the magnitude of negative correlations, *r trial.1* = −.14, and *r trial.340* = −.31. The magnitude in covariances from the fastest to the slowest trial increased monotonically (see Figure 3A), whereas the correlations peaked in their magnitude after approximately 85 percent of the trials (maximum correlation: *r trial.346* = −.31). Afterwards, the magnitude of correlations decreased again (see Figure 3B). This right tail of the RT distribution is particularly interesting, because it reveals a simultaneous increase in covariations and a decrease in correlations in the slowest 15 percent of RT distribution. Together, this pattern of results indicates that the inter-individual variance in RTs increased substantially in the right tail of the RT distribution, for unknown reasons, without an accompanying increase in the relationship between RTs and intelligence. Because this pattern of results was highly surprising and violates the core prediction of the WPR to observe a monotonic increase in both covariances and correlations across the whole RT distribution, we excluded the slowest 15 percent of the trials from all further analyses. However, we will discuss this unexpected finding and its implications in the General Discussion.

2.2.3. The Worst Performance Rule with Unstandardized Coefficients (Covariances)

We analyzed the data with multilevel analyses to test if covariances between RT and intelligence revealed a significant worst performance pattern from faster to slower trials (Table 3). This analysis revealed a significant main effect of intelligence, $b = -44.18$, $t(85) = -2.77$, $p = .007$, which indicated that more intelligent participants showed faster RTs than less intelligent ones. Moreover, we found a significant worst performance interaction between intelligence and trial number, $b = -14.93$, $t(85) = -2.85$, $p = .005$, which confirms the presence of a statistically robust increase of the magnitude in covariances between RTs and intelligence over the RT distribution in our data. The worst performance interaction showed a medium effect size of η^2part = 0.09. This result can be interpreted as follows: In the central trial with the sorting number of 170 (it corresponds to trial number 0 after rescaling between −2 and 2), a participant with an intelligence test score one *SD* above the mean was about 44 ms faster in their responses than an average intelligent participant. However, in a slow trial (trial number 255, which corresponds to the rescaled trial number 1), the same participant was even 59 ms faster than an average intelligent participant, whereas their RT difference was relatively negligible in a fast trial (trial number 85, which corresponds to the rescaled trial number −1), with only a difference of about 29 ms. Taken together, our baseline model indicated a significant WPR on the level of covariances. In the next steps we examined the influences of several behavioral and self-reported measures of attentional lapses on the unstandardized WPR.

Table 3. Baseline multilevel model of the WPR on an unstandardized level.

RT On	*b*-Weight (Standard Error)	*df*	*t*-Value	Random Effect *SD*	*p*
Intercept	835.82 (15.86)	85	52.62	146.45	<.001
intelligence	−44.18 (15.98)	85	−2.77		.007
trial number	146.99 (5.20)	85	28.26	47.95	<.001
trial number × intelligence = WPR	−14.93 (5.23)	85	−2.85		.005

Note: $N = 85$. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between −2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

2.2.4. Do Individual Differences in Behavioral and Self-Reported Measures of Attentional Lapses Account for the WPR with Unstandardized Coefficients (Covariances)

In the next step, we analyzed if the increase of the magnitude in covariation disappeared after controlling for behavioral and self-report measurements of attentional lapses (TUT rates, Q-SMW scores, RT variability in the MRT). Therefore, we controlled participants' RTs for individual differences in attentional lapses. Afterwards, we tested in multilevel analyses if the covariances between RT and intelligence still revealed a significant worst performance pattern. Figure 4A shows the descriptive course of covariances between RT and intelligence over the sorted trials before and after controlling for behavioral and self-reported attentional lapses. The two-way interaction between trial number and intelligence was no longer significant after controlling for individual differences in behavioral and self-report measures of attentional lapses, $b = -8.88$, $t(85) = -1.82$, $p = .073$ (Table S1 in the Supplementary Materials).

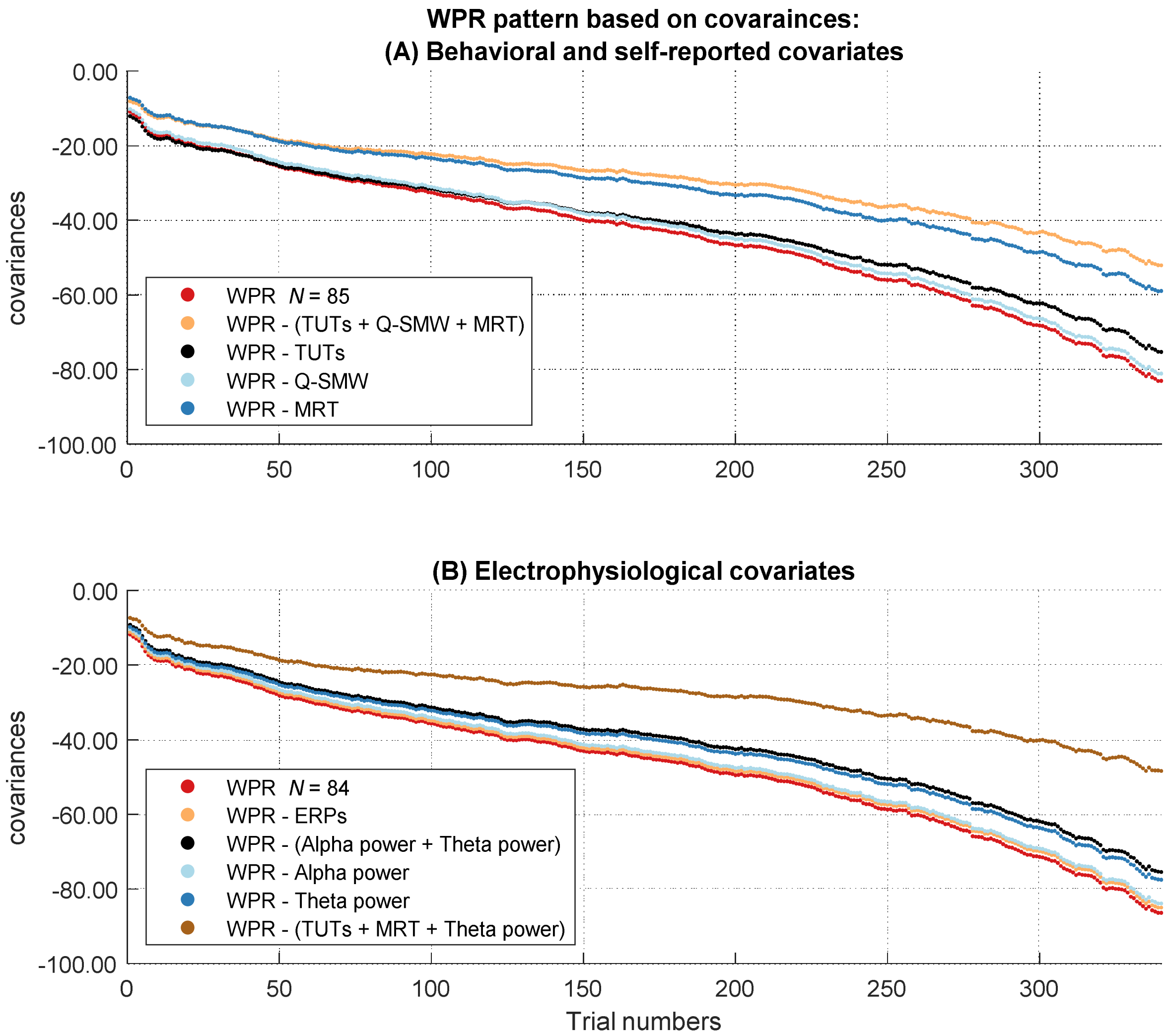


Figure 4. Course of the covariances over the RT distribution before and after controlling for the influence of the attentional lapses covariates. The figure describes the worst performance pattern in covariances before (red lines) and after (other lines) the different covariates or their combinations were partialized out of the RT variable (labeled in the boxes on the side of the dashes in the figure legend). (**A**) shows the results of the behavioral and self-reported covariates in the full sample of $N = 85$. (**B**) shows the results of the electrophysiological covariates in the subsample of $N = 84$.

To test if the changes in the WPR after controlling for individual differences in attentional lapses were significant, we merged both data sets (not controlled and controlled for attentional lapses) together and introduced a dummy-coded level-2 factor named "control" for moderation analyses in our multilevel model (Table 4). Hence, the RT variable in this multilevel model either reflected raw RTs or the residuals of those RTs after controlling for the influence of the covariates. A significant interaction between intelligence, trial number, and the control factor indicated that the increase of the magnitude in covariation between intelligence and RTs from faster to slower trials changed significantly after controlling for attentional lapses. This three-way interaction between intelligence, trial number, and the control factor was indeed significant, $b = 6.05$, $t(57630) = 25.70$, $p < .001$. The effect size of the three-way interaction revealed a small effect, η^2part = 0.01.

To additionally determine whether including the three-way interaction significantly improved the model fit, we compared our model to a more parsimonious model without this three-way interaction. Model comparison revealed a significantly better fit for the model with the three-way interaction as indicated by smaller AIC values, $\Delta_{AIC} = 655$. Taken together, these results indicate that the behavioral and self-reported attentional lapses covariates together explained substantial parts of the worst performance pattern in covariances. To assess more specifically which behavioral and self-reported indicator of attentional lapses was most relevant, we examined the specific influence of each behavioral and self-report covariate on the WPR using the same procedure.

Table 4. Full multilevel model, which tests the effect of attentional lapses covariates (TUTs + Q-SMW + MRT) on the WPR on an unstandardized level.

RT On	*b*-Weight (Standard Error)	*df*	*t*-Value	Random Effect *SD*	*p*
intercept	835.82 (15.40)	85	54.29	96.56	<.001
intelligence	−44.18 (15.49)	85	−2.85		.005
trial number	146.99 (4.91)	85	29.91	47.38	<.001
control	−835.82 (0.27)	57630	−3091.39		<.001
trial number × intelligence = WPR	−14.93 (4.94)	85	−3.02		.003
intelligence × control	15.10 (0.27)	57630	55.53		<.001
trial number × control	−146.99 (0.23)	57630	−627.78		<.001
trial number × intelligence × control	6.05 (0.24)	57630	25.70		<.001

Note: N = 85. For each participant, 340 trials were included in the analysis. Data were centered to the trial with the sorted number of 170 and rescaled between −2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the covariance.

Task-Unrelated Thoughts (TUTs)

If we controlled for TUTs, we still observed a significant worst performance interaction in the baseline model, $b = -12.98$, $t(85) = -2.55$, $p = .013$ (Table S2 in the Supplementary Materials). Nevertheless, the significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated that TUTs had an effect on the worst performance pattern, $b = 1.95$, $t(57630) = 12.24$, $p < .001$ (Table S3 in the Supplementary Materials). Model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 147$. The effect size was very small, η^2part = 0.00. Taken together, these results indicate that self-reported TUTs accounted for small parts of the WPR in covariances.

Questionnaire of Spontaneous Mind Wandering (Q-SMW)

If we controlled for Q-SMW scores, we still observed a significant worst performance interaction in the baseline model, $b = -14.73$, $t(85) = -2.81$, $p = .006$ (Table S4 in the Supplementary Materials). The three-way interaction between intelligence, trial number, and the control factor in the full model was not significant, indicating that the worst performance pattern did not change after controlling for Q-SMW scores, $b = 0.20$, $t(57630) = 1.34$, $p = .179$, η^2part = 0.00 (Table S5 in the Supplementary Materials). Consequently, model comparison did not indicate a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 0$. Taken together, these results indicate that Q-SMW scores did not contribute to the WPR in covariances.

Metronome Response Task (MRT)

After controlling for the RT variability in the MRT, we still observed a significant worst performance interaction in the baseline model, $b = -10.57$, $t(85) = -2.09$, $p = .039$ (Table S6 in the Supplementary Materials). The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a smaller worst performance pattern after controlling for RT variability in the MRT, $b = 4.36$, $t(57630) = 19.60$, $p < .001$ (Table S7 in the Supplementary Materials). Also, model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 380$. Effect size estimation revealed a small effect, η^2part = 0.01. Taken together, these results indicate that RT variability in the MRT accounted for some parts of the WPR in covariances.

2.2.5. Do Individual Differences in Electrophysiological Measures of Attentional Lapses Account for the WPR with Unstandardized Coefficients (Covariances)

Figure 4B shows the descriptive course of covariances between RT and intelligence over the sorted trials before and after controlling for the electrophysiological covariates representing attentional lapses. The baseline multilevel model indicated a significant interaction between trial number and intelligence in this subset, $b = -15.21$, $t(84) = -2.88$, $p = .005$, η^2part = 0.09 (Table S8 in the Supplementary Materials).

ERP Analyses

If we controlled for individual differences in mean occipital P1 and mean centro-parietal P3 amplitudes, the two-way interaction between trial number and intelligence remained significant in the baseline model, $b = -14.99$, $t(84) = -2.84$, $p = .006$ (Table S9 in the Supplementary Materials). We observed no significant three-way interaction between intelligence, trial number, and the control factor in the full model, indicating that the size of the WPR did not change after controlling for the ERP mean amplitudes, $b = 0.22$, $t(56952) = 1.42$ $p = .156$, η^2part = 0.00 (Table S10 in the Supplementary Materials). Consequently, model comparison did not reveal a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 1$. Taken together, these results indicate that the mean occipital P1 amplitude and the mean parietal P3 amplitude did not account for the WPR in covariances.

Time-Frequency Analyses

If we controlled for individual differences in alpha and theta power, the two-way interaction between trial number and intelligence remained significant in the baseline model, $b = -13.14$, $t(84) = -2.55$, $p = .013$ (Table S11 in the Supplementary Materials). Still, the significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a decrease in the worst performance pattern after controlling for alpha and theta power, $b = 2.06$, $t(56952) = 9.98$ $p < .001$ (Table S12 in the Supplementary Materials). Model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 98$. However, this effect was very small, η^2part = 0.00. Taken together, these results indicate that the time-frequency covariates accounted for small parts of the WPR in covariances. To detect the unique influence of the two different time-frequency covariates on the WPR, we estimated the models for both covariates separately.

Alpha-Power

After controlling for individual differences in alpha power, the two-way interaction between trial number and intelligence remained significant in the baseline model, $b = -14.96$, $t(84) = -2.87$, $p = .005$ (Table S13 in the Supplementary Materials). More importantly, there was no significant three-way interaction between intelligence, trial number, and the control factor in the full model, indicating that the size of the WPR did not change after controlling for alpha power, $b = 0.24$, $t(56952) = 1.41$ $p = .159$, η^2part = 0.00 (Table S14 in the Supplementary Materials). Model comparison did not reveal a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 0$. Taken together, these results indicate that individual differences in inter-trial alpha power did not account for the WPR in covariances.

Theta-Power

After controlling for individual differences in theta power, the two-way interaction between trial number and intelligence remained significant in the baseline model, $b = -13.48$, $t(84) = -2.57$, $p = .012$ (Table S15 in the Supplementary Materials). The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a significant change of the worst performance pattern after controlling for theta power, $b = 1.72$, $t(56952) = 9.73$, $p = .001$ (Table S16 in the Supplementary Materials). Model comparison also showed a better fit for the model with the three-way interaction, $\Delta_{AIC} = 96$, but the effect size of the three-way interaction was very small, η^2part = 0.00. Taken together, these results indicate that theta power accounted for small parts of the WPR in covariances.

The Combined Effect on the Unstandardized Worst Performance Pattern of All Predictors with a Substantial Contribution (TUTs, MRT, Theta-Power)

After controlling for individual differences in covariates with a unique contribution to the explanation of the WPR, we examined their combined influence. The two-way interaction between trial number and intelligence was no longer significant in the baseline model, $b = -7.76$, $t(84) = -1.59$, $p = .116$ (Table S17 in the Supplementary Materials). The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a substantial change of the worst performance pattern after controlling for all three predictors, $b = 7.45$, $t(56952) = 28.68$ $p < .001$ (Table S18 in the Supplementary Materials). Model comparison revealed a significantly better fit for the full model with the three-way interaction, $\Delta_{AIC} = 815$. The estimation of the effect size indicated a small effect, η^2part = 0.01. All in all, these results indicate that TUT rates, variability in the MRT, and theta power together fully explained the worst performance pattern in covariances.

2.2.6. The Worst Performance Rule with Standardized Coefficients (Correlations)

On the level of correlations, we did not find a significant worst performance pattern in the baseline multilevel model, $b = -0.02$, $t(85) = -1.10$, $p = .276$ (Table S19 in the Supplementary Materials). We also did not find a significant worst performance interaction, $b = -0.02$, $t(84) = -0.91$, $p = .359$, in the subset with psychophysiological covariates (Table S28 in the Supplementary Materials). The worst performance interaction revealed a small effect size of η^2part = 0.01. We observed a small descriptive increase in the magnitude of negative correlations from the first to the last trial of $\Delta_r = .08$ (Figure 3B).

2.2.7. Do Individual Differences in Behavioral and Self-Reported Measures of Attentional Lapses Account for the WPR with Standardized Coefficients (Correlations)

Because there was no significant worst performance interaction in the baseline multilevel model with standardized coefficients and we found no suppressor effect of the covariates on this interaction, we will not report the baseline models without the effect of any covariates (they can be found in Tables S20, S22, S24, S26, S29 and S31 in the Supplementary Materials). The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a change in the worst performance pattern after controlling for the behavioral and self-reported covariates, $b = 0.01$, $t(57630) = 8.70$, $p < .001$ (Table S21 in the Supplementary Materials). Model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 73$. However, the effect size of η^2part = 0.00 suggested that this effect was very small. Taken together, the behavioral and self-reported attentional lapses covariates together explained very small parts of the (not significant) worst performance pattern in correlations. To assess more specifically which behavioral and self-reported indicator of attentional lapses was most relevant for this effect, we additionally examined the individual influence of each of these covariates on the WPR in correlations by using the already known procedure (Figure 5A).

Task-Unrelated Thoughts (TUTs)

The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a smaller worst performance pattern after controlling for TUTs, $b = 0.01$, $t(57630) = 9.49$, $p < .001$ (Table S23 in the Supplementary Materials). Model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 88$. The effect size of η^2part = 0.00 indicated a very small effect of TUTs on the WPR. Taken together, these results indicate that self-reported TUTs accounted for a very small part of the WPR in correlations.

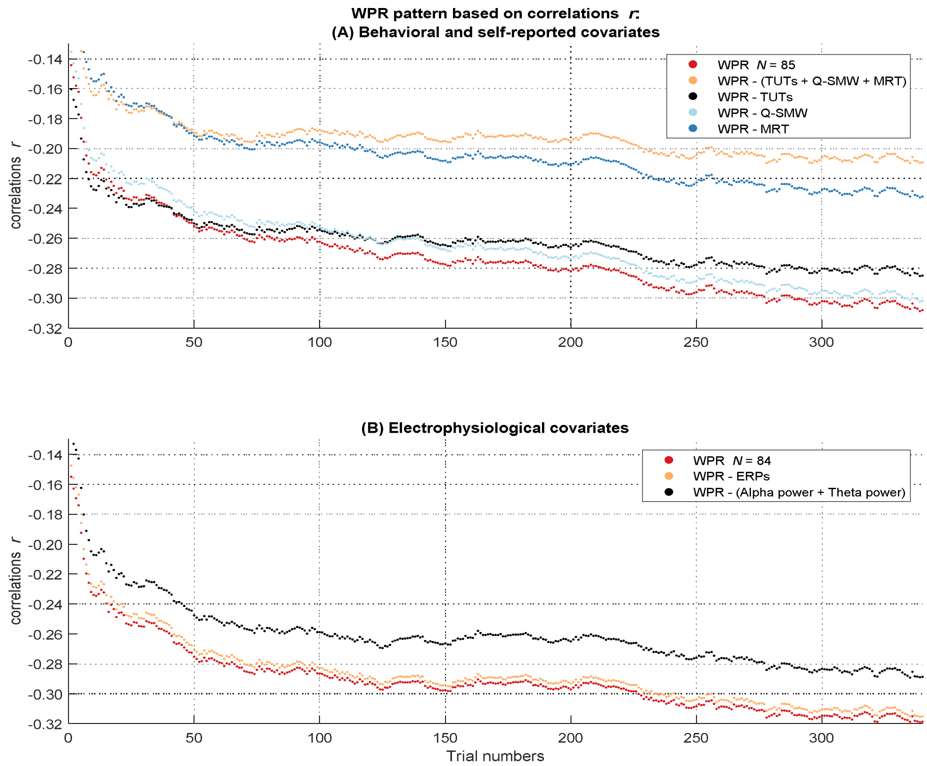


Figure 5. Course of the correlations over the RT distribution before and after controlling for the influence of the attentional lapses covariates**.** The figure describes the worst performance pattern in correlations before (red lines) and after (other lines) the different covariates or their combinations were partialized out of the RT variable (labeled in the boxes on the side of the dashes in the figure legend). (**A**) shows the results of the behavioral and self-reported covariates in the full sample of $N = 85$. (**B**) shows the results of the electrophysiological covariates in the subsample of $N = 84$.

Questionnaire of Spontaneous Mind Wandering (Q-SMW)

The three-way interaction between intelligence, trial number, and the control factor in the full model was not significant, indicating that the worst performance pattern did not change after controlling for Q-SMW scores, $b = 0.00$, $t(57630) = 1.39$, $p = .165$, η^2part = 0.00 (Table S25 in the Supplementary Materials). Consequently, model comparison did not indicate a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 0$. Taken together, these results indicate that Q-SMW scores did not contribute to the WPR in correlations.

Metronome Response Task (MRT)

The significant three-way interaction between intelligence, trial number, and the control factor in the full model indicated a smaller worst performance pattern after controlling for RT variability in the MRT, $b = 0.00$, $t(57630) = 3.47$, $p < .001$ (Table S27 in the Supplementary Materials). Model comparison revealed a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 10$. We found only a very small effect of the MRT on the WPR, η^2part = 0.00. Taken together, these results indicate that RT variability in the MRT accounted only for a very small part of the WPR.

2.2.8. Do Individual Differences in Electrophysiological Measures of Attentional Lapses Account for the WPR with Standardized Coefficients (Correlations)

ERP Analyses

There was no significant three-way interaction between intelligence, trial number, and the control factor in the full model, indicating that the size of the WPR did not change if we controlled for the ERP amplitudes, $b = 0.00$, $t(56952) = -0.32$ $p = .749$, η^2part = 0.00

(Table S30 in the Supplementary Materials). Consequently, model comparison did not show a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 2$. Taken together, these results indicate that mean occipital P1 and centro-parietal P3 amplitudes could not account for the WPR in correlations (see Figure 5B).

Time-Frequency Analyses

The three-way interaction between intelligence, trial number, and the control factor in the full model indicated that there was no difference in the worst performance pattern after controlling for the combined influence of mean alpha and theta power, $b = 0.00$, $t(56952) = 1.17$ $p = .243$, η^2part = 0.00 (Table S32 in the Supplementary Materials). Model comparison also did not show a better fit for the full model with the three-way interaction, $\Delta_{AIC} = 1$. Taken together, these results suggest that the time-frequency covariates could not account for the WPR in correlations.

2.3. Discussion

Our findings provided some evidence for the attentional lapses account of the worst performance rule. We found a significant increase in the magnitude of covariances between intelligence and RTs from the fastest to the slowest RTs (i.e., a WPR). This increase was less strong when we controlled for inter-individual differences in several of the self-reported attentional lapses measures. Notably, after combining different attentional lapses measures and controlling for these, the WPR disappeared. Thus, inter-individual differences in the propensity of attentional lapses did fully explain the WPR in the present data set on the level of covariances. Nevertheless, it has to be stressed that the combined effect of attentional lapses on the WPR was very small (η^2part = 0.01). It is possible that we were only able to detect this small effect of attentional lapses on the WPR due to the high statistical power of the multi-level account and the trial-by-trial analyses.

However, there was no significant WPR on the level of correlations. Nevertheless, descriptively, there was still an increase in the negative correlations with a magnitude of about $r = -.08$, which is consistent with former research investigating the WPR on a descriptive level (e.g., Fernandez et al. 2014). Again, the increase in the magnitude was reduced after controlling for self-reported attentional lapses but the present data do not address the question of the extent to which attentional lapses can account for the WPR on the level of correlations, as we did not find a significant WPR on that level. Apparently, the statistical power was rather low for a detection of an effect with the magnitude of the WPR on the correlational level. Thus, one reason for why we did not observe a significant WPR on the correlational level probably was our somewhat low sample size. We tackled this problem with our second study.

2.3.1. Influence of Covariates on the WPR in Covariances

Different covariates of attentional lapses showed a significant influence on the WPR and controlling for them reduced the increasing magnitude in covariances. In particular, controlling for self-reported attentional lapses led to a reduction of the worst performance pattern and provided evidence for the attentional lapses account. However, we found some unexpected relations between self-reported attentional lapses and participants' mean RTs/RT variability as well as between TUTs and intelligence. These correlations between the measures were not in line with former findings and contrary to theoretical predictions. In detail, individuals who reported more attentional lapses, measured by TUTs, showed faster RTs and less RT variability as well as higher intelligence test scores in our data. The attentional lapses account, in contrast, states that individuals with lower cognitive abilities should experience more attentional lapses and should be slower in their responses. Also, individuals with lower cognitive abilities should show more variability of their responses within a certain task. Previous studies showed typically the opposite direction of correlations compared to our findings (e.g., Kane et al. 2016; McVay and Kane 2009, 2012; Randall et al. 2014; Robison et al. 2020; Welhaf et al. 2020). Possible reasons for these surprising correlations may be the size or composition of our sample and will be discussed below.

Besides self-reported attentional lapses, one of the objective measures (i.e., the RT variability in the MRT) also contributed to the explanation of the WPR. The MRT is typically used as an alternative, more objective measure of attentional lapses (Anderson et al. 2021; Seli et al. 2013). However, Figures 4 and 5 show that the MRT explained not only the slope of the WPR but also large parts of the covariances and correlations over the whole RT distribution. It is plausible that the MRT and the assigned decision making task for assessing the WPR possess some overlaps. Performances in both tasks were measured via RTs, which are determined by different processes, such as the encoding of stimuli and the preparation of the motor response. Thus, controlling for MRT variability in our reaction time task means that we also have controlled for some variance resulting from these processes. This could be the reason for the similar reduction of the covariances and correlations over the whole RT distribution after controlling for the MRT.

It must be noted that several of our covariates did not contribute to the WPR. This was especially surprising in case of the Q-SMW, as the underlying construct (i.e., mind wandering tendencies) are supposed to be strong predictors of attentional lapses. In the present sample, questionnaire scores were moderately correlated with self-reported attentional lapses during the task. This is consistent with earlier studies showing that mind wandering trait questionnaires predict the frequency with which attentional lapses are experienced while participants work on an experimental task (Mrazek et al. 2013; Schubert et al. 2020). Mind wandering is, however, a broad construct covering a range of attentional phenomena. This may explain why the thought-probing measure of attentional lapses but not the global mind wandering questionnaire explained parts of the WPR.

On the electrophysiological level, the mean amplitudes of the lateral-occipital P1 and the centro-parietal P3 as well as mean parieto-occipital inter-trial alpha power showed no effects on the WPR. Only mean stimulus-evoked fronto-central theta power changed the course of covariances over the RT distribution. It is surprising that the electrophysiological covariates did not change the worst performance pattern, because former studies found relations of the centro-parietal P3 to TUTs (Kam and Handy 2013; Smallwood et al. 2008), to sustained attention (O'Connell et al. 2009), and to the allocation of cognitive resources (Allison and Polich 2008; Kok 2001). Likewise, former studies demonstrated that attentional lapses and neural processing of stimuli via the occipital visual P1 are related (Baird et al. 2014; Kam et al. 2011). Also, inter-trial alpha power, which reflects internally directed mental states and which was shown to be strongly predictive for the experience of attentional lapses (Arnau et al. 2020), could not explain the WPR. Altogether, it seems that the chosen electrophysiological covariates did not account for the WPR, except for the very small effect of mean theta power.

2.3.2. Influence of Covariates on the WPR in Correlations

Self-reported attentional lapses and intra-individual RT variability of the MRT accounted for the WPR on the level of correlations. Descriptively it seemed that the MRT explained large parts of the correlations, but the effect of the MRT on the WPR in the multilevel models was slightly smaller compared to the effect of self-reported attentional lapses. This underlines the just discussed proposition that the MRT accounts for RT properties unrelated to the slope of the WPR. In contrast to the analyses of the covariances, on the level of correlations mean fronto-central theta power could not account for the worst performance pattern. Again, all other covariates revealed no effect on WPR.

2.3.3. Low Correlation and Unpredicted Correlations with Attentional Lapses Measures

There were hardly any correlations between different attentional lapses measures or their psychophysiological correlates. It is well known that individual occurrences of attentional lapses depend on personal and context-related variables, which means that the construct of attentional lapses shows a multiverse structure (Robison et al. 2020).

Nevertheless, beyond the multiverse structure of the attentional lapses construct, the low correlations should also be considered as challenging for attentional lapses research. The absence of relations between different attentional lapses measures raises the question of construct validity. If we try to capture a certain ability or a state of attention with a multimethod approach, these measures should all reflect a common latent construct. This assumption should be empirically reflected in—at least—small correlations between those measures. A problem of attentional lapses research is the vague definition of attentional lapses, which leads to more degrees of freedom in its operationalization. Future research should further examine the construct validity of attentional lapses.

In contrast to former findings (e.g., Kane et al. 2016; McVay and Kane 2009, 2012; Randall et al. 2014; Robison et al. 2020; Welhaf et al. 2020) and to predicted relations, we found that TUTs and cognitive abilities as well as RT and RT variability measures were not related or that their correlations pointed in the unpredicted direction.

2.3.4. Interim Conclusion

Generally, each attentional lapses indicator explained unique parts of the worst performance pattern. When we examined the common influence of different attentional lapses covariates on the WPR, the WPR disappeared fully on the level of covariances (Figure 4). On a descriptive level, we also observed a clear change in the pattern of correlations from the fastest to the slowest RTs (Figure 5). Our findings are in line with the idea that attentional lapses have different facets, which should be captured by different indicators (Robison et al. 2020). Due to the multiverse structure, measures of attentional lapses do not need to converge (e.g., Mrazek et al. 2013; Schubert et al. 2020; Seli et al. 2013). We found the same pattern in our results with weak or absent correlations between the different measures of the attentional states (Table 2). This underscores the necessity of the multimethod approach, which we chose in the present study by assessing attentional lapses with self-reports, objective indicators, and psychophysiological measures to capture individual differences in this construct as comprehensively as possible, which is as a major advantage of our study.

Nevertheless, despite the clear descriptive worst performance pattern in correlations in our study and despite the recent meta-analysis by Schubert (2019), who reported robust evidence for the presence of the WPR, we did not find a significant WPR on the level of correlations. There are several possible explanations for this. First, the sample size in this study was small and consequently the statistical power was too low to detect a significant WPR in our multilevel models given the small effect size. Additionally, the multilevel approach, proposed by Frischkorn et al. (2016), considers the uncertainty in correlation estimates. In a small sample, the confidence intervals of the estimators are quite large, and therefore the differences in correlations may not have become significant in our analyses. A larger sample size would minimize the uncertainty in the estimators (Schönbrodt and Perugini 2013).

Second, the absence of the WPR may also be attributed to the heterogeneity of our sample. It is known that student samples differ in many psychological variables compared to general population or even representative samples (Hanel and Vione 2016). In addition, age may have affected participants' response behavior in self-reported attentional lapses and RTs. For example, previous studies found fewer instances of attentional lapses in older people as compared to younger people (e.g., Arnicane et al. 2021; Frank et al. 2015; Krawietz et al. 2012; Maillet et al. 2018, 2020; Maillet and Schacter 2016). Furthermore, it is well established that older participants respond slower compared to younger participants (e.g., Verhaeghen and Salthouse 1997). As we have recruited an age-heterogeneous sample, age differences may have obscured our covariance structure. We found no evidence for an age-related decline in the frequency of reported attentional lapses in our sample ($r = -.14$, $p = .201$), but older participants showed slower responses ($r = .26$, $p = .016$).

Third, the measurement took place in a highly controlled laboratory situation. In order to achieve a clear measure of brain activity with the EEG, participants were individually seated in a shielded cabin so that any kind of noise was reduced to a minimum. Consequently, participants of our study probably experienced fewer distractions than in standard behavioral laboratory studies. It is possible that the special laboratory situation of our study influenced the occurrence and experience of attentional lapses and in consequence the magnitude of the WPR.

Because of the mentioned shortcomings of our first study (low power resulting from the small sample size, heterogeneity of the sample, and unexpected correlations between intelligence, RTs or RT variability and self-reported attentional lapses), we reanalyzed an already published data set with our approach to test whether the results and descriptive patterns would replicate in an independent larger and more homogenous student sample. In Study 2, we were particularly interested if we would find a significant WPR (and a reduction thereof when controlling for inter-individual differences in attentional lapses) on the correlational level when the statistical power was increased.

3. Study 2

3.1. Materials and Methods

To replicate our results in an independent sample, we reanalyzed the data set from two previously published studies by Kane et al. (2016) and Welhaf et al. (2020). From these previous studies it is already known that the correlations between TUTs, RTs, and intelligence are in accordance with expectations, which we consider an advantage of this data set. The data for Study 2 are available via the Open Science Framework. Use https://osf.io/9qcmx/ (accessed on 5 February 2021) to access the raw data and use https://osf.io/5pafg (accessed on 23 December 2021) to get access to additional data, which are not provided via the previous link.

3.1.1. Participants

At three measurement occasions, Kane et al. (2016) recruited a total sample of 545 undergraduates, aged between 17 and 35 years, from the University of North Carolina at Greensboro and Minority-Serving state university. For the present analyses, the number of available data-sets differed between the tasks (arrow-flanker N = 481, letter-flanker N = 426, number-stroop N = 481, sustained attention to response task [SART] N = 486). In consequence of outlier analyses, different numbers of participants remained for each task (see Data Preparation below for specific information). We analyzed the data with the same analysis strategy as used in Study 1. The mean age of the analyzed subsample was 18.92 (SD = 1.91), 66.94 percent of the sample were female. Five participants did not disclose their gender.

3.1.2. Materials

Sustained Attention Task (SART)

Participants had to press the space bar in go-trials (89% of 675 trials) and to withhold their response in no-go-trials (11% of 675 trials). Go-trials were indicated by words of the category “animals” and no-go trials were indicated by words of the category “vegetables”. We used RTs of go-trials as dependent variable, consistent with the analyses by Welhaf et al. (2020).

Letter-Flanker

Participants had to decide whether the presented target letter “F” appeared normally or backwards. The letter was presented amid six distractors on the horizontal line. In total, participants had to respond in 144 trials, which consisted of 24 neutral trials (the target letter was presented amid dots), 48 congruent trials (the target and the distractors were the same letters and pointed in the same direction), 24 trials of an incongruent condition (the target and the distractors were the same letters, but only five out of the six distractors pointed in the same direction as the target), 24 stimulus-response incongruent trials (the target and the distractors were the same letters but pointed in the opposite directions), and 24 stimulus-stimulus incongruent trials (the distractors consists of the letters “E” and “T”, which were additionally tilted by 90 and 270 degrees). We used the RTs of correctly solved congruent and neutral trials as dependent variable, consistent with the analyses by Welhaf et al. (2020).

Arrow-Flanker

Participants had to decide whether a centrally presented arrow pointed to the right or to the left. The arrow was presented amid four distractors on the horizontal line. In total participants had to respond in 192 trials, which consisted of 48 neutral trials (the target was presented amid dots), 48 congruent trials (the target and the distractors pointed in the same direction), 48 stimulus-response incongruent trials (the target and the distractors pointed in the opposite directions), and 48 stimulus-stimulus incongruent trials (the distractor arrows pointed upwards). We used the RTs of correctly solved congruent and neutral trials as dependent variable, consistent with the analyses by Welhaf et al. (2020).

Number-Stroop

In each trial, two to four digits were presented in a row. Participants had to count the quantity of presented digits, while they had to ignore their meaning. They responded by pressing one of three labeled keys. The condition could be congruent, if the quantity of presented digits was equal to their meaning (e.g., 4444 or 333), or incongruent, if the quantity of presented digits differed from their meaning (e.g., 2222 or 44). Twenty percent of the 300 trials were incongruent trials. We used the RTs of correctly solved congruent trials as dependent variable, consistent with the analyses by Welhaf et al. (2020).

Working Memory Capacity

In Study 2 we used WMC as an independent variable to measure cognitive abilities. This is unproblematic, because the WPR was also observed in the relations between RTs an WMC (McVay and Kane 2012; Schmiedek et al. 2007; Unsworth et al. 2010; Welhaf et al. 2020). Furthermore, WMC is highly related to intelligence (Conway et al. 2002; Kane et al. 2005; Kyllonen and Christal 1990; Oberauer et al. 2005) and therefore a suitable alternative measure of cognitive abilities beside intelligence. Moreover, individual differences in attentional lapses should account for individual differences in both WMC as well as intelligence (Kane et al. 2008; Shipstead et al. 2016). WMC was measured with six different tasks. Four of these tasks required maintaining serially presented memory items while participants had to repeatedly engage in an unrelated secondary task (Operation-Span, Sentence-Span, Symmetry-Span, and Rotation-Span). Participants' responses were coded as correct if they recognized memory items in their correct serial position. The two remaining tasks measuring WMC required participants' ability for updating previously memorized items (Running-Span-Task and Updating-Counters). Participants' responses were coded as correct if they recognized the updated memory items. For more detailed information on the tasks, see Kane et al. (2016). We used the latent WMC scores calculates by Welhaf et al. (2020). These were estimated with confirmatory factor analyses and full information maximum likelihood was used to account for missing data when the factor scores were computed.

Online Thought-Probing Procedure

At each online thought probe, participants were asked: "What are you thinking about?" and had to answer by pressing one of eight keys which most closely matched their thought content. They could choose between: (1) The task—on-task thoughts; (2) Task experience/performance—thoughts about one's own task performance; (3) Everyday things—thoughts about routine things; (4) Current state of being—thoughts about one's own current physical or emotional state; (5) Personal worries—thoughts about one's worries and concerns; (6) Daydreaming—fantastic thoughts, which are decoupled from reality; (7) External environment—thoughts about the immediate external environment; (8) Other—thoughts which do not fit in one of the other seven categories. Kane et al. (2016) as well as Welhaf et al. (2020) coded all answers of the categories 1 and 2 as on-task and all answers of the categories 3 to 8 as off-task thoughts (TUTs). We used the rate of these TUTs as a measure of attentional lapses. The attentional lapses covariate contained 45 thought probes from the SART, 20 from the Number-Stroop task, 20 from the Arrow-Flanker task, and 12 from the Letter-Flanker task, as well as 12 from an otherwise-not further reported and analyzed 2-back task.

3.1.3. Data Preparation and Analyses

Within each task, we removed participants with fewer than 50 percent of correctly answered trials. In the next step, the two trials following thought probes, responses faster than 150 ms and slower as 3000 ms, incorrect responses, and trials of the non-analyzed conditions were discarded within each task. Afterwards, we removed all participants who showed higher logarithmical accuracy z-scores than 3 *SD*s from the sample mean within each task. After that, we conducted an intra-individual outlier analysis and discarded all trials with RTs that deviated more than 3 *SD*s from the mean of the intra-individual logarithmized RT distribution within each task. Finally, within each task, we removed the participants with higher mean RT z-scores than 3 *SD*s from the sample mean.

We sorted all of the remaining trials within each participant in each task in the ascending order according to their RTs. All participants with at least 60 remaining trials in the arrow-flanker task, 50 remaining trials in the letter-flanker task, 170 remaining trials in the number-stroop task, as well as at least 200 remaining trials in the SART were included to ensure a sufficient and comparable number of trials on the one hand and to minimize the number of participants with fewer trials who had to be excluded from the analyses on the other hand. In consequence of this minimal amount of trials criterion, we removed different numbers of participants within each task from further analyses. This led to a final sample of 463 participants in the arrow-flanker task (28 participants were removed as outliers), 416 participants in the letter-flanker task (10 participants were removed as outliers), 460 participants in the stroop task (21 participants were removed as outliers), and 441 participants in the SART (45 participants were removed as outliers). We used the middle trials of each participant's RT distribution in each task and removed the remaining trials symmetrically from both ends of the intraindividual distribution. Multilevel analyses were conducted in the same way as in Study 1. We included all of the four tasks in one model and added the task as an additional effect-coded level-3 factor. The factor levels of the task-factor were contrasted to the SART. All multilevel models were estimated using the "nlminb" optimizer, except for the two full models in which the WPR was controlled for TUT rates, because those two models only converged with the "L-BFGS-B" optimizer algorithm.

3.2. Results

3.2.1. Descriptive Analyses

Descriptive statistics are shown in Table 5 and the correlations between all relevant variables are shown in Table 6. Mean RTs as well as RT variability of the four different tasks were highly correlated. In contrast to Study 1, the correlations between TUTs and RTs, TUTs and RT variability, as well as between TUTs and cognitive abilities (in this case WMC) pointed in the hypothesized directions. For WMC, reliability estimation across the working memory tasks revealed an acceptable internal consistency with Cronbach's α= .78.

Table 5. Descriptive statistics of all RT variables in Study 2.

	Mean	***SD***	**Reliability**	***N***
RT AF	461.03	49.65	.99	463
RT LF	532.35	85.93	.99	416
RT Stroop	508.34	49.86	.99	460
RT SART	510.62	81.94	.99	441

Note: RT AF = reaction time in the arrow-flanker task, RT LF = reaction time in the letter-flanker task, RT Stroop = reaction time in the number-stroop task, RT SART = reaction time in the SART, reliabilities were estimated with Spearman-Brown corrected odd-even split correlations.

Over the RT distributions, we found the same pattern of correlations in most of the four tasks as we did in Study 1. After about 85 percent of the selected range of the RT distributions, the negative increases in the magnitude of the covariances accelerated, whereas the magnitude of the negative correlations decreased at this point (Figure 6). These descriptive findings were consistent over the different tasks and replicated our unexpected results from Study 1. For the comparability to the results of Study 1, we only analyzed the fastest 85 percent of each participant's trials. Every participant contributed 51 trials from the arrow-flanker task, 43 trials from the letter-flanker task, 145 trials from the number-stroop task, and 170 trials from the SART to the multilevel models. Again, in each task, we centered the data to participants' central trials and rescaled the trial numbers between −2 and 2.

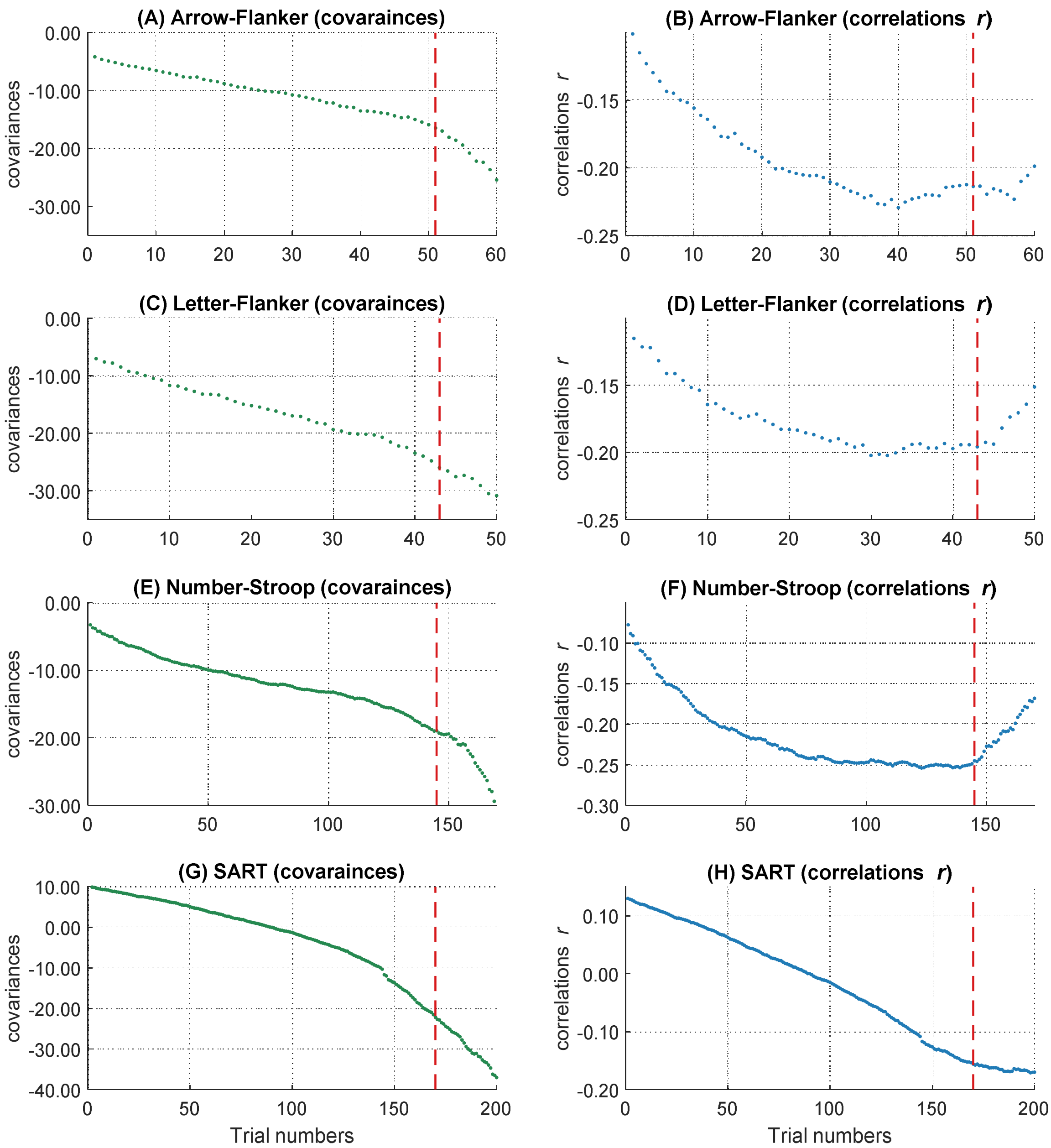


Figure 6. The increasing magnitude of negative correlations and covariances over RT distributions. The courses of the covariances in the four different tasks are shown on the left side (**A**,**C**,**E**,**G**). The courses of the correlations in the four different tasks are shown on the right side (**B**,**D**,**F**,**H**). The dashed lines represent the 85 percent thresholds. Only the left parts of the dashed lines were analyzed in the following multi-level analyses.

Table 6. Correlations between all variables.

	1	2	3	4	5	6	7	8	9
1. Mean RT AF									
2. *SD* RT AF	.65 ***								
3. Mean RT LF	.53 ***	.42 ***							
4. *SD* RT LF	.34 ***	.40 ***	.73 ***						
5. Mean RT Stroop	.63 ***	.40 ***	.49 ***	.33 ***					
6. *SD* RT Stroop	.31 ***	.48 ***	.30 ***	.32 ***	.52 ***				
7. Mean RT SART	.11 *	−.04	.12 *	.05	.24 ***	.02			
8. *SD* RT SART	.13 **	.18 ***	.14 **	.16 **	.23 ***	.28 ***	.21 ***		
9. WMC	−.20 ***	−.22 ***	−.19 ***	−.20 ***	−.23 ***	−.25 ***	−.01	−.23 ***	
10. TUT	.12 *	.20 ***	.19 ***	.26 ***	.16 **	.22 ***	−.02	.21 ***	−.23 ***

Note: Mean RT AF = mean reaction times in the arrow-flanker task, *SD* RT AF = standard deviation of reaction times in the arrow-flanker task, Mean RT LF = mean reaction times in the letter-flanker task, *SD* RT LF = standard deviation of reaction times in the letter-flanker task, Mean RT Stroop = mean reaction times in the number-stroop task, *SD* RT Stroop = standard deviation of reaction times in the number-stroop task, Mean RT SART = mean reaction times in the SART, *SD* RT SART = standard deviation of reaction times in the SART, TUT = task unrelated thoughts, WMC = working memory capacity, * $p < .05$; ** $p < .01$, *** $p < .001$.

3.2.2. The Worst Performance Rule with Unstandardized Coefficients (Covariances)

On the level of unstandardized coefficients, the baseline multilevel model indicated a significant interaction between trial number and WMC, $b = -4.46$, $t(496) = -6.53$, $p < .001$ (Table S33 in the Supplementary Materials). The worst performance interaction revealed a medium effect size of η^2part = 0.08. There were significant interactions between the factor task and the worst performance effect (interaction with arrow-flanker task: $b = 1.31$ $t(182674) = 5.14$, $p < .001$; no interaction with letter-flanker task: $b = 0.17$, $t(182657) = 0.61$, $p = .543$; interaction with number-stroop task: $b = 1.18$, $t(182711) = 6.28$, $p < .001$), suggesting that the strength of the WPR varied between tasks. Separate follow-up analyses for each of the four tasks revealed that a significant worst performance interaction was present in each of the four tasks (all *p*s < .001).

After controlling for individual differences in attentional lapses, we still observed a significant two-way interaction between trial number and WMC in the baseline model, $b = -3.44$, $t(496) = -5.20$, $p < .001$ (Figure 7 left side, Table S34 in the Supplementary Materials). The significant three-way interaction between WMC, trial number, and the control factor in the full model indicated a small but significant change of the worst performance pattern after controlling for attentional lapses, $b = 0.94$, $t(365374) = 5.07$, $p < .001$ (Table S35 in the Supplementary Materials). Also, model comparison revealed a significantly better fit for the full model with the three-way interaction in comparison to a model without the three-way interaction, $\Delta_{AIC} = 38$. Effect size estimation found a very small effect, η^2part = 0.00. We found no effects of the task on the three-way interaction, which indicates that the influence of TUTs on the worst performance pattern was comparable for all tasks (all four-way interactions were not significant, all *p*s > .192). Taken together, these results indicate that TUTs accounted for a small part of the worst performance pattern in multilevel models with unstandardized coefficients.

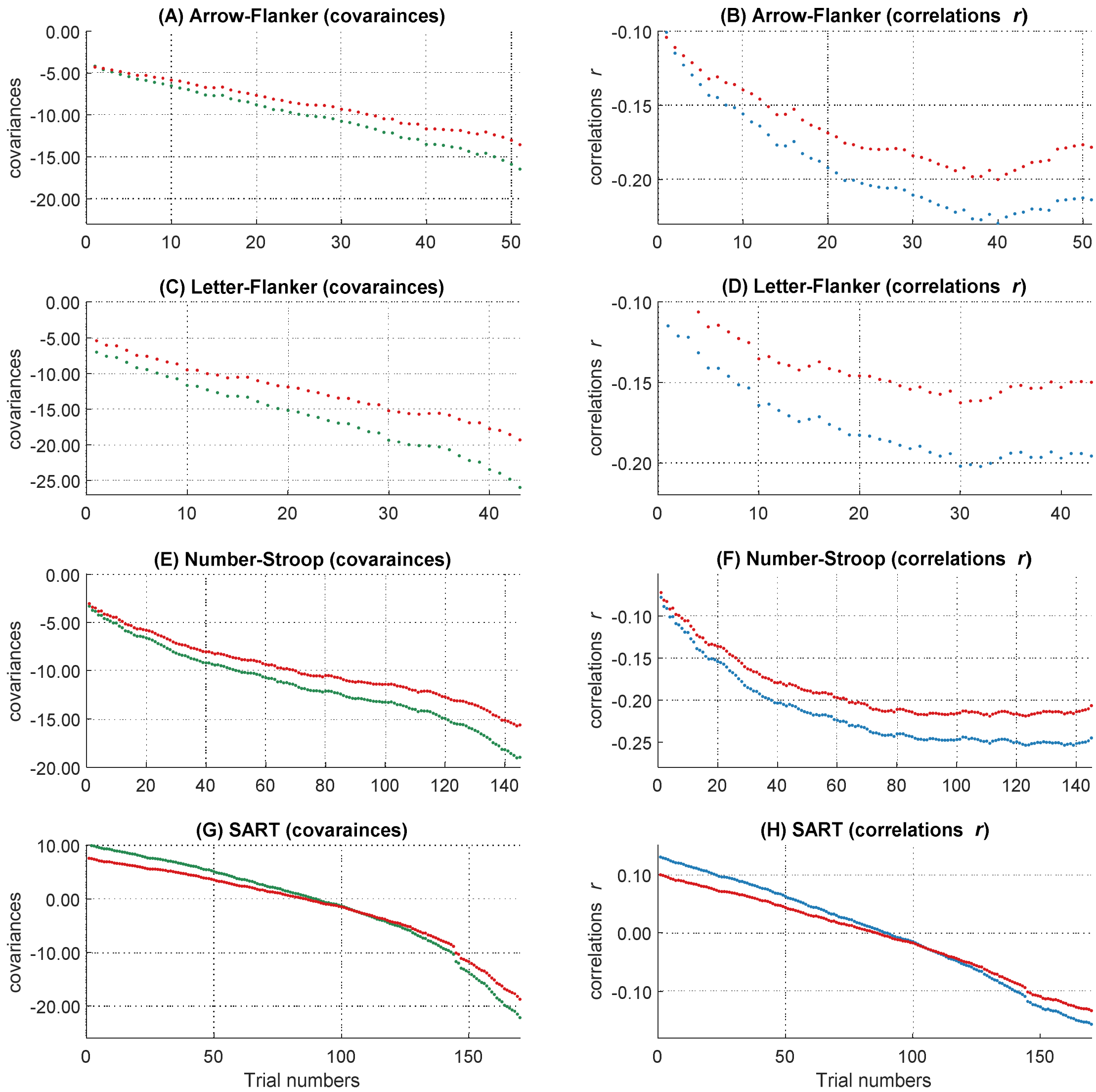


Figure 7. Course of the covariances and correlations over the RT distributions before and after controlling for the influence of the attentional lapses covariates. The courses of the covariances in the four different tasks are shown on the left side (**A**,**C**,**E**,**G**). The courses of the correlations in the four different tasks are shown on the right side (**B**,**D**,**F**,**H**). The figure describes the worst performance pattern before (green and blue lines) and after (red lines) the TUTs covariate were partialized out of the covariance.

3.2.3. The Worst Performance Rule with Standardized Coefficients (Correlations)

On the level of standardized coefficients, the baseline multilevel model indicated a significant interaction between trial number and WMC, $b = -0.04$, $t(499) = -5.13$, $p < .001$ (Table S36 in the Supplementary Materials). The worst performance interaction revealed a small effect size of η^2part = 0.05. Again, we observed interactions between the task factor and the WPR (interaction with arrow-flanker task: $b = 0.01$, $t(182643) = 3.28$, $p = .001$; interaction with letter-flanker task: $b = 0.02$, $t(182.633) = 5.41$, $p < .001$; no interaction with number-stroop task: $b = 0.00$, $t(182687) = 1.79$, $p = .074$) but baseline models for all tasks showed significant worst performance interactions (all *ps* < .017).

After controlling for individual differences in attentional lapses, we still observed a significant two-way interaction between trial number and WMC in the baseline model, $b = -0.03$, $t(499) = -4.05$, $p < .001$ (Figure 7 right side, Table S37 in the Supplementary Materials). The significant three-way interaction between WMC, trial number, and the control factor in the full model indicated a small but significant change of the worst performance pattern after controlling for attentional lapses, b = 0.01, t(365373) = 3.42, p = .001 (Table S38 in the Supplementary Materials). Also, model comparison revealed a significantly better fit for the full model with the three-way interaction in comparison to a model without the three-way interaction, Δ_{AIC} = 19. Effect size estimation revealed an effect close to zero, η^2part = 0.00. We found no effects of the task factor on the tree-way interaction, which indicates that the influence of TUTs on the worst performance pattern was comparable for all tasks (all four-way interactions were non-significant, all ps >.538). Taken together, these results indicate that TUTs accounted for very small parts of the worst performance pattern in the multilevel models with standardized coefficients (i.e., the WPR on the correlational level).

3.3. Discussion

The results of Study 2 substantiated the main results of Study 1 that attentional lapses can explain the increasing magnitude of covariation of the WPR to a significant degree. The large sample size and the greater homogeneity of the sample (students; mean age = 18.92, SD = 1.91) are the main characteristics different from the Study 1 sample. In Study 2, we found a significant WPR in our multilevel models, both on the level of covariances as well as on the level of correlations. We found a larger effect of attentional lapses on the WPR on the level of covariances than on the level of correlational analyses. This confirms the choice of our strategy to examine the WPR on both levels and suggests that attentional lapses contribute not only to the relation between RTs and cognitive abilities, but also to the variance in RTs, which is independent of cognitive abilities. As in Study 1, the single measure of self-reported attentional lapses explained only a small part of the WPR. The WPR remained significant after controlling for TUTs, independent of whether we analyzed covariances or correlations. We therefore conclude that TUTs as the sole measurement of attentional lapses explain a small part of the worst performance pattern and substantial parts of the WPR remain unexplained.

Taken together, we found significant worst performance patterns in the data and replicated our multilevel model findings of Study 1 in a large and age-homogenous sample. As already known from former findings by Kane et al. (2016) and Welhaf et al. (2020), the relations between all variables (TUTs, WMC, RTs) were consistent with previous research and our predictions. Self-reported attentional lapses, measured as TUTs, explained some significant—albeit very small—part of the WPR.

4. General Discussion

We analyzed two independent data sets and found support for Larson and Alderton's (1990) idea that attentional lapses can explain parts of the worst performance pattern (Larson and Alderton 1990). According to our results, the contribution of attentional lapses to the WPR varied for each of the covariates and the effects of the single covariates appeared to be very small, which in turn led to a small but significant reduction of the WPR. Considering the multiverse structure of attentional lapses, we combined different covariates and examined their common influence on the WPR. The influence of self-reported attentional lapses and an objective attentional lapses indicator together led to a full explanation of the phenomenon. In Study 1, we found a significant reduction of the worst performance pattern in covariances and a significant decrease of the worst performance slope in correlations. To address statistical power issues and to replicate our findings, we applied the same analysis strategy in a larger independent student sample in Study 2. The results of this replication study were in line with our former findings and also statistically significant on both levels. Taken together, we found evidence for the attentional lapses account, which claims that the origin of the WPR is based on inter-individual differences in the experience of attentional lapses.

Across both studies, we found that controlling for attentional lapses affected the WPR more strongly on the level of covariances than on the level of correlations. This result has important theoretical implications, because it indicates that the occurrence of attentional lapses affects the inter-individual variance in the right tail of the RT distribution. In other words, inter-individual differences in attentional lapses affected the amount of between-subject variability in the right tail of the RT distribution and could thus account for a large part of the WPR on the level of covariances. On the level of correlations, however, they only accounted for a small part of the WPR, because here the WPR was calculated based on standardized measures (i.e., controlled for between-subject variability in RTs). The idea that between-subject variability may differ across RT bands is not new (see Coyle 2003; Larson and Alderton 1990). The present study demonstrates that these differences in between-subject variability across RT bands are not merely a statistical artifact, but substantially related to individual differences in elementary attentional processes.

However, there is an alternative and simpler mathematical explanation that could account for the different results on the level of covariances and correlations. We found that RTs in faster and in slower trials are highly correlated. In consequence, it is plausible that fast responses are nearly proportional to slow responses. Furthermore, the nature of slower RTs is that their variance is larger in comparison to faster responses. Consequentially, we would assume that individual differences in RTs would fan out and the variance of individual differences become larger in slower RTs. Given that the intelligence score of each individual remains the same while the RT variance increases over the RT distribution, the covariance between intelligence and RTs grows monotonically larger towards slower RTs. In contrast, correlations would not necessarily increase in the same pattern, because they are standardized. Considering this pure mathematical explanation of the different results in covariances and correlations, one could either conclude that covariances are more sensitive than correlations or that correlations are more reliable than covariances.[1]

Our results are in line with Coyle's (2003) claim that the WPR is not driven by outlier or extreme values. Depending on the task, we extracted a certain number of trials out of the middle of participants' RT distributions. Additionally, we applied a careful intra- and inter-individual outlier analysis. In both studies, we found a robust increase of the magnitude in covariances that is consistent with the WPR. Moreover, we found a significant WPR effect on the standardized/correlational level in Study 2. In contrast, we did not find this significant worst performance pattern in the correlations in Study 1. Possible reasons for this may be the already discussed low statistical power and small sample size. However, we clearly observed a similar course of correlations over the RT distribution in both studies (see Figures 3 and 6). Notably, several previous studies used a descriptive approach for specifying the WPR. Although a test of significance is certainly warranted to test the existence of the WPR against chance (see Frischkorn et al. 2016), it is not uncommon to rely on descriptive evidence for the investigation of the WPR.

Effect sizes of the moderating role of the attentional lapses covariates on the WPR were small. Some of these estimates were η^2part < 0.01, especially in the analyses with standardized coefficients, which should be interpreted as very small effects. The reason why those small effects were significant is that those interaction terms were tested with a very large number of degrees of freedom, due to the trial-by-trial analyses and the repeated-measures design. As a consequence, the standard errors became very small and small *b*-weights reached the significance level more quickly. This may be considered as curse and blessing at the same time. On the one hand, we had enough power to detect small influences of attentional lapses on the WPR; on the other hand, statistical tests may have been overpowered, leading to the adoption of irrelevant effects as an explanation for the WPR. That is, the multilevel approach to the WPR is a powerful instrument that bears the risk of overpowering. An alternative approach could be to use Fisher's *Z*-test (e.g., Edwards 1976) as a more conservative method, which has less statistical power but requires a problematic two-stage estimation processes to assess the statistical significance of the WPR.

However, especially in study 2 some significant parts of the worst performance pattern remained unexplained after controlling for attentional lapses. It is important to conclude that some parts of the increasing magnitude in covariances and correlations between RTs and intelligence could not be explained by attentional lapses. There could be additional reasons for the origin of the WPR.

4.1. Alternative Accounts of the Worst Performance Rule

Beyond the attentional lapses account, there are two prominent alternative explanations of the WPR. They cannot be rule out as alternative explanations by our findings. To some degree these accounts are additional explanations for the remaining unexplained parts of the worst performance patterns and to some other degree they complement each other and can even be transferred into each other.

The *drift diffusion model account* claims that inter-individual differences in the evidence accumulation process could explain the WPR (Ratcliff et al. 2008). The drift diffusion model is a mathematical model that describes binary decision making as a random walk process through which evidence is accumulated until one of two decision thresholds is reached (Ratcliff 1978). The basic diffusion model consists of four parameters, namely the drift rate, which describes the strength and direction of the evidence accumulation process, the boundary separation, which describes how much information needs to be accumulated before a decision is being made, the starting point, which describes biases in decision making, and the non-decision time, which encompasses the time needed for all non-decisional processes such as encoding and response execution. The drift rate parameter in particular has been repeatedly shown to be associated with individual differences in mental abilities, working memory capacity, and intelligence (Ratcliff et al. 2010, 2011; Schmiedek et al. 2007; Schubert et al. 2015). More intelligent individuals show higher drift rates across several tasks (Schmiedek et al. 2007; Schubert et al. 2015, 2016). In their simulation study, Ratcliff et al. (2008) showed that the drift rate parameter of the diffusion model is more negatively related to slower quantiles compared to faster quantiles of the RT distribution, which means that the drift rate parameter and its underlying processes were better described by slower compared to faster RTs. The drift rate parameter is typically considered as a measure of the speed of information uptake. Hence, it is possible that the speed of information uptake is more validly measured in slower responses, which in turn would lead to higher negative correlations between RT and intelligence in slower than in faster responses. The higher validity of slower responses for the speed of information uptake could be an alternative explanation of the WPR. In other words, one could say that individual differences in the speed of evidence accumulation (measured by drift rates) may also account for the pattern of the WPR, as they give rise to individual differences in slowest RTs and are also strongly related to individual differences in cognitive abilities. However, drift rates are likely affected by a number of lower-level cognitive processes that may also include attentional processes. The drift diffusion model account of the WPR is not necessarily irreconcilable with the attentional lapses account. In this sense, it is also possible that attentional lapses are related to differences in the evidence accumulation process (see also Boehm et al. 2021).

Another explanation of the WPR focuses on its statistical characteristics (Sorjonen et al. 2020, 2021). With simulated data, Sorjonen et al. showed that the WPR is a special case of the *correlation of sorted scores rule* (Sorjonen et al. 2020, 2021). This rule states that the correlation between a sorted measure of performance (e.g., binned mean RTs or trial-wise sorted RTs) and intelligence will depend on the direction of the correlation between the variability in performance (e.g., intra-individual standard deviation in RTs) and intelligence. Because of the negative correlation between intra-individual standard deviation in RTs and intelligence, the rule predicts the emergence of the WPR. If there were a positive correlation between intra-individual variability in the respective performance measure and intelligence, the rule would instead predict a best performance rule. It is well-established that more intelligent individuals show a smaller standard deviation in RTs (Doebler and Scheffler 2016), which was also the case in our sample. We found negative correlations between the variance in RTs and cognitive abilities, $r = -.30$, $p = .003$, in Study 1, and from $r = -.20$ to $r = -.25$, all ps < .001, in Study 2. Hence, the WPR could also be (statistically) accounted for by the correlation of sorted scores rule. In turn, the correlation of sorted scores rule does not rule out the attentional lapses account of the WPR, because it is possible that the larger intra-individual RT variability in individuals with lower cognitive abilities results as the consequence of their more frequent experience of attentional lapses.

4.2. The Curious Course in Very Slow RTs

A novel and surprising finding in this study was the observed decrease in the magnitude of negative correlations and the simultaneous accelerated increase in the magnitude of negative covariances, respectively, in the slowest 15 percent of the responses (Figures 3 and 6). Apparently, some unknown process unrelated to intelligence increased the variance in RTs in the right tail of the RT distribution, which puts the WPR in a different light. Our observations are consistent with the meta-analysis of Schubert (2019), who described a logarithmic trend of the increases in the magnitude of negative correlations. This meta-analysis found that the increases in the magnitude of negative correlations is largest from the fastest to the mean performances and flattens from the mean to the slowest performances. Because of this observation, it was suggested to rename the WPR as *the not-best performance rule*, which is arguably a more appropriate name for this phenomenon. Welhaf et al. (2020)[2] replicated the not-best performance rule. With our trial-by-trial analyses, it was possible to draw a more detailed picture of this phenomenon and we found Schubert's (2019) observed logarithmic trend of correlations over the RT bins. There was an unexpected decline in the negative correlations in the slowest trials. Surprisingly, the increase in covariances accelerated at the same time. Based on these observations, we can conclude that some unknown process unrelated to cognitive abilities gave rise to RT variance in the slowest responses. The observed decline in correlations is also consistent with many previous studies that revealed a decrease or stagnation in the magnitude of the negative correlations in the slowest RT bins (Fernandez et al. 2014; Ratcliff et al. 2010; Salthouse 1998; Saville et al. 2016; Schmitz et al. 2018). Taken together, it seems that our observation is not an isolated case but a replicable phenomenon. Further studies may address the reasons for this conundrum.

5. Conclusions

Taken together, our results support the attentional lapses account of the WPR. Using multilevel models, we demonstrated that different single measures of attentional lapses accounted for some parts of the increasing magnitude in covariances and correlations between intelligence and RTs from the fastest to the slowest responses. The combined influence of several self-reported and objective attentional lapses measures accounted fully for this phenomenon, which in turn underlines the multiverse nature of the attentional lapses construct. Our results suggested that the WPR is caused by inter-individual differences in attentional lapses. Thus, it seems that individual differences in attentional control processes are an important factor contributing to individual differences in cognitive abilities.

Supplementary Materials: The following figure and tables are available via the Open Science Framework (https://osf.io/5pafg/, accessed on 23 December 2021), Figure S1: EEG Electrode assembly; Tables S1–S38: Results of multilevel analyses.

Author Contributions: Conceptualization, C.L., G.T.F., J.R. and A.-L.S.; methodology, C.L., G.T.F. and A.-L.S.; software, C.L. and G.T.F.; formal analysis, C.L.; investigation, C.L.; resources, A.-L.S., J.R. and D.H.; data curation, C.L.; writing—original draft preparation, C.L.; writing—review and editing C.L., G.T.F., J.R., D.H. and A.-L.S.; visualization, C.L.; supervision, A.-L.S., J.R. and D.H.; project administration, C.L. and A.-L.S.; funding acquisition, A.-L.S. and J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Excellence Initiative of the German Research Foundation (DFG) (grant number ZUK 49/Ü 5.2.178).

Institutional Review Board Statement: The present study was conducted according to the guidelines of the Declaration of Helsinki and approved by the ethics committee of behavioral and cultural studies of Heidelberg University.

Informed Consent Statement: At the beginning of each experimental session, participants signed an informed consent.

Data Availability Statement: The preprocessed data supporting the findings of Study 1 and the code for the statistical analysis used in this manuscript are available via the Open Science Framework (https://osf.io/5pafg/, accessed on 23 December 2021). Access to raw data of Study 1 will be granted upon request. The data supporting the findings of Study 2 are available via the Open Science Framework (https://osf.io/9qcmx/, accessed on 5 February 2021).

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

Notes

1 Many thanks to an anonymous reviewer for this suggestion.

2 We used the same data used by Welhaf et al. (2020) in Study 2.

References

Akaike, Hirotogu. 1998. Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike*. Edited by Emanuel Parzen, Kunio Tanabe and Genshiro Kitagawa. New York: Springer, pp. 199–213. [CrossRef]

Allison, Brendan Z., and John Polich. 2008. Workload Assessment of Computer Gaming Using a Single-Stimulus Event-Related Potential Paradigm. *Biological Psychology* 77: 277–83. [CrossRef]

Anderson, Thomas, Rotem Petranker, Hause Lin, and Norman A. S. Farb. 2021. The Metronome Response Task for Measuring Mind Wandering: Replication Attempt and Extension of Three Studies by Seli et al. *Attention, Perception, & Psychophysics* 83: 315–30. [CrossRef]

Arnau, Stefan, Christoph Löffler, Jan Rummel, Dirk Hagemann, Edmund Wascher, and Anna-Lena Schubert. 2020. Inter-Trial Alpha Power Indicates Mind Wandering. *Psychophysiology* 57: e13581. [CrossRef] [PubMed]

Arnicane, Andra, Klaus Oberauer, and Alessandra S. Souza. 2021. Validity of Attention Self-Reports in Younger and Older Adults. *Cognition* 206: 104482. [CrossRef]

Atchley, Rachel, Daniel Klee, and Barry Oken. 2017. EEG Frequency Changes Prior to Making Errors in an Easy Stroop Task. *Frontiers in Human Neuroscience* 11: 521. [CrossRef]

Baird, Benjamin, Jonathan Smallwood, Antoine Lutz, and Jonathan W. Schooler. 2014. The Decoupled Mind: Mind-Wandering Disrupts Cortical Phase-Locking to Perceptual Events. *Journal of Cognitive Neuroscience* 26: 2596–607. [CrossRef] [PubMed]

Baldwin, Carryl L., Daniel M. Roberts, Daniela Barragan, John D. Lee, Neil Lerner, and James S. Higgins. 2017. Detecting and Quantifying Mind Wandering during Simulated Driving. *Frontiers in Human Neuroscience* 11: 406. [CrossRef]

Barron, Evelyn, Leigh M. Riby, Joanna Greer, and Jonathan Smallwood. 2011. Absorbed in Thought: The Effect of Mind Wandering on the Processing of Relevant and Irrelevant Events. *Psychological Science* 22: 596–601. [CrossRef] [PubMed]

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using Lme4. *Journal of Statistical Software* 67: 1–48. [CrossRef]

Baumeister, Alfred, and George Kellas. 1968. Reaction Time and Mental Retardation. In *International Review of Research in Mental Retardation*, 3rd ed. New York and London: Academic Press, pp. 163–93.

Ben-Shachar, Mattan, Daniel Lüdecke, and Dominique Makowski. 2020. Effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software* 5: 2815. [CrossRef]

Berger, Hans. 1929. Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten* 87: 527–70. [CrossRef]

Boehm, Udo, Maarten Marsman, Han L. J. van der Maas, and Gunter Maris. 2021. An Attention-Based Diffusion Model for Psychometric Analyses. *Psychometrika* 86: 938–72. [CrossRef]

Braboszcz, Claire, and Arnaud Delorme. 2011. Lost in Thoughts: Neural Markers of Low Alertness during Mind Wandering. *NeuroImage* 54: 3040–47. [CrossRef] [PubMed]

Burnham, Kenneth P., and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer. [CrossRef]

Carriere, Jonathan S. A., Paul Seli, and Daniel Smilek. 2013. Wandering in Both Mind and Body: Individual Differences in Mind Wandering and Inattention Predict Fidgeting. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale* 67: 19–31. [CrossRef]

Cavanagh, James F., and Michael J. Frank. 2014. Frontal Theta as a Mechanism for Cognitive Control. *Trends in Cognitive Sciences* 18: 414–21. [CrossRef]

Cavanagh, James F., Laura Zambrano-Vazquez, and John J. B. Allen. 2012. Theta Lingua Franca: A Common Mid-frontal Substrate for Action Monitoring Processes. *Psychophysiology* 49: 220–38. [CrossRef]

Compton, Rebecca J., Dylan Gearinger, and Hannah Wild. 2019. The Wandering Mind Oscillates: EEG Alpha Power Is Enhanced during Moments of Mind-Wandering. *Cognitive, Affective, & Behavioral Neuroscience* 19: 1184–91. [CrossRef]

Conway, Andrew R. A., Nelson Cowan, Michael F. Bunting, David J. Therriault, and Scott R. B. Minkoff. 2002. A Latent Variable Analysis of Working Memory Capacity, Short-Term Memory Capacity, Processing Speed, and General Fluid Intelligence. *Intelligence* 30: 163–83. [CrossRef]

Cooper, Nicholas R., Rodney J. Croft, Samuel J. J. Dominey, Adrian P. Burgess, and John H. Gruzelier. 2003. Paradox Lost? Exploring the Role of Alpha Oscillations during Externally vs. Internally Directed Attention and the Implications for Idling and Inhibition Hypotheses. *International Journal of Psychophysiology* 47: 65–74. [CrossRef]

Coyle, Thomas R. 2001. IQ Is Related to the Worst Performance Rule in a Memory Task Involving Children. *Intelligence* 29: 117–29. [CrossRef]

Coyle, Thomas R. 2003. A Review of the Worst Performance Rule: Evidence, Theory, and Alternative Hypotheses. *Intelligence* 31: 567–87. [CrossRef]

Delorme, Arnaud, and Scott Makeig. 2004. EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. *Journal of Neuroscience Methods* 134: 9–21. [CrossRef]

Diascro, Matthew N., and Nathan Brody. 1993. Serial versus Parallel Processing in Visual Search Tasks and IQ. *Personality and Individual Differences* 14: 243–45. [CrossRef]

Doebler, Philipp, and Barbara Scheffler. 2016. The Relationship of Choice Reaction Time Variability and Intelligence: A Meta-Analysis. *Learning and Individual Differences* 52: 157–66. [CrossRef]

Dutilh, Gilles, Joachim Vandekerckhove, Alexander Ly, Dora Matzke, Andreas Pedroni, Renato Frey, Jörg Rieskamp, and Eric-Jan Wagenmakers. 2017. A Test of the Diffusion Model Explanation for the Worst Performance Rule Using Preregistration and Blinding. *Attention, Perception, & Psychophysics* 79: 713–25. [CrossRef]

Edwards, Allen L. 1976. *An Introduction to Linear Regression and Correlation*. San Francisco: W. H. Freeman.

Fernandez, Sébastien, Delphine Fagot, Judith Dirk, and Anik de Ribaupierre. 2014. Generalization of the Worst Performance Rule across the Lifespan. *Intelligence* 42: 31–43. [CrossRef]

Frank, David J., Brent Nara, Michela Zavagnin, Dayna R. Touron, and Michael J. Kane. 2015. Validating Older Adults' Reports of Less Mind-Wandering: An Examination of Eye Movements and Dispositional Influences. *Psychology and Aging* 30: 266–78. [CrossRef] [PubMed]

Frischkorn, Gidon, Anna-Lena Schubert, Andreas Neubauer, and Dirk Hagemann. 2016. The Worst Performance Rule as Moderation: New Methods for Worst Performance Analysis. *Journal of Intelligence* 4: 9. [CrossRef]

Frischkorn, Gidon, Anna-Lena Schubert, and Dirk Hagemann. 2019. Processing Speed, Working Memory, and Executive Functions: Independent or Inter-Related Predictors of General Intelligence. *Intelligence* 75: 95–110. [CrossRef]

Gelman, Andrew, and Hal Stern. 2006. The Difference Between 'Significant' and 'Not Significant' Is Not Itself Statistically Significant. *The American Statistician* 60: 328–31. [CrossRef]

Hanel, Paul H. P., and Katia C. Vione. 2016. Do Student Samples Provide an Accurate Estimate of the General Public?" Edited by Martin Voracek. *PLoS ONE* 11: e0168354. [CrossRef]

Hanslmayr, Simon, Joachim Gross, Wolfgang Klimesch, and Kimron L. Shapiro. 2011. The Role of Alpha Oscillations in Temporal Attention. *Brain Research Reviews* 67: 331–43. [CrossRef] [PubMed]

Jäger, Adolf O. 1984. Intelligenzstrukturforschung: Konkurrierende Modelle, Neue Entwicklungen, Perspektiven. [Structural Research on Intelligence: Competing Models, New Developments, Perspectives]. *Psychologische Rundschau* 35: 21–35.

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

Jensen, Arthur R. 1992. The Importance of Intraindividual Variation in Reaction Time. *Personality and Individual Differences* 13: 869–81. [CrossRef]

Kam, Julia W. Y., and Todd C. Handy. 2013. The Neurocognitive Consequences of the Wandering Mind: A Mechanistic Account of Sensory-Motor Decoupling. *Frontiers in Psychology* 4: 725. [CrossRef]

Kam, Julia W. Y., Elizabeth Dao, James Farley, Kevin Fitzpatrick, Jonathan Smallwood, Jonathan W. Schooler, and Todd C. Handy. 2011. Slow Fluctuations in Attentional Control of Sensory Cortex. *Journal of Cognitive Neuroscience* 23: 460–70. [CrossRef]

Kane, Michael J., David Z. Hambrick, Stephen W. Tuholski, Oliver Wilhelm, Tabitha W. Payne, and Randall W. Engle. 2004. The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General* 133: 189–217. [CrossRef] [PubMed]

Kane, Michael J., David Z. Hambrick, and Andrew R. A. Conway. 2005. Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131: 66–71. [CrossRef]

Kane, Michael J., Andrew R. A. Conway, David Z. Hambrick, and Randall W. Engle. 2008. Variation in Working Memory Capacity as Variation in Executive Attention and Control. In *Variation in Working Memory*. Edited by Andrew Conway, Chris Jarrold, Michael Kane, Akira Miyake and John Towse. Oxford: Oxford University Press, pp. 21–48. [CrossRef]

Kane, Michael J., Matt E. Meier, Bridget A. Smeekens, Georgina M. Gross, Charlotte A. Chun, Paul J. Silvia, and Thomas R. Kwapil. 2016. Individual Differences in the Executive Control of Attention, Memory, and Thought, and Their Associations with Schizotypy. *Journal of Experimental Psychology. General* 145: 1017–48. [CrossRef]

Kleiner, Mario, David Brainard, Denis Pelli, A. Ingling, R. Murray, and C. Broussard. 2007. What's New in Psychtoolbox-3. *Perception* 36: 1–16.

Kok, Albert. 2001. On the Utility of P3 Amplitude as a Measure of Processing Capacity. *Psychophysiology* 38: 557–77. [CrossRef] [PubMed]
Kranzler, John H. 1992. A Test of Larson and Alderton's (1990) Worst Performance Rule of Reaction Time Variability. *Personality and Individual Differences* 13: 255–61. [CrossRef]
Krawietz, Sabine A., Andrea K. Tamplin, and Gabriel A. Radvansky. 2012. Aging and Mind Wandering during Text Comprehension. *Psychology and Aging* 27: 951–58. [CrossRef] [PubMed]
Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. LmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82: 1–26. [CrossRef]
Kyllonen, Patrick C., and Raymond E. Christal. 1990. Reasoning Ability Is (Little More than) Working-Memory Capacity?!". *Intelligence* 14: 389–433. [CrossRef]
Larson, Gerald E., and David L. Alderton. 1990. Reaction Time Variability and Intelligence: A 'Worst Performance' Analysis of Individual Differences. *Intelligence* 14: 309–25. [CrossRef]
Leite, Fábio P. 2009. Should IQ, Perceptual Speed, or Both Be Used to Explain Response Time?". *The American Journal of Psychology* 122: 517–26.
Leszczynski, Marcin, Leila Chaieb, Thomas P. Reber, Marlene Derner, Nikolai Axmacher, and Juergen Fell. 2017. Mind Wandering Simultaneously Prolongs Reactions and Promotes Creative Incubation. *Scientific Reports* 7: 10197. [CrossRef]
Lopez-Calderon, Javier, and Steven J. Luck. 2014. ERPLAB: An Open-Source Toolbox for the Analysis of Event-Related Potentials. *Frontiers in Human Neuroscience* 8: 213. [CrossRef]
Maillet, David, and Daniel L. Schacter. 2016. From Mind Wandering to Involuntary Retrieval: Age-Related Differences in Spontaneous Cognitive Processes. *Neuropsychologia* 80: 142–56. [CrossRef]
Maillet, David, Roger E. Beaty, Megan L. Jordano, Dayna R. Touron, Areeba Adnan, Paul J. Silvia, Thomas R. Kwapil, Gary R. Turner, R. Nathan Spreng, and Michael J. Kane. 2018. Age-Related Differences in Mind-Wandering in Daily Life. *Psychology and Aging* 33: 643–53. [CrossRef]
Maillet, David, Lujia Yu, Lynn Hasher, and Cheryl L. Grady. 2020. Age-Related Differences in the Impact of Mind-Wandering and Visual Distraction on Performance in a Go/No-Go Task. *Psychology and Aging* 35: 627–38. [CrossRef] [PubMed]
McVay, Jennifer C., and Michael J. Kane. 2009. Conducting the Train of Thought: Working Memory Capacity, Goal Neglect, and Mind Wandering in an Executive-Control Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35: 196–204. [CrossRef] [PubMed]
McVay, Jennifer C., and Michael J. Kane. 2010. Does Mind Wandering Reflect Executive Function or Executive Failure? Comment on Smallwood and Schooler (2006) and Watkins (2008). *Psychological Bulletin* 136: 188–97. [CrossRef]
McVay, Jennifer C., and Michael J. Kane. 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: 525–49. [CrossRef] [PubMed]
Miller, Edward M. 1994. Intelligence and Brain Myelination: A Hypothesis. *Personality and Individual Differences* 17: 803–32. [CrossRef]
Mognon, Andrea, Jorge Jovicich, Lorenzo Bruzzone, and Marco Buiatti. 2011. ADJUST: An Automatic EEG Artifact Detector Based on the Joint Use of Spatial and Temporal Features. *Psychophysiology* 48: 229–40. [CrossRef]
Mrazek, Michael D., Dawa T. Phillips, Michael S. Franklin, James M. Broadway, and Jonathan W. Schooler. 2013. Young and Restless: Validation of the Mind-Wandering Questionnaire (MWQ) Reveals Disruptive Impact of Mind-Wandering for Youth. *Frontiers in Psychology* 4: 560. [CrossRef] [PubMed]
Nash J.C., and R. Varadhan. Unifying Optimization Algorithms to Aid Software System Users: Optimx for R. *Journal of Statistical Software* 43: 1–14. Available online: http://www.jstatsoft.org/v43/i09/ (accessed on 23 December 2021).
O'Connell, Redmond G., Paul M. Dockree, Ian H. Robertson, Mark A. Bellgrove, John J. Foxe, and Simon P. Kelly. 2009. Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal of Neuroscience* 29: 8604–11. [CrossRef]
Oberauer, Klaus, Ralf Schulze, Oliver Wilhelm, and Heinz-Martin Süß. 2005. Working Memory and Intelligence–Their Correlation and Their Relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* 131: 61–65. [CrossRef] [PubMed]
Polich, John. 2007. Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 118: 2128–48. [CrossRef]
R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
Rammsayer, Thomas, and Stefan Troche. 2016. Validity of the Worst Performance Rule as a Function of Task Complexity and Psychometric g: On the Crucial Role of g Saturation. *Journal of Intelligence* 4: 5. [CrossRef]
Randall, Jason G., Frederick L. Oswald, and Margaret E. Beier. 2014. Mind-Wandering, Cognition, and Performance: A Theory-Driven Meta-Analysis of Attention Regulation. *Psychological Bulletin* 140: 1411–31. [CrossRef]
Ratcliff, Roger. 1978. A Theory of Memory Retrieval. *Psychological Review* 85: 59–108. [CrossRef]
Ratcliff, Roger, Florian Schmiedek, and Gail McKoon. 2008. A Diffusion Model Explanation of the Worst Performance Rule for Reaction Time and IQ. *Intelligence* 36: 10–17. [CrossRef] [PubMed]
Ratcliff, Roger, Anjali Thapar, and Gail McKoon. 2010. Individual Differences, Aging, and IQ in Two-Choice Tasks. *Cognitive Psychology* 60: 127–57. [CrossRef]

Ratcliff, Roger, Anjali Thapar, and Gail McKoon. 2011. The Effects of Aging and IQ on Item and Associative Memory. *Journal of Experimental Psychology. General* 140: 464–87. [CrossRef]

Revelle, W. 2020. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. R Package Version 2.0.12. Evanston: Northwestern University, Available online: https://CRAN.R-project.org/package=psych (accessed on 16 March 2021).

Robison, Matthew K., Ashley L. Miller, and Nash Unsworth. 2020. A Multi-Faceted Approach to Understanding Individual Differences in Mind-Wandering. *Cognition* 198: 104078. [CrossRef] [PubMed]

Salthouse, Timothy A. 1993. Attentional Blocks Are Not Responsible for Age-Related Slowing. *Journal of Gerontology* 48: P263–P270. [CrossRef]

Salthouse, Timothy A. 1998. Relation of Successive Percentiles of Reaction Time Distributions to Cognitive Variables and Adult Age. *Intelligence* 26: 153–66. [CrossRef]

Saville, Christopher W. N., Kevin D. O. Beckles, Catherine A. MacLeod, Bernd Feige, Monica Biscaldi, André Beauducel, and Christoph Klein. 2016. A Neural Analogue of the Worst Performance Rule: Insights from Single-Trial Event-Related Potentials. *Intelligence* 55: 95–103. [CrossRef]

Scharfen, Jana, Judith Marie Peters, and Heinz Holling. 2018. Retest Effects in Cognitive Ability Tests: A Meta-Analysis. *Intelligence* 67: 44–66. [CrossRef]

Schmiedek, Florian, Klaus Oberauer, Oliver Wilhelm, Heinz-Martin Süß, and Werner W. Wittmann. 2007. Individual Differences in Components of Reaction Time Distributions and Their Relations to Working Memory and Intelligence. *Journal of Experimental Psychology: General* 136: 414–29. [CrossRef] [PubMed]

Schmitz, Florian, and Oliver Wilhelm. 2016. Modeling Mental Speed: Decomposing Response Time Distributions in Elementary Cognitive Tasks and Correlations with Working Memory Capacity and Fluid Intelligence. *Journal of Intelligence* 4: 13. [CrossRef]

Schmitz, Florian, Dominik Rotter, and Oliver Wilhelm. 2018. Scoring Alternatives for Mental Speed Tests: Measurement Issues and Validity for Working Memory Capacity and the Attentional Blink Effect. *Journal of Intelligence* 6: 47. [CrossRef] [PubMed]

Schönbrodt, Felix D., and Marco Perugini. 2013. At What Sample Size Do Correlations Stabilize? " *Journal of Research in Personality* 47: 609–12. [CrossRef]

Schubert, Anna-Lena. 2019. A Meta-Analysis of the Worst Performance Rule. *Intelligence* 73: 88–100. [CrossRef]

Schubert, Anna-Lena, Dirk Hagemann, Andreas Voss, Andrea Schankin, and Katharina Bergmann. 2015. Decomposing the Relationship between Mental Speed and Mental Abilities. *Intelligence* 51: 28–46. [CrossRef]

Schubert, Anna-Lena, Gidon T. Frischkorn, Dirk Hagemann, and Andreas Voss. 2016. Trait Characteristics of Diffusion Model Parameters. *Journal of Intelligence* 4: 7. [CrossRef]

Schubert, Anna-Lena, Gidon T. Frischkorn, and Jan Rummel. 2020. The Validity of the Online Thought-Probing Procedure of Mind Wandering Is Not Threatened by Variations of Probe Rate and Probe Framing. *Psychological Research* 84: 1846–56. [CrossRef]

Seli, Paul, James Allan Cheyne, and Daniel Smilek. 2013. Wandering Minds and Wavering Rhythms: Linking Mind Wandering and Behavioral Variability. *Journal of Experimental Psychology: Human Perception and Performance* 39: 1–5. [CrossRef] [PubMed]

Seli, Paul, Jonathan S. A. Carriere, David R. Thomson, James Allan Cheyne, Kaylena A. Ehgoetz Martens, and Daniel Smilek. 2014. Restless Mind, Restless Body. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40: 660–68. [CrossRef] [PubMed]

Sheppard, Leah D., and Philip A. Vernon. 2008. Intelligence and Speed of Information-Processing: A Review of 50 Years of Research. *Personality and Individual Differences* 44: 535–51. [CrossRef]

Shipstead, Zach, Tyler L. Harrison, and Randall W. Engle. 2016. Working Memory Capacity and Fluid Intelligence: Maintenance and Disengagement. *Perspectives on Psychological Science* 11: 771–99. [CrossRef] [PubMed]

Skrondal, Anders, and Petter Laake. 2001. Regression among Factor Scores. *Psychometrika* 66: 563–75. [CrossRef]

Smallwood, Jonathan, and Jonathan W. Schooler. 2006. The Restless Mind. *Psychological Bulletin* 132: 946–58. [CrossRef]

Smallwood, Jonathan, and Jonathan W. Schooler. 2015. The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology* 66: 487–518. [CrossRef]

Smallwood, Jonathan, Merrill McSpadden, and Jonathan W. Schooler. 2007. The Lights Are on but No One's Home: Meta-Awareness and the Decoupling of Attention When the Mind Wanders. *Psychonomic Bulletin & Review* 14: 527–33. [CrossRef]

Smallwood, Jonathan, Emily Beach, Jonathan W. Schooler, and Todd C. Handy. 2008. Going AWOL in the Brain: Mind Wandering Reduces Cortical Analysis of External Events. *Journal of Cognitive Neuroscience* 20: 458–69. [CrossRef]

Sorjonen, Kimmo, Guy Madison, Bo Melin, and Fredrik Ullén. 2020. The Correlation of Sorted Scores Rule. *Intelligence* 80: 101454. [CrossRef]

Sorjonen, Kimmo, Guy Madison, Tomas Hemmingsson, Bo Melin, and Fredrik Ullén. 2021. Further Evidence That the Worst Performance Rule Is a Special Case of the Correlation of Sorted Scores Rule. *Intelligence* 84: 101516. [CrossRef]

Steindorf, Lena, and Jan Rummel. 2020. Do Your Eyes Give You Away? A Validation Study of Eye-Movement Measures Used as Indicators for Mindless Reading. *Behavior Research Methods* 52: 162–76. [CrossRef]

Sudevan, Padmanabhan, and David A. Taylor. 1987. The Cuing and Priming of Cognitive Operations. *Journal of Experimental Psychology: Human Perception and Performance* 13: 89–103. [CrossRef]

Thomson, David R., Paul Seli, Derek Besner, and Daniel Smilek. 2014. On the Link between Mind Wandering and Task Performance over Time. *Consciousness and Cognition* 27: 14–26. [CrossRef] [PubMed]

Thut, Gregor, Annika Nietzel, Stephan A. Brandt, and Alvaro Pascual-Leone. 2006. α-Band Electroencephalographic Activity over Occipital Cortex Indexes Visuospatial Attention Bias and Predicts Visual Target Detection. *Journal of Neuroscience* 26: 9494–502. [CrossRef] [PubMed]

Unsworth, Nash, Thomas S. Redick, Chad E. Lakey, and Diana L. Young. 2010. Lapses in Sustained Attention and Their Relation to Executive Control and Fluid Abilities: An Individual Differences Investigation. *Intelligence* 38: 111–22. [CrossRef]

Verhaeghen, Paul, and Timothy A. Salthouse. 1997. Meta-Analyses of Age–Cognition Relations in Adulthood: Estimates of Linear and Nonlinear Age Effects and Structural Models. *Psychological Bulletin* 122: 231–49. [CrossRef]

Verleger, Rolf. 2020. Effects of Relevance and Response Frequency on P3b Amplitudes: Review of Findings and Comparison of Hypotheses about the Process Reflected by P3b. *Psychophysiology* 57: e13542. [CrossRef]

Watkins, Edward R. 2008. Constructive and Unconstructive Repetitive Thought. *Psychological Bulletin* 134: 163–206. [CrossRef] [PubMed]

Weinstein, Yana. 2018. Mind-Wandering, How Do I Measure Thee with Probes? Let Me Count the Ways. *Behavior Research Methods* 50: 642–61. [CrossRef]

Welhaf, Matthew S., Bridget A. Smeekens, Matt E. Meier, Paul J. Silvia, Thomas R. Kwapil, and Michael J. Kane. 2020. The Worst Performance Rule, or the Not-Best Performance Rule? Latent-Variable Analyses of Working Memory Capacity, Mind-Wandering Propensity, and Reaction Time. *Journal of Intelligence* 8: 25. [CrossRef] [PubMed]

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy DAgostino McGowan, Romain François, and Garrett Grolemund. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4: 1686. [CrossRef]

Supplementary material

Figure S1

Electrode assembly with relevant channels for the different psychophysiological attentional lapses covariates

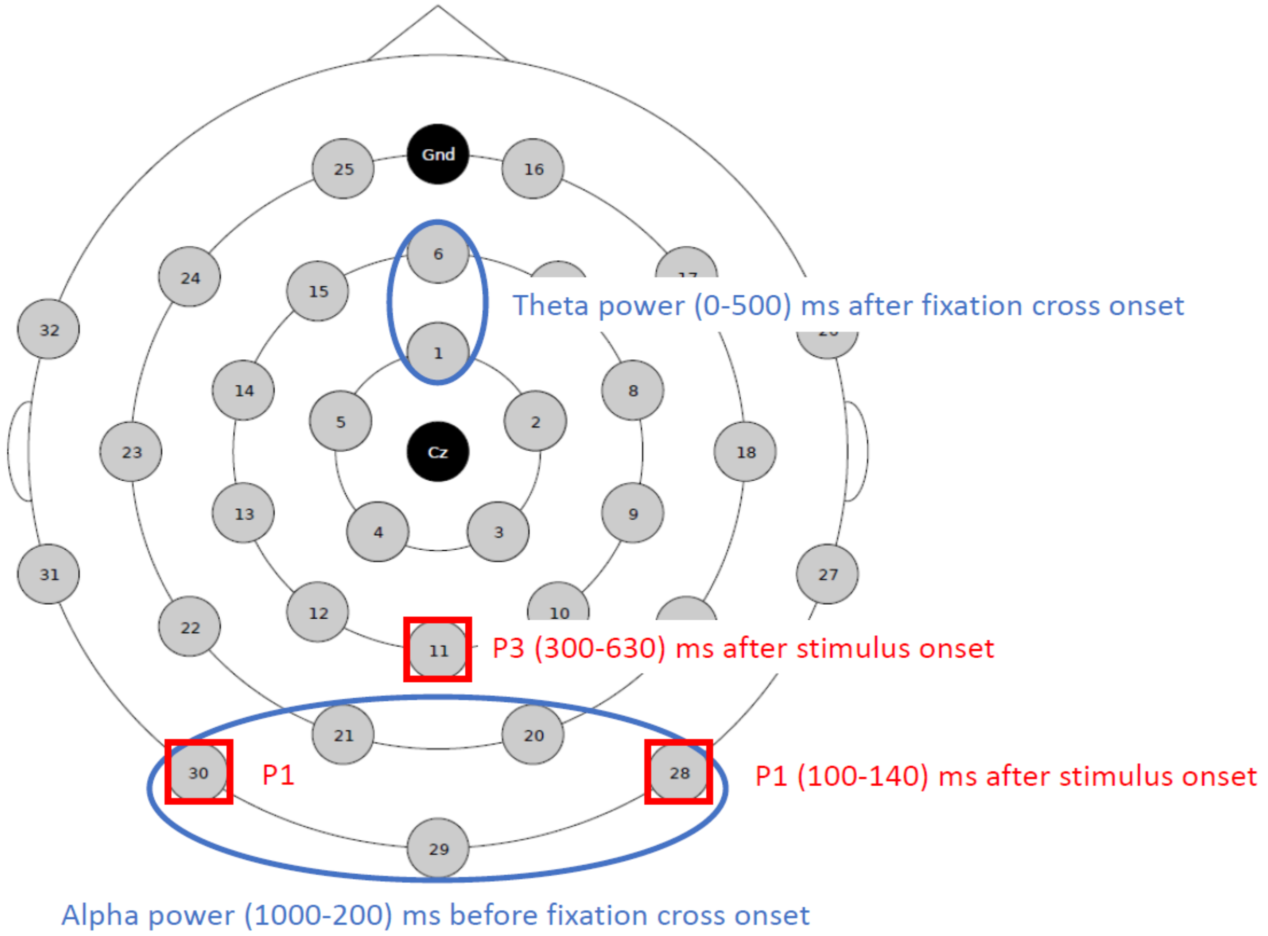


Table S1

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (TUTs + Q-SMW + MRT)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.34)	85	0.00	141.42	>.999
intelligence	-29.08 (15.43)	85	-1.88		.063
trial number	0.00 (4.86)	85	0.00	44.76	>.999
trial number x intelligence = WPR	-8.88 (4.88)	85	-1.82		.073

Note. N = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S2

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (TUTs)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.85)	85	0.00	146.11	>.999
intelligence	-41.55 (15.94)	85	-2.61		.011
trial number	0.00 (5.06)	85	0.00	46.66	>.999
trial number x intelligence = WPR	-12.98 (5.09)	85	-2.55		.013

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S3

Full multilevel model, which tests the effect of attentional lapses covariate (TUTs) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	835.82 (15.84)	85	52.77	146.01	< .001
intelligence	-44.18 (15.93)	85	-2.77		.007
trial number	146.99 (5.08)	85	28.92	46.85	< .001
control	-835.82 (0.18)	57630	-4577.19		< .001
trial number x intelligence = WPR	-14.93 (5.11)	85	-2.92		.004
intelligence x control	2.64 (0.18)	57630	14.35		<.001
trial number x control	-146.99 (0.16)	57630	-929.50		< .001
trial number x intelligence x control	1.95 (0.16)	57630	12.24		<.001

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of covariances.

Table S4

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (Q-SMW)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.82)	85	0.00	145.89	>.999
intelligence	-42.61 (15.92)	85	-2.68		.009
trial number	0.00 (5.20)	85	0.00	47.97	>.999
trial number x intelligence = WPR	-14.73 (5.23)	85	-2.81		.006

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S5

Full multilevel model, which tests the effect of attentional lapses covariate (Q-SMW) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	835.82 (15.83)	85	52.81	145.92	< .001
intelligence	-44.18 (15.92)	85	-2.76		.007
trial number	146.99 (5.20)	85	28.26	47.95	< .001
control	-835.82 (0.17)	57630	-4882.83		< .001
trial number x intelligence = WPR	-14.93 (5.23)	85	-2.85		.005
intelligence x control	1.58 (0.17)	57630	9.15		<.001
trial number x control	-146.99 (0.15)	57630	-991.57		< .001
trial number x intelligence x control	0.20 (0.15)	57630	1.34		.179

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of covariances.

Table S6

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (MRT)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.41)	85	0.00	142.08	>.999
intelligence	-31.55 (15.50)	85	-2.04		.045
trial number	0.00 (5.03)	85	0.00	46.33	>.999
trial number x intelligence = WPR	-10.57 (5.06)	85	-2.09		.039

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S7

Full multilevel model, which tests the effect of attentional lapses covariate (MRT) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	835.82 (15.46)	85	54.08	142.50	< .001
intelligence	-44.18 (15.55)	85	-2.84		.006
trial number	146.99 (5.05)	85	29.13	46.49	< .001
control	-835.82 (0.26)	57630	-3274.10		< .001
trial number x intelligence = WPR	-14.93 (5.08)	85	-2.94		.004
intelligence x control	12.63 (0.26)	57630	49.18		<.001
trial number x control	-146.99 (0.22)	57630	-664.88		< .001
trial number x intelligence x control	4.36 (0.22)	57630	19.60		<.001

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of covariances.

Table S8

Baseline multilevel model of the WPR on an unstandardized level – electrophysiological subsample

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.97)	84	52.37	146.40	<.001
intelligence	-46.95 (16.07)	84	-2.92		.004
trial number	146.91 (5.26)	84	27.95	48.17	<.001
trial number x intelligence = WPR	-15.21 (5.29)	84	-2.88		.005

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S9

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (ERPs = P1 + P3)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.86)	84	0.00	145.40	>.999
intelligence	-45.91 (15.96)	84	-2.88		.005
trial number	0.00 (5.25)	84	0.00	48.10	>.999
trial number x intelligence = WPR	-14.99 (5.28)	84	-2.84		.006

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S10

Full multilevel model, which tests the effect of attentional lapses covariate (ERPs = P1 + P3) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.88)	84	52.69	145.57	< .001
intelligence	-46.95 (15.98)	84	-2.94		.004
trial number	146.91 (5.25)	84	27.98	48.10	< .001
control	-836.54 (0.18)	56952	-4725.41		< .001
trial number x intelligence = WPR	-15.21 (5.28)	84	-2.88		.005
intelligence x control	1.04 (0.18)	56952	5.84		<.001
trial number x control	-146.91 (0.15)	56952	-958.24		< .001
trial number x intelligence x control	0.22 (0.15)	56952	1.42		.156

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of covariances.

Table S11

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (Power = Alpha +Theta)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.56)	84	0.00	142.57	>.999
intelligence	-40.69 (15.65)	84	-2.60		.011
trial number	0.00 (5.14)	84	0.00	47.07	>.999
trial number x intelligence = WPR	-13.14 (5.17)	84	-2.55		.013

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S12

Full multilevel model, which tests the effect of attentional lapses covariate (Power = Alpha +Theta) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.61)	84	53.58	145.57	< .001
intelligence	-46.95 (15.71)	84	-2.99		.004
trial number	146.91 (5.15)	84	28.51	48.10	< .001
control	-836.54 (0.24)	56952	-3525.50		< .001
trial number x intelligence = WPR	-15.21 (5.18)	84	-2.93		.004
intelligence x control	6.26 (0.24)	56952	26.34		<.001
trial number x control	-146.91 (0.21)	56952	-714.91		< .001
trial number x intelligence x control	2.06 (0.21)	56952	9.98		<.001

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of covariances.

Table S13

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (Alpha power)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.71)	84	0.00	143.98	>.999
intelligence	-46.11 (15.80)	84	-2.92		.005
trial number	0.00 (5.19)	84	0.00	47.57	>.999
trial number x intelligence = WPR	-14.96 (5.22)	84	-2.87		.005

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S14

Full multilevel model, which tests the effect of attentional lapses covariate (Alpha power) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.77)	84	53.05	144.52	< .001
intelligence	-46.95 (15.86)	84	-2.96		.004
trial number	146.91 (5.21)	84	28.22	47.71	< .001
control	-836.54 (0.20)	56952	-4211.03		< .001
trial number x intelligence = WPR	-15.21 (5.24)	84	-2.90		.005
intelligence x control	0.84 (0.20)	56952	4.21		<.001
trial number x control	-146.91 (0.17)	56952	-853.93		< .001
trial number x intelligence x control	0.24 (0.17)	56952	1.41		.159

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of covariances.

Table S15

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (Theta power)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.85)	84	0.00	143.98	>.999
intelligence	-41.85 (15.94)	84	-2.63		.010
trial number	0.00 (5.21)	84	0.00	47.57	>.999
trial number x intelligence = WPR	-13.48 (5.24)	84	-2.57		.012

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S16

Full multilevel model, which tests the effect of attentional lapses covariate (Theta power) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.84)	84	52.81	145.57	< .001
intelligence	-46.95 (15.94)	84	-2.95		.004
trial number	146.91 (5.21)	84	28.20	48.10	< .001
control	-836.54 (0.20)	56952	-4197.36		< .001
trial number x intelligence = WPR	-15.21 (5.24)	84	-2.91		.005
intelligence x control	5.08 (0.20)	56952	25.42		<.001
trial number x control	-146.91 (0.17)	56952	-851.15		< .001
trial number x intelligence x control	1.72 (0.17)	56952	9.73		<.001

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of covariances.

Table S17

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (all significant covariates)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (15.20)	84	0.00	139.30	>.999
intelligence	-27.73 (15.29)	84	-1.81		.073
trial number	0.00 (4.86)	84	0.00	44.49	>.999
trial number x intelligence = WPR	-7.76 (4.88)	84	-1.59		.116

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariation according to the WPR.

Table S18

Full multilevel model, which tests the effect of attentional lapses covariate (all significant covariates) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	836.54 (15.30)	84	54.69	145.57	< .001
intelligence	-46.95 (15.39)	84	-3.05		.003
trial number	146.91 (4.93)	84	29.82	48.10	< .001
control	-836.54 (0.30)	56952	-2807.62		< .001
trial number x intelligence = WPR	-15.21 (4.96)	84	-3.07		.003
intelligence x control	19.22 (0.30)	56952	64.13		<.001
trial number x control	-146.91 (0.26)	56952	-569.40		< .001
trial number x intelligence x control	7.45 (0.26)	56952	28.68		<.001

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of covariances.

Table S19

Baseline multilevel model of the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.93	>.999
intelligence	-0.27 (0.10)	85	-2.70		.008
trial number	0.00 (0.02)	85	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.10		.276

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S20

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (TUTs + Q-SMW + MRT)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.95	>.999
intelligence	-0.19 (0.10)	85	-1.86		.067
trial number	-0.00 (0.02)	85	0.00	0.17	>.999
trial number x intelligence = WPR	-0.01 (0.02)	85	-0.52		.605

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S21

Full multilevel model, which tests the effect of attentional lapses covariates (TUTs + Q-SMW + MRT) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.93	>.999
intelligence	-0.27 (0.10)	85	-2.70		.008
trial number	0.00 (0.02)	85	0.00	0.17	> .999
control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.14		.258
intelligence x control	0.08 (0.00)	57630	52.74		<.001
trial number x control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence x control	0.01 (0.00)	57630	8.70		<.001

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of correlations.

Table S22

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (TUTs)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.94	>.999
intelligence	-0.26 (0.10)	85	-2.56		.012
trial number	-0.00 (0.02)	85	0.00	0.17	>.999
trial number x intelligence = WPR	-0.01 (0.02)	85	-0.75		.456

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S23

Full multilevel model, which tests the effect of attentional lapses covariates (TUTs) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.93	>.999
intelligence	-0.27 (0.10)	85	-2.69		.009
trial number	0.00 (0.02)	85	0.00	0.17	> .999
control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.14		.258
intelligence x control	0.01 (0.00)	57630	12.98		<.001
trial number x control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence x control	0.01 (0.00)	57630	9.49		<.001

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of correlations.

Table S24

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (Q-SMW)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.92	>.999
intelligence	-0.26 (0.10)	85	-2.60		.011
trial number	-0.00 (0.02)	85	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.15		.253

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S25

Full multilevel model, which tests the effect of attentional lapses covariates (Q-SMW) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.93	>.999
intelligence	-0.27 (0.10)	85	-2.70		.008
trial number	0.00 (0.02)	85	0.00	0.18	> .999
control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.10		.276
intelligence x control	0.01 (0.00)	57630	9.58		<.001
trial number x control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence x control	0.00 (0.00)	57630	1.39		.165

Note. N = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of correlations.

Table S26

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (MRT)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.95	>.999
intelligence	-0.20 (0.10)	85	-1.98		.051
trial number	-0.00 (0.02)	85	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-0.84		.402

Note. N = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S27

Full multilevel model, which tests the effect of attentional lapses covariates (MRT) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	85	0.00	0.93	>.999
intelligence	-0.27 (0.10)	85	-2.70		.008
trial number	0.00 (0.02)	85	0.00	0.18	> .999
control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	85	-1.08		.284
intelligence x control	0.07 (0.00)	57630	48.05		<.001
trial number x control	0.00 (0.00)	57630	0.00		>.999
trial number x intelligence x control	0.00 (0.00)	57630	3.47		>.001

Note. *N* = 85. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariate on the WPR on the level of correlations.

Table S28

Baseline multilevel model of the WPR on a standardized level – electrophysiological subsample

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	84	0.00	0.93	>.999
intelligence	-0.29 (0.10)	84	-2.86		.005
trial number	0.00 (0.02)	84	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	84	-0.92		.359

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S29

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (ERPs = P1 + P3)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	84	0.00	0.93	>.999
intelligence	-0.29 (0.10)	84	-2.82		.006
trial number	-0.00 (0.02)	84	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	84	-0.93		.355

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S30

Full multilevel model, which tests the effect of attentional lapses covariates (ERPs = P1 + P3) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	84	0.00	0.92	>.999
intelligence	-0.29 (0.10)	84	-2.87		.005
trial number	0.00 (0.02)	84	0.00	0.18	> .999
control	0.00 (0.00)	56952	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	84	-0.92		.360
intelligence x control	0.00 (0.00)	56952	4.61		<.001
trial number x control	0.00 (0.00)	56952	0.00		>.999
trial number x intelligence x control	-0.00 (0.00)	56952	-0.32		.749

Note. *N* = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of correlations.

Table S31

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (Power = Alpha +Theta)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	84	0.00	0.93	>.999
intelligence	-0.26 (0.10)	84	-2.55		.013
trial number	-0.00 (0.02)	84	0.00	0.18	>.999
trial number x intelligence = WPR	-0.02 (0.02)	84	-0.83		.410

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S32

Full multilevel model, which tests the effect of attentional lapses covariates (Power = Alpha +Theta) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.10)	84	0.00	0.92	>.999
intelligence	-0.29 (0.10)	84	-2.88		.005
trial number	0.00 (0.02)	84	0.00	0.18	> .999
control	0.00 (0.00)	56952	0.00		>.999
trial number x intelligence = WPR	-0.02 (0.02)	84	-0.91		.366
intelligence x control	0.03 (0.00)	56952	22.72		<.001
trial number x control	0.00 (0.00)	56952	0.00		>.999
trial number x intelligence x control	0.00 (0.00)	56952	1.17		.243

Note. N = 84. 340 trials of each participant were included for analysis. Data were centered to the trial with the sorted number of 170 and afterwards rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariates. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the covariates on the WPR on the level of correlations.

Table S33

Baseline multilevel model of the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	503.16 (2.20)	485	228.99	48.22	<.001
WMC	-9.11 (2.17)	485	-4.19		<.001
trial number	43.73 (0.69)	495	63.40	14.88	<.001
Arrow-Flanker task (AF)	-41.46 (0.29)	182692	-143.77		<.001
Letter-Flanker task (LF)	29.92 (0.33)	88356	90.45		<.001
Stroop task (Stroop)	4.58 (0.21)	180117	21.51		<.001
trial number x WMC = WPR	-4.46 (0.68)	496	-6.53		<.001
WMC x AF	-1.77 (0.29)	180963	-6.10		<.001
WMC x LF	-5.16 (0.32)	181798	-15.95		<.001
WMC x Stroop	-1.94 (0.21)	182097	-9.12		<.001
trial number x AF	-9.69 (0.25)	182525	-38.10		<.001
trial number x LF	6.06 (0.29)	180382	21.22		<.001
trial number x Stroop	3.85 (0.19)	182711	20.64		<.001
WMC x trial number x AF	1.31 (0.26)	182674	5.14		<.001
WMC x trial number x LF	0.17 (0.29)	182657	0.61		.543
WMC x trial number x Stroop	1.18 (0.19)	182711	6.28		<.001

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariances according to the WPR.

Table S34

Baseline multilevel model of the WPR on an unstandardized level without the influence of attentional lapses (TUTs)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.04 (2.19)	485	0.02	48.13	.985
WMC	-7.48 (2.17)	485	-3.45		.001
trial number	0.21 (0.67)	495	0.31	14.43	.757
Arrow-Flanker task (AF)	0.54 (0.29)	182692	1.90		.058
Letter-Flanker task (LF)	0.74 (0.33)	87371	2.24		.025
Stroop task (Stroop)	-0.79 (0.21)	180039	-3.75		<.001
trial number x WMC = WPR	-3.44 (0.66)	496	-5.20		<.001
WMC x AF	-2.02 (0.39)	180907	-7.00		<.001
WMC x LF	-3.28 (0.32)	181765	-10.29		<.001
WMC x Stroop	-1.91 (0.21)	182074	-9.01		<.001
trial number x AF	0.08 (0.25)	182528	0.30		.765
trial number x LF	0.41 (0.28)	180593	1.45		.147
trial number x Stroop	0.23 (0.19)	182711	-1.25		.213
WMC x trial number x AF	0.92 (0.25)	182674	3.62		<.001
WMC x trial number x LF	0.39 (0.28)	182658	1.38		.167
WMC x trial number x Stroop	0.91 (0.19)	182712	4.86		<.001

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in covariances according to the WPR.

Table S35

Full multilevel model, which tests the effect of attentional lapses covariate (TUTs) on the WPR on an unstandardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	503.08 (2.19)	487	229.28	48.15	<.001
WMC	-9.06 (2.17)	487	-4.17		<.001
trial number	43.69 (0.68)	508	64.63	14.58	<.001
Arrow-Flanker task (AF)	-41.42 (0.29)	365486	-144.47		<.001
Letter-Flanker task (LF)	30.03 (0.33)	111402	91.50		<.001
Stroop task (Stroop)	4.57 (0.21)	351029	21.76		<.001
control	-502.96 (0.21)	365374	-2388.27		<.001
trial number x WMC = WPR	-4.42 (0.70)	510	-6.62		<.001
WMC x AF	-1.81 (0.29)	355178	-6.28		<.001
WMC x LF	-5.19 (0.32)	359263	-16.21		<.001
WMC x Stroop	-1.96 (0.21)	360978	-9.04		<.001
trial number x AF	-9.64 (0.25)	365686	-38.12		<.001
trial number x LF	6.09 (0.28)	356047	21.50		<.001
trial number x Stroop	3.81 (0.19)	365886	20.61		<.001
WMC x control	1.53 (0.21)	365374	7.25		< .001
trial number x control	-43.44 (0.19)	365374	-233.38		<.001
AF x control	41.93 (0.40)	365374	103.84		<.001
LF x control	-29.40 (0.45)	365374	-65.56		<.001
Stroop x control	-5.38 (0.29)	365374	-18.30		<.001
WMC x trial number x AF	1.28 (0.25)	365875	5.05		<.001
WMC x trial number x LF	0.16 (0.28)	365852	0.55		.579
WMC x trial number x Stroop	1.21 (0.19)	365878	6.54		<.001
WMC x trial number x control	0.94 (0.19)	365374	5.07		<.001
WMC x AF x control	-0.17 (0.40)	365374	-0.43		.669
WMC x LF x control	1.95 (0.50)	365374	4.43		<.001
WMC x Stroop x control	0.07 (0.29)	365374	0.24		.809
trial number x AF x control	9.67 (0.36)	365374	27.12		<.001
trial number x LF x control	-5.70 (0.40)	365374	-14.36		<.001

trial number x Stroop x control	-4.01 (0.26)	365374	-15.43	<.001
WMC x trial number x AF x control	-0.33 (0.36)	365374	-0.93	.350
WMC x trial number x LF x control	0.25 (0.40)	365374	0.63	.532
WMC x trial number x Stroop x control	-0.34 (0.26)	365374	-1.30	.193

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the TUTs covariate on the WPR on the level of covariances.

Table S36

Baseline multilevel model of the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.03)	485	0.11	0.69	.914
WMC	-0.14 (0.03)	485	-4.42		<.001
trial number	0.00 (0.01)	498	0.51	0.17	.610
Arrow-Flanker task (AF)	0.01 (0.00)	182672	2.14		.032
Letter-Flanker task (LF)	0.01 (0.00)	69753	1.66		.096
Stroop task (Stroop)	-0.01 (0.00)	178200	-4.76		<.001
trial number x WMC = WPR	-0.04 (0.01)	499	-5.13		<.001
WMC x AF	-0.06 (0.00)	179560	-17.12		<.001
WMC x LF	-0.02 (0.00)	180951	-4.33		<.001
WMC x Stroop	-0.07 (0.00)	181481	-25.71		<.001
trial number x AF	-0.00 (0.00)	182521	-0.13		.901
trial number x LF	0.00 (0.00)	181309	1.13		.259
trial number x Stroop	-0.00 (0.00)	182677	-0.66		.510
WMC x trial number x AF	0.01 (0.00)	182643	3.28		.001
WMC x trial number x LF	0.02 (0.00)	182633	5.41		<.001
WMC x trial number x Stroop	0.00 (0.00)	182687	1.79		.074

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S37

Baseline multilevel model of the WPR on a standardized level without the influence of attentional lapses (TUTs)

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.03)	485	0.04	0.70	.968
WMC	-0.12 (0.03)	485	-3.71		<.001
trial number	0.00 (0.01)	498	0.43	0.16	.666
Arrow-Flanker task (AF)	0.01 (0.00)	182672	2.42		.015
Letter-Flanker task (LF)	0.01 (0.00)	69800	2.37		.018
Stroop task (Stroop)	-0.01 (0.00)	178206	-4.46		<.001
trial number x WMC = WPR	-0.03 (0.01)	499	-4.05		<.001
WMC x AF	-0.06 (0.00)	179565	-16.46		<.001
WMC x LF	-0.00 (0.00)	180954	-0.85		.396
WMC x Stroop	-0.07 (0.00)	181483	-23.39		<.001
trial number x AF	0.00 (0.00)	182526	0.18		.861
trial number x LF	0.01 (0.00)	181424	1.25		.212
trial number x Stroop	-0.00 (0.00)	182679	-0.95		.343
WMC x trial number x AF	0.01 (0.00)	182644	2.68		.007
WMC x trial number x LF	0.02 (0.00)	182637	4.58		<.001
WMC x trial number x Stroop	0.00 (0.00)	182687	1.18		.240

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. A significant interaction between *trial number* and *intelligence* represents a significant increase of the magnitude in correlations according to the WPR.

Table S38

Full multilevel model, which tests the effect of attentional lapses covariate (TUTs) on the WPR on a standardized level

RT on:	*b*-weight (standard error)	*df*	*t*-value	random effect *SD*	*p*
intercept	0.00 (0.03)	486	0.07	0.69	.942
WMC	-0.14 (0.03)	486	-4.39		<.001
trial number	0.00 (0.01)	516	0.49	0.17	.627
Arrow-Flanker task (AF)	0.01 (0.00)	365035	2.30		.022
Letter-Flanker task (LF)	0.01 (0.00)	83228	2.04		.042
Stroop task (Stroop)	0.01 (0.00)	342656	-4.65		<.001
control	0.00 (0.00)	365373	0.00		>.999
trial number x WMC = WPR	-0.04 (0.01)	517	-5.15		<.001
WMC x AF	-0.07 (0.00)	348743	-17.32		<.001
WMC x LF	-0.02 (0.00)	355053	-4.44		<.001
WMC x Stroop	-0.07 (0.00)	357739	-26.04		<.001
trial number x AF	0.00 (0.00)	365600	0.03		.977
trial number x LF	0.00 (0.00)	360311	1.22		.224
trial number x Stroop	-0.00 (0.00)	365855	-0.03		.407
WMC x control	0.02 (0.00)	365373	7.37		<.001
trial number x control	0.00 (0.00)	365373	0.00		>.999
AF x control	0.00 (0.01)	365373	0.00		>.999
LF x control	0.00 (0.01)	365373	0.00		>.999
Stroop x control	0.00 (0.00)	365373	0.00		>.999
WMC x trial number x AF	0.01 (0.00)	365834	3.17		.002
WMC x trial number x LF	0.02 (0.00)	365807	5.39		<.001
WMC x trial number x Stroop	0.01 (0.00)	365841	1.94		.053
WMC x trial number x control	0.01 (0.00)	287232	3.42		.001
WMC x AF x control	0.00 (0.01)	365373	0.66		.509
WMC x LF x control	0.02 (0.01)	365373	2.64		.008
WMC x Stroop x control	0.01 (0.00)	365373	1.86		.063
trial number x AF x control	0.00 (0.01)	365373	0.00		>.999
trial number x LF x control	0.00 (0.01)	365373	0.00		>.999

trial number x Stroop x control	0.00 (0.00)	365373	0.00	>.999
WMC x trial number x AF x control	-0.00 (0.01)	365373	-0.27	.785
WMC x trial number x LF x control	-0.00 (0.01)	365373	-0.54	.591
WMC x trial number x Stroop x control	-0.00 (0.00)	365373	-0.61	.539

Note. Data were centered to the central trial within every task and afterwards the trial numbers were rescaled between -2 and 2. *Control* is a dummy coded factor, which represents raw RTs or RTs residualized by the corresponding attentional lapses covariate. A significant three-way interaction between *trial number*, *intelligence* and *control* represents a moderating influence of the TUTs covariate on the WPR on the level of correlations.

Appendix Manuscript II

Psychological Research (2024) 88:1092–1114
https://doi.org/10.1007/s00426-023-01924-7

RESEARCH

The common factor of executive functions measures nothing but speed of information uptake

Christoph Löffler[1,2] · Gidon T. Frischkorn[3] · Dirk Hagemann[1] · Kathrin Sadus[1] · Anna-Lena Schubert[2]

Received: 2 June 2023 / Accepted: 27 December 2023 / Published online: 19 February 2024

Abstract
There is an ongoing debate about the unity and diversity of executive functions and their relationship with other cognitive abilities such as processing speed, working memory capacity, and intelligence. Specifically, the initially proposed unity and diversity of executive functions is challenged by discussions about (1) the factorial structure of executive functions and (2) unfavorable psychometric properties of measures of executive functions. The present study addressed two methodological limitations of previous work that may explain conflicting results: The inconsistent use of (a) accuracy-based vs. reaction time-based indicators and (b) average performance vs. difference scores. In a sample of 148 participants who completed a battery of executive function tasks, we tried to replicate the three-factor model of the three commonly distinguished executive functions shifting, updating, and inhibition by adopting data-analytical choices of previous work. After addressing the identified methodological limitations using drift–diffusion modeling, we only found one common factor of executive functions that was fully accounted for by individual differences in the speed of information uptake. No variance specific to executive functions remained. Our results suggest that individual differences common to all executive function tasks measure nothing more than individual differences in the speed of information uptake. We therefore suggest refraining from using typical executive function tasks to study substantial research questions, as these tasks are not valid for measuring individual differences in executive functions.

The common factor of executive function tasks measures nothing else but speed of information uptake

The umbrella term “executive functions” summarizes many top-down regulated abilities known under several synonyms, such as executive control, cognitive control, attentional control, and executive attention (Rey-Mermet et al., 2019). The most popular model of executive functions proposed by Miyake et al. (2000) includes three of these abilities: Shifting describes one’s ability to shift attention between different tasks or different mental sets; updating describes one’s ability to monitor memory contents and store new contents to the memory; inhibition describes one’s ability to block irrelevant information or interferences from the attentional focus (Friedman & Miyake, 2017; Friedman et al., 2008; Miyake & Friedman, 2012; Miyake et al., 2000; Rey-Mermet et al., 2018). This three-factor model has become the predominant model for describing and separating executive functions. Although the selection of abilities classified as executive functions by Miyake et al. (2000) was not exhaustive, it was based on practical and neuroanatomical considerations, as each of the three executive functions is associated with specific areas of the neocortex.

Several theoretical accounts of processes underlying individual differences in cognitive abilities claim that differences in executive functions determine differences in higher order cognitive abilities (Kane et al., 2008; Kovacs & Conway, 2016). Moreover, there is a large body of empirical findings reporting correlations between higher order cognitive processes (intelligence and working memory capacity [WMC]) and executive functions (e.g., Friedman et al., 2006, 2008, 2011).

To measure executive functions, researchers usually contrast two experimental conditions, i.e., one *condition with*

✉ Christoph Löffler
christoph.loeffler@uni-mainz.de

1 Institute of Psychology, Heidelberg University, Heidelberg, Germany

2 Department of Psychology, University of Mainz, Mainz, Germany

3 Department of Psychology, University of Zurich, Zurich, Switzerland

lower processing demands and one *condition with greater processing demands*. For example, in an Arrow Flanker task (Eriksen & Eriksen, 1974), participants see an arrow in the center of a screen pointing to the left or right side. Participants have to indicate in which direction this arrow points. The arrow is shown amid further flanking stimuli. In the condition with lower processing demands (neutral condition), the central target arrow is surrounded by dashes not containing any directional or spatial information, which therefore have lower distracting effects on participants' performance. In the condition with greater processing demands, the flanking stimuli consist of arrows pointing in the opposite direction of the target arrow (incongruent condition). Participants' task is to ignore the irrelevant information of the flanker stimuli, which is more distracting in the condition with greater processing demands than in the neutral condition. This makes the decision in the condition with greater processing demands more difficult and leads to slower responses and higher error rates compared to the condition with lower processing demands. This decrease in performance between the two conditions, which are identical except for added demands on inhibition, indicates the specific strain on executive function demands (inhibitory processes). This strain occurs when participants have to ignore irrelevant flankers. In addition, performance in both conditions is also affected by *task-specific processes* as well as *task-general processes*. Following the logic of selective additivity, the performance decrement from the condition with less to the conditions with greater processing demands (e.g., the difference between reaction times [RTs] or accuracy rates) can be used to measure inhibitory demands, and individual differences in the performance decrement reflect individual differences in inhibition. Typical updating and shifting tasks are created following the same logic of selective additivity, allowing to analyze individual differences in the performance decrement. Because researchers have the choice between using RTs or accuracy rates as performance measures, there is much heterogeneity how individual differences in executive functions are assessed (von Bastian et al., 2020), even within single studies (Friedman et al., 2006, 2008; Himi et al., 2019, 2021; Ito et al., 2015; Krumm et al., 2009; Miyake et al., 2000; Schnitzspahn et al., 2013; Vaughan & Giovanello, 2010; Wongupparaj et al., 2015).

Empirical findings on the three-factor model of executive functions

The three executive functions introduced by Miyake et al. (2000) represent distinct but interrelated factors (Friedman et al., 2006, 2008; Himi et al., 2019, 2021; Ito et al., 2015; Miyake et al., 2000; Schnitzspahn et al., 2013; Vaughan & Giovanello, 2010). Substantial correlations between the three latent factors raised the question of a higher-order factor of executive functions, often labeled as common executive functions. Hence, Friedman et al., (2008, 2011) further developed the model of three distinct factors into a model with two distinct factors of shifting and updating and an additional common factor of executive functions (see also Himi et al., 2019). This common factor supposedly represents the "ability to maintain task goals and goal-related information" (Miyake & Friedman, 2012, p. 3), which is considered as a general ability required in all cognitive tasks.

Despite the seemingly robust findings on the three executive functions model, recent research questions this factor structure and casts doubt on the existence of meaningful individual differences in specific executive functions, in particular inhibition (Frischkorn et al., 2019; Hedge et al., 2018; Hull et al., 2008; Karr et al., 2018; Klauer et al., 2010; Krumm et al., 2009; Rey-Mermet et al., 2018, 2019; Rouder & Haaf, 2019; Stahl et al., 2014; von Bastian et al., 2020). A recently published review by Karr et al. (2018) reported that previous studies showed evidence for both unidimensional and multidimensional factor structures of executive functions in adults. Karr et al. (2018) reanalyzed data from nine adult samples with different types of model composition to evaluate which type of model best describes executive functions data. They compared unidimensional models, nested-factor models (a special kind of bi-factor models), two-factor models, and three-factor models. Karr et al. (2018) found that none of the different model compositions was clearly superior and could be selected as the best model describing executive functions, although the authors observed slightly more evidence for nested-factor models than for the other model types. They attributed these inconsistencies in the dimensionality of models to a publication bias for well-fitting but possibly non-replicating models with underpowered sample sizes (Karr et al., 2018). This review clearly demonstrated that the factorial structure of executive functions is still an open research question.

Previous research did not only focus on the factor structure across, but also within specific executive functions. In particular, there is a lot of research on the factor structure of inhibition, with many papers demonstrating that inhibition tasks do not form a coherent latent factor (e.g., Krumm et al., 2009; Rey-Mermet et al., 2018, 2019; Rouder & Haaf, 2019; Stahl et al., 2014). For example, Rey-Mermet et al. (2018) used a battery of 11 inhibition tasks to analyze correlations between RT-based performance decrements, but could not find a coherent pattern of correlations between the performances in the different inhibition tasks. Instead, they found that inhibition abilities formed two correlated factors, one that reflected inhibition of prepotent responses and another that reflected inhibition of distractor interferences. Also, in a follow-up study using accuracy-based scores to measure inhibition, Rey-Mermet et al. (2019) could not observe a coherent factor structure among inhibition tasks. Likewise, Krumm et al. (2009) tried to replicate the three-factor model of executive functions using tasks from Miyake et al. (2000) with RT- and accuracy-based dependent variables, but they did not find a latent factor of inhibition. These results are exemplary for further studies that failed to find a coherent factor of inhibition even after accounting for trial-to-trial measurement noise (Rouder & Haaf, 2019; Stahl et al., 2014).

Further research suggested that the shared variance of executive function tasks is mainly driven by task-general process demands and not by demands specific to executive functions. For example, Frischkorn et al. (2019) separated the variance of experimental manipulations in executive function tasks from the shared variance of task-general processes, which are required in nearly every task and not specific to the experimental manipulation. The authors used adapted versions of a shifting task (Sudevan & Taylor, 1987), of an N-Back task (Scharinger et al., 2015), and of an Attentional Network task (Fan et al., 2002), and found that manipulation-specific variance (reflecting added executive demands) barely contributed to performance in executive function tasks. Instead, task-general processing abilities captured the majority of variance in task performance. Hence, performance in executive function tasks reflected task-general cognitive processes instead of specific executive functions (Frischkorn et al., 2019). In sum, executive function tasks, especially inhibition tasks, hardly measure a coherent construct or individual differences specific to executive functions. Instead, individual differences in general processing abilities explain most of the variance in performance in executive function tasks.

These inconsistent findings pose a problem for individual difference research and theoretical frameworks of executive functions: If it is impossible to find coherent factors of executive functions, it is impossible to assess covariations between these factors and other psychological constructs. A current literature review on attentional control and executive functions suggested that these inconstancies regarding the factor structure of executive functions may result from the psychometric properties of performance measures generated from executive function tasks (see von Bastian et al., 2020).

The inconsistent use of dependent variables

There is much heterogeneity in how performance is assessed in executive function tasks (von Bastian et al., 2020). Usually, researchers use RT-based scores as measures in inhibition and shifting tasks, whereas they commonly use accuracy-based scores as measures in updating tasks (e.g., Friedman et al., 2006, 2008; Himi et al., 2019, 2021; Krumm et al., 2009; Miyake et al., 2000; Wongupparaj et al., 2015). We refer to such studies using different types of performance scores within their study designs as studies with *heterogeneous measurement scores.* In a recent review of 76 studies, von Bastian et al. (2020) showed that RT-based and accuracy-based scores were used more or less interchangeably to measure inhibition and shifting, whereas updating was typically assessed using accuracy-based scores. This inconsistent use of different types of performance scores can generate unexpected side effects because accuracy- and RT-based measures are often only weakly correlated, even in the same task (Hedge et al., 2018).

Furthermore, several studies measured individual differences in specific executive functions as difference scores, as the performance in the condition with higher task demands (e.g., RTs of the the incongruent condition in the Stroop task, Wongupparaj et al., 2015), or as the average performance over all conditions (e.g., averaged proportion correct across trials with different updating demands as updating scores; Miyake et al., 2000; for an overview, see also von Bastian et al., 2020). The issue with using either of the latter two measures is that other processes contribute to individual differences in performance in addition to the specific executive function demands. In particular, task-specific and task-general process parameters such as perceptual processing speed, the speed of decision-making, the speed of response preparation, and the speed of response execution contribute to individual differences in both condition-specific and task-general average performances. Consequently, using condition-specific or task-general average scores lowers the validity of the resulting measures if those are intended to only reflect specific executive functions. In consequence, correlations between these variables and other constructs do not necessarily reflect correlations between specific executive processes and other constructs but also of these other constructs with general performance parameters reflected in the measurement scores.

Difference scores: high validity or further psychometric concerns?

Despite the seemingly greater face validity of difference scores in comparison to condition-specific or task-average scores, voices have been cautioning against the blind use of difference scores for two reasons. First, the use of difference scores relies on the assumption that each individual cognitive process added to an experimental task is independent of other processes and that each process has an additive effect on the performance measure (i.e., RTs and accuracy rates). In the Arrow Flanker task, for example, subtracting the RTs of the condition with lower processing demands (neutral condition) from the condition with greater processing demands (incongruent condition) should isolate specific inhibitory demands from more general processing demands affecting performance in both conditions (Donders, 1869). However, the assumption of additive processes has been challenged. For example, Miller and Ulrich (2013) introduced a model demonstrating that different processes contributing to RTs do not act independently from each other, but interact with each other, which is contrary to the assumption of their additivity. In a certain task, the specific executive function processes may, for example, interact with general processing demands. Following their reasoning, the subtraction of RTs from two conditions does not purely isolate executive function processes, because the influence of the interaction between general processing demands and executive function demands also remains in the difference score (Miller & Ulrich, 2013).

Second, some researchers caution against using difference scores because they tend to show low reliabilities (Ackerman & Hambrick, 2020; Draheim et al., 2019, 2023; Hedge et al., 2018; Miller & Ulrich, 2013; von Bastian et al., 2020; Weigard et al., 2021). Von Bastian et al. (2020) summarized the reliabilities of 406 measures of executive functions and found that the difference scores of inhibition tasks showed particularly low reliabilities with a mean reliability of 0.63 and a range from close to zero to close to one, whereas the reliability for shifting difference scores and updating scores were markedly higher (with mean reliabilities of 0.78). Low reliabilities are problematic for individual differences research, because they limit the strength of correlations with other measures (Cronbach & Furby, 1970; Spearman, 1904). Taken together, these issues of validity and reliability suggest that difference scores may not yield psychometrically sound measures of executive functions.

Overcoming these problems

We summarized two problems of previous research measuring individual differences in executive functions, namely: (1) The inconsistent use of accuracy- and RT-based scoring methods, and (2) the psychometric problems of difference scores. Here we propose another analytical strategy to overcome these problems by combining cognitive modeling approaches with structural equation modeling. To address the first issue, we will use the drift rate parameter (v) of the diffusion model (Ratcliff, 1978), which is a mathematical model parameter that represents the speed of the evidence accumulation. For parameter estimation, the drift–diffusion model takes the distributions of correct as well as incorrect response times into account and thus integrates information about accuracies and RTs. To address the second issue, we will not control for general processing efficiency by controlling for performance in the condition with lower processing demands (by, e.g., calculating difference scores). Instead, we will control for general processing efficiency by using *elementary cognitive tasks*. The battery of three elementary cognitive tasks used in this study consists of three tasks with minimal executive demands often used in individual differences research (Frischkorn et al., 2019; Neubauer & Knorr, 1998; Schubert et al., 2015, 2017). We aim to use these tasks to measure individual differences in basic abilities of information processing largely free of executive demands, allowing us to control individual differences in performance in executive function tasks for individual differences in basic processing abilities (Frischkorn et al., 2019; Neubauer & Knorr, 1998; Schubert et al., 2015). This way, we overcome reliability problems of performance measures stemming from contrasting two conditions of the same task. These analytical choices will increase the likelihood of obtaining reliable and valid individual differences in executive functions.

Cognitive modeling to generate integrated measures of accuracies and RTs

To address the first problem—the inconsistent use of accuracies vs. RTs as indicator variables —, we used the drift parameter (v) of the drift–diffusion model to quantify participants' task performances. The drift–diffusion model (Ratcliff, 1978) describes individuals' cognitive processes in binary decision-making tasks, and distinguishes between decisional and non-decisional processes. By taking the whole intra-individual RT-distribution of correct and incorrect responses into account, we can estimate different parameters (for an illustration of the drift–diffusion model see Fig. 1). This means that the drift–diffusion model accounts for participants' RTs and accuracy equally.

The model describes the decision-making process over time as a random walk during which information is taken up and evidence for a decision gets accumulated. In the drift–diffusion model, v describes the speed of information uptake and the strength and direction of the evidence accumulation process, that is, the average increase of evidence supporting one of the two choices per time unit. The decision process starts at the starting point (z), which can be used to model biases in decision making. During information uptake, the decision process approaches one of two decision thresholds. One threshold describes the correct and the other the alternative response. The boundary separation parameter (a) represents the distance between the two thresholds. In the course of time, v reaches one of two thresholds. Once it crosses a threshold, the decision process is terminated, and the response gets executed. The non-decision time parameter (t_0) describes the speed of all non-decisional processes, such as the speed of motor-response execution and the speed of perceptional processes (see also Ratcliff et al., 2008; Voss et al., 2013).

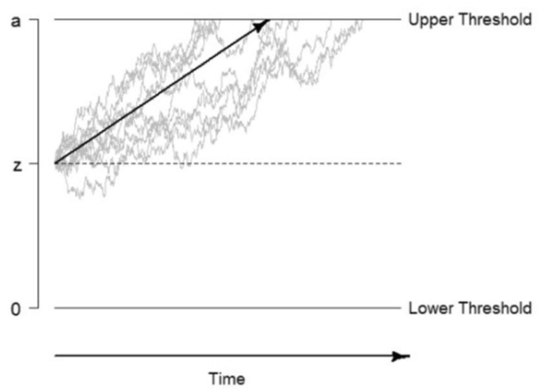


Fig. 1 Graphical illustration of the drift–diffusion mode. *Note.* The decision process begins at the starting point z. Over the time more and more information will be accumulated until one of both thresholds is reached. The drift parameter v represents the strength and direction of the evidence accumulation process (represented by the black arrow). The parameter a describes the distance between both thresholds. The figure does not display the non-decision time t_0. Figure with permission from Frischkorn and Schubert (2018), licensed under CC BY

Generally, participants' drift rates in the standard drift–diffusion model are considered as measures of the speed of information uptake over time (Schmiedek et al., 2007; Voss et al., 2013) and are thus ability parameters that differ between individuals. Previous research on the psychometric properties of the drift rate parameter has shown that drift rates reflect both task-general and condition-specific processes (Lerche et al., 2020; Schubert et al., 2016). These specific components can be considered as specific abilities independent of the task-general speed of information uptake. In a previous study, Lerche et al. (2020) used a battery of different tasks measuring several domains of abilities (numerical, verbal, and figural) and separated the domain-specific processes (numerical, verbal, and figural abilities) from the domain-general processes (i.e., speed of information uptake) reflected in drift rates using bi-factor models. Following this logic, we aim to isolate variance specific to executive functions in drift rates after controlling for task-general processes (i.e., the speed of information uptake) on a latent level.

In contrast to the standard drift–diffusion model, recent research developed specific mathematical models to describe and measure the processes specific to executive functions more accurately: the shrinking spotlight diffusion model (White et al., 2011), the dual-stage two-phase model (Hübner et al., 2010), and the diffusion model for conflict task (Ulrich et al., 2015). However, in contrast to the standard drift–diffusion model, these newly developed models are specific to certain tasks (e.g., Arrow Flanker task or inhibition tasks) and not generalizable for all executive function tasks (e.g., updating and shifting tasks). In this study, our goal was to use one homogenous measurement score for all executive function tasks, which is why we chose the standard drift–diffusion model, fully aware that v represents only an approximation of the underlying processes.

First results of studies using drift rates to measure individual differences in executive functions are promising. For example, a recently published study showed that drift rates estimated from performances in seven different cognitive control tasks formed a common task-general factor of cognitive efficiency, which was related to self-reported cognitive control[1] (Weigard et al., 2021). However, it remains unclear to what degree this factor reflected variance specific to executive processes and to what degree it reflected participants' general speed of information uptake. Therefore, to capture individual differences in executive functions by drift rates, it is necessary to control for participants' task-general speed of information uptake.

Structural equation modeling approach to avoid difference scores

To address the second problem of executive function research—the use of potentially problematical manifest difference scores—we proposed a structural equation modeling approach. In detail, we used the drift rates from the condition with greater processing demands of the different executive function tasks as *homogenous measurement scores* and controlled on the latent level for the influence of task-general processes, because the drift rates of the conditions with greater processing demands reflected not only processes specific to executive functions. We chose this structural equation modeling account because this method allows separating different kinds of variances and in particular distinguishing the variance in drift rates unique to executive function demands from the variance reflecting task-general processing demands. In our study, we therefore controlled the latent executive function factors for the influence of task-general speed of information uptake.

The present study

The aim of the present study was to examine the factor structure of executive functions, whereby we attempted to address the two identified problems of executive function measures: (1) The inconsistent use of accuracy- and RT-based scoring methods and (2) the use of psychometrically unsatisfying difference scores. By applying a cognitive mathematical modeling approach and using v from the drift–diffusion model, we used a homogeneous

[1] We consider cognitive control as a construct that is closely related to executive functions.

scoring method for all executive function tasks. Additionally, we used structural equation models to separate the variance of task-general process demands from the variance of specific executive function demands.

Using data from 148 participants who completed a battery of different cognitive and experimental tasks, we first tried to replicate the factor structure of the seminal paper by Miyake et al. (2000) using accuracy rates and RT-based scores as performance measures (heterogenous measurement scores). Second, we estimated individuals' task performances of all tasks with the drift rate parameter from the drift–diffusion model to integrate both RTs and accuracy rates simultaneously into one performance score. Third, we examined the factor structure of inhibition, updating, and shifting based on these parameter estimates. Fourth, we tested whether executive functions showed divergent validity to task-general speed of information uptake. Fifth, we evaluated the predictive validity of executive functions by relating them to individual differences in WMC and cognitive abilities. Taken together, our goal was to assess the factor structure of executive functions using error-free and valid measures of individual differences in inhibition, updating, and shifting.

Materials and methods

Openness and transparency

We provide access to the preprocessed data and the statistical analysis code used for this paper via the Open Science Framework (https://osf.io/6c4pu/). In addition, we provide access to the raw data and to the materials via the Open Science Framework (https://osf.io/4pvz3/; except for the materials of the BIS, which are commercially licensed).

Statements and declarations

We declare no conflicts of interest. The study was approved by the ethics committee of the faculty of behavioral and cultural studies of Heidelberg University (reference number: Löf 2019/1–3). At the beginning of the first study session, participants signed an informed consent. All procedures were conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013). This study was not preregistered.

Participants

We recruited 151 participants from the general population via advertisements in different local newspapers, distribution of flyers, and acquisition by the participant pool of the department. Three participants declared their withdrawal from participation, which leads to a total sample of $N = 148$ participants (♀ 96, ♂ 51, one person declared no affiliation to either gender). We included participants between 18 and 60 years ($M_{age} = 31.52$, $SD_{age} = 13.91$) to generate a sample with heterogeneous cognitive abilities. Four participants stated having a different native language, but they were fluent in German. Thirty-nine percent of the sample had a university degree.

A minimum sample size of $N = 95$ would be needed to the hypothesis of close fit (H0: $\varepsilon \leq 0.05$, H1: $\varepsilon \geq 0.08$) as suggested by Browne and Cudeck (1992) for the most extensive structural equation model in this paper, displayed in Fig. 5 B ($df = 166$, alpha error: $\alpha = 0.05$, power [1-β] = 0.80). The actual sample size of 148 participants yielded a power > 96% to test the hypothesis of close fit.[2] Participants received 75 € and personal feedback about their performances in intelligence and working memory tests as compensation for participation.

Materials

Table S1 in the supplementary materials shows the stimuli presentation times of the following 12 RT tasks. All computer-based tasks were programmed in MATLAB (The MathWorks Inc., Natick, Massachusetts) with the open source software package Psychtoolbox version 3.0.13 (Kleiner et al., 2007). We presented all the stimuli in the RT tasks in the center of the screen on a black background. In each task, we instructed the participants to respond as quickly and as accurately as possible. Before the experimental part of each task, participants worked on practice trials with feedback.

Inhibition

Stroop task In each trial, participants saw one of four color words presented in one of four colors. The meaning of the word could be the same as the color in which the word was presented (congruent condition, 50% of the trials) or not (incongruent/inhibition condition, 50% of the trials). By pressing one of four keys on the keyboard, participants had to state the color of the word while they had to ignore its meaning (Stroop, 1935). Colored stickers on certain keys

[2] We followed the recommendations by MacCallum et al. (1996) and conducted the power analysis by comparing the null hypothesis RMSEA (RMSEA = .05) with an alternative hypothesis RMSEA (RMSEA = .08). Using both RMSEA values, the given sample size of $N = 148$, and the degrees of freedom of the model (e.g., $df = 166$), we calculated the non-centrality parameters for both hypotheses. With these parameters, an $\alpha = .05$, and the given dfs, we calculated the critical χ^2 value and subsequently the observed power using the cumulative distribution function of the χ^2 distribution.

of the keyboard indicated the key mapping. We randomized the trials, with none of the conditions occurring more than three times in a row and none of the colors or words occurring twice in a row. Participants worked on 20 practice trials and 192 experimental trials.

Arrow flanker task In each trial, one target arrow appeared in the center of the screen, pointing to the left or to the right direction. This target stimulus appeared in the middle of four flanker stimuli, two on each of both horizontal sides. The distractors could either point in the same direction as the target (congruent condition) or in the opposite direction (incongruent/inhibition condition; Eriksen & Eriksen, 1974). Participants had to indicate the side to which the target stimulus pointed while ignoring the distractors by pressing one of two keys on the keyboard. We randomized the trials, with none of the conditions or target directions occurring more than three times in a row. Participants worked first on 20 practice trials followed by 200 experimental trials.

Negative priming task In each trial, two horizontal lines appeared on both sides next to the center of the screen. Subsequently, an X and an O appeared simultaneously on two of these lines. Participants had to indicate the position where the O appeared by pressing one of four keys while ignoring the X. In 50% of the trials, the O appeared at the position where the X appeared one trial before. To respond to an O shown at such a negatively primed position, participants had to redirect their attention to the positions previously associated with the distractor and overcome the transient residual inhibition (Tipper & Cranston, 1985). We randomized the trials, with none of the conditions (negatively primed vs. not negatively primed) occurring more than three times in a row and none of the stimuli appearing more than three times in a row on the same position. Participants worked on 20 practice trials and 192 experimental trials.

Updating

Keep track task We adopted this task from the study by Miyake et al. (2000). Participants completed two blocks with different updating steps. The stimulus material consisted of four categories (letters, numbers, colors, geometric figures) and six stimuli within each category. Before each trial started, participants received an instruction about which of the four categories they had to keep track of. Depending on the block, they had to keep track of one or on three target-categories (updating steps: one or three). After that, participants saw a sequence of seven stimuli. This sequence contained stimuli from each of the four categories. Subsequently, a probe stimulus from one of the target categories followed. Participants had to indicate whether the probe stimulus was the last presented stimulus of the target category/categories (50% of the trials, matching condition) or not by pressing one of two keys on the keyboard. In 50% of the trials the target category was updated (updating condition). Within each block, participants worked on 10 practice trials and 96 experimental trials. We randomized the trials, with none of the conditions (matching and updating) and none of the target categories occurred more than three times in a row.

Running span task We adopted this task from the study by Broadway and Engle (2010). In each trial, the stimuli of the memory and the stimuli of the updating set appeared sequentially in the center of the screen. Afterwards, participants saw a probe stimulus and had to decide whether this probe was part of the last three or last five stimuli, depending on the set size of the block. Participants completed two blocks with different set sizes. In both blocks the updating steps ranged from zero to three. Within the first block, the memory set consisted of three memory-letters followed by zero to three updating-letters. Within the second block, the memory set consisted of five memory-letters followed by zero to three updating-letters. Participants responded by pressing one of two keys on the keyboard. Half of the trials had zero updating steps. The other half of the trials included all one to three updating steps with equal frequency. In each block, participants worked on 10 practice trials and 120 experimental trials. We randomized the trials, with none of the updating steps and none of the probe stimuli occurring more than three times in a row.

N-back task We adopted this task from the verbal working memory conditions of the task by Gevins et al. (1996). Participants completed three blocks, which included a different number of updating steps. In the first block, participants completed a 0-back task. Before the first block started, a target letter appeared, followed by 96 trials. In these trials either the target or a different letter was presented in the center of the screen. Specific target and non-target letters varied between participants. Participants had to decide whether the presented letter was the target or not by pressing one of two keys on the keyboard. Before the experimental part of the first block started, participants had to work on 20 practice trials. Data of the 0-back condition were not included in our analyses. In the second block, participants completed a 1-back task. In each trial, participants saw one of four letters in the center of the screen and had to decide whether or not this letter was equal to the stimulus that had appeared one trial before by pressing one of two keys on the keyboard. In total, participants completed 96 trials. In the third block, participants completed a 2-back task. In each trial, participants saw one of four letters in the center of the screen and had to decide whether or not this letter was equal to the stimulus that had appeared two trials before by pressing one of two keys on the keyboard. Again, we used 96 trials. Before the experimental part of the second and third block started, participants worked on 30 practice trials. Within each block the probe stimulus matched with the target stimulus in 50% of the trials (the stimulus one or two trials before = match condition). We randomized the trials, with none of the stimuli and none of the matching conditions occurring more than three times in a row.

Shifting

In each of the three shifting tasks, the color of the fixation cross at the beginning of each trial was the same as the color of the following probe stimulus.

Switching task In each trial, a number between one and nine (except five) appeared either in red or in green in the center of the screen. Depending on the color of the presented stimulus, participants had to perform different tasks (Sudevan & Taylor, 1987). They had either to decide whether the number was less or more than five (red) or the number was odd or even (green) by pressing one of two keys on the keyboard. Both tasks appeared with equal frequency. In 50% of the trials, the task was the same as one trial before (repeat condition); in the other 50% of the trials, the color was different to the last trial (shifting condition). We randomized the trials, with none of the tasks and none of the conditions occurring more than three times in a row and none of the numbers appearing twice in a row. Participants worked on 10 task-pure practice trials for each of the two tasks, followed by 20 practice trials with both tasks intermixed. After the practice block, participants worked on 384 experimental trials.

Number letter task In each trial, one number between one and nine (except five) together with one letter out of a set of eight letters appeared either in red or in green in the center of the screen. The letter set consisted of the letters A, E, I, U, G, K, M, and R. Depending on the color of the presented stimuli, participants had to perform different tasks. They had either to decide whether the number was less or more than five (red) or the letter was a consonant or a vocal (green) by pressing one of two keys on the keyboard (Rogers & Monsell, 1995). Both tasks appeared with equal frequency. Additionally, in 50% of the trials the task was the same as one trial before (repeat condition); in the other 50% of the trials, the color was different to the last trial (shifting condition). We randomized the trials, with none of the tasks and none of the conditions occurring more than three times in a row and none of the numbers and letters appearing twice in a row. Participants worked on 10 task-pure practice trials for each of the two tasks, followed by 20 practice trials with both tasks included. After the practice block, participants worked on 256 experimental trials.

Global local task We adopted this task from the study by Miyake et al. (2000). In each trial, one of four geometrical shapes (circle, triangle, square, cross) appeared either in red or in green in the center of the screen. This figure was composed of small geometric shapes from the same set of shapes, better known as Navon-figures (Navon, 1977). The larger figure (global) and the smaller figure (local) could never have the same geometrical shape. Depending on the color, participants had to perform different tasks. They had either to identify the shape of the large figure (red) or the shape of the small figures (green) by pressing one of four keys on the keyboard. Both tasks appeared with equal frequency. In 50% of the trials the condition was the same as one trial before (repeat condition) in the other 50% of the trials the color was different to the last trial (shifting condition). We randomized the trials, with none of the tasks and none of the conditions occurring more than three times in a row and none of the large figures appearing twice in a row. Participants worked on 10 task-pure practice trials for each of the two tasks, followed by 20 practice trials with both tasks intermixed. After the practice block, participants worked on 384 experimental trials.

Processing speed

Two choice reaction time task In each trial, participants had to focus on a centrally presented fixation cross, which was amid two quadratic frames. A plus sign appeared either in the left or in the right frame (e.g., Chen et al., 2012). Participants had to indicate whether the plus appeared in the left or in the right frame by pressing one of two response keys on the keyboard. The plus appeared on both sides with equal frequency. We randomized the trials, with none of the stimulus presentation sides repeating more than three times in a row. Participants worked on 20 practice trials, followed by 100 experimental trials.

Sternberg task In each trial, five numbers between zero and nine appeared sequentially in the center of the screen. Following this sequence, a probe stimulus appeared and participants had to decide whether this probe was part of the formerly presented set or not (Sternberg, 1969) by pressing one of two response keys on the keyboard. In 50% of the trials the probe stimulus was part of the set (match condition). All numbers occurred with equal frequency as probe stimulus. We randomized the trials, with none of the conditions (match vs. no match) occurring more than three times in a row and none of the probe stimuli occurring twice in a row. Participants worked on 20 practice trials, followed by 100 experimental trials.

Posner task In each trial, two letters appeared in the center of the screen. The stimulus set included the letters A, B, F, H, Q, a, b, f, h, q. Participants had to decide whether the meaning of the two letters was identical or not (e.g., Aa or AA = identical, AB or Ab = not identical; Posner & Mitchell, 1967). In 50% of the trials the letters had identical names. We randomized the trials, with none of the conditions (identical vs. not identical) and none of the letters occurring more than three times in a row. Participants worked on 20 practice trials, followed by 120 experimental trials.

Working memory capacity (WMC)

We used the memory updating task, the operation span task, the sentence span task, and the spatial short-term memory task from the working memory test battery by Lewandowsky et al. (2010) to assess participants' WMC. In addition, we used the location-letter binding task by Wilhelm et al. (2013). All participants except five completed this letter binding task. For each of the different set sizes in the working memory tasks, we calculated participants' mean proportion of correctly solved items as the dependent variable. Due to a programming error, we could not use the data of the spatial short-term memory task in our analyses.

Fluid intelligence

We used the short version of the Berlin Intelligence Structure Test (BIS, Jäger et al., 1997) as an assessment for fluid intelligence, which is a particularly suitable instrument for measuring higher-order cognitive abilities in a relatively short amount of time (about 50–60 min). The short version consists of a heterogeneous test battery including 15 different tasks. Four operation-related (processing capacity [PC], processing speed [PS], memory [M], creativity [C]) and three content-related (verbal, numerical, figural) components of intelligence can be assessed with the short version of the BIS. For our analyses, we calculated participants' operation-related component scores by aggregating the normalized *z*-scores of all subtests measuring the respective component. Participants had a mean IQ of 96 ($SD = 15.86$).

Procedure

Participants completed three measurement occasions within one year. At the beginning of the first session, participants signed an informed consent and completed the Ishihara-Test (Ishihara, 2000) to rule out that they were colorblind. Following that, we prepared participants' EEG and seated them in a dimly lit cabin during the first and second measurement occasions. The EEG data are not reported in the current paper (see Sadus et al., 2023; Schubert et al., 2022a, 2022b). Subsequently, participants worked on the 12 tasks in the following order. Measurement occasion one: Sternberg task, Arrow Flanker task, Global Local task, N-Back task, Switching task, and Stroop task. Measurement occasion two: Running Span task, Two Choice Reaction Time task, Number Letter task, Negative Priming task, Keep Track task, and Posner task. In addition, participants completed a questionnaire about their demographical data at the end of the first occasion. Each occasion lasted approximately 3.5 h. To avoid between-subjects error variance by balancing the task order, we decided to present all tasks for all participants in the same order, well knowing that this procedure might result in fatigue, reduced motivation, or sequence effects systematically affecting performance measures (Goodhew & Edwards, 2019). During the third measurement occasion, participants first completed the intelligence test followed by the working memory test battery and the letter binding task. In addition, participants also completed two short tests measuring their higher-order cognitive abilities, a mind-wandering questionnaire, and a pretzel task (these data are not reported here).

Data analysis

We used the statistics software R—version 4.1.0 (R. Core Team, 2022) for data preprocessing and analyses and used the following packages: For preparation and data management the package "tidyverse" (Wickham et al., 2019), for descriptive statistics the package "psych" (Revelle, 2020), for correlations the package "Hmisc" (Harrell, 2019), for structural equation model analyses the package "lavaan" (Rosseel, 2012), for confidence interval estimations the package "MBESS" (Kelley, 2007), and for the preparation of the correlation matrices the package "patchwork" (Pedersen, 2020).

Outlier analysis and data processing

Before we conducted the main analyses, we performed univariate intra- and inter-individual outlier analyses. The procedure was identical for each participant and variable. The detected outliers (trials or participants) were excluded only from the corresponding conditions of the respective task.

For the intra-individual outlier analysis, we applied the following steps to each condition in each executive function and processing speed task. Initially, responses faster than 150 ms were discarded. Subsequently, we logarithmized and *z*-transformed the RT variables for each participant and removed the trials with *z*-values greater than 3 or smaller than -3. On average, 0.69% of the trials were removed within each condition of the 12 reaction time tasks (range: 0.33% to 1.06%).

Next, we conducted inter-individual outlier analyses based on both RT and accuracy scores for each condition in the twelve tasks. Participants with accuracy scores below the guessing probability threshold were discarded. This threshold was determined based on the number of trials and response options of the corresponding condition, assuming a binomial distribution. In addition, we identified mean RTs or logit-transformed accuracy values that deviated from the average by more than 3 standard deviations as inter-individual outliers. These participants were removed from the corresponding task.

Following the outlier analyses, we modified the data for subsequent analyses according to our requirements. This involved estimating participants' drift–diffusion model parameters for all conditions of all tasks separately (details below). In addition, to replicate the model of three interrelated executive functions by Miyake et al. (2000), we removed all incorrect trials and calculated participants' RT-difference scores for the shifting and inhibition tasks, their mean RTs for the inhibition tasks, and their arcsine-transformed probability scores for the updating tasks. Before we inserted the variables in the structural equation models, we discarded the values deviating from the average by more than three standard deviations. Accumulated over all these steps of the inter-individual outlier detection, we removed, on average, 3.63% of the participants within each of the variables (range: 0.70% to 7.09%).

Drift–diffusion modeling

We fitted the diffusion model parameters with *fast-dm-30* (Voss et al., 2015) using the Kolmogorov–Smirnov criterion for optimization. For each participant, we estimated (v), the boundary separation (a), the non-decision time (t_0), and the inter-trial-variability of the non-decision time (st_0) in the conditions with greater processing demands of the executive function tasks. Further, we followed the recommendations of Lerche and Voss (2016) and fixed all other parameters to zero except the starting point z, which we centered between the two decision thresholds ($z = 0.5$).

Subsequently, we assessed if the drift–diffusion models provided a good account to the observed data by evaluating the models using simulated RT and accuracy data based on model parameters. The correlations between the observed and predicted scores were between $r = 0.94$ and $r = 0.99$ for the RTs in the 25th, 50th and 75th percentile of the RT distributions and between $r = 0.47$ and $r = 0.87$ for the overall accuracy scores (except for the accuracies in the Two Choice Reaction Time task, $r = 0.05$; see for further discussion the limitations section), which indicated that there was overall no evidence for a systematically biased model prediction. For a visual inspection of the model fits see Fig. S1 to Fig. S4 in the supplementary materials.

In three decision tasks, participants had to respond by pressing four instead of two keys, which is not a binary choice in the classical way (Stroop task, Negative Priming task, Global Local task). However, Voss et al. (2015) argued that diffusion modeling of tasks with more than two response keys is possible under some assumptions: The responses have to be re-coded as either correct or incorrect, drift rates should not differ between stimulus types, there should be no bias in response behavior, and these tasks should have a sufficient number of errors (Voss et al., 2015). The three tasks with more than two response options in our study met these assumptions. In addition, the parameter recovery indicated no systematically lower predictions of these scores compared with the classical binary choice tasks.

Structural equation modeling

First, we wanted to replicate the original model of three interrelated executive function factors by Miyake et al. (2000). For this, we used similar scores for the manifest variables as in the original study. Second, we estimated the three-factor model of executive functions, with drift parameters difference scores. Therefore, the drift rate parameters were estimated separately for the two conditions of each task, while the other parameters of the drift–diffusion model were kept constant. Afterwards, we contrasted the drift parameters to get the drift differences scores for each of the nine executive function tasks and inserted these difference scores as indicators in the three-factor model of executive functions. Third, we estimated the drift parameters only for the conditions with greater processing demands, which were used as indicators for the following analyses and models. Again, we specified the three-factor model of executive functions based on these drift rate parameters. Fourth, we estimated a model with a second-order factor as well as a model with a first-order factor of *common executive functions* and examined in the following step the relations of this common factor to higher-order cognitive abilities. Fifth, to control for task-general speed of information uptake, we regressed the common factor of executive functions on a task-general *speed* factor estimated from three elementary cognitive tasks and examined again the relations of the latent variables to intelligence and WMC.

To account for missing data, we used full information maximum likelihood (FIML). We fixed one of the loadings of each factor to one and estimated the variances of the latent factors. The goodness-of-fit was evaluated by the comparative fit index (CFI; Bentler, 1990) and the root mean square error of approximation (RMSEA; Browne & Cudeck, 1992). Following the recommendations by Browne and Cudeck, (1992) as well as Hu and Bentler (1999), we considered CFI values > 0.90 and RMSEA values ≤ 0.08 as an acceptable model fit and CFI values > 0.95 and RMSEA values ≤ 0.06 as good model fit. In direct model comparisons, AIC differences ≥ 10 indicated substantial advantages (Burnham &

Table 1 Descriptive statistics of the heterogeneous measurement scores

Task name	Measurement score	Mean	*SD*	Reliability
Negative priming task	RT difference	0.02	0.02	0.28[a]
Flanker task	RT difference	0.03	0.02	0.66[a]
Stroop task	RT difference	0.11	0.06	0.77[a]
Negative priming task, priming cond	RT	0.61	0.11	0.99[a]
Flanker task, incong. cond	RT	0.50	0.08	0.99[a]
Stroop task, incong. cond	RT	0.82	0.14	0.98[a]
Keep track task, updating cond	a. t. proportion correct (percent correct)	1.29 (91.40)	0.11 (6.38)	0.57[a]
Running span task, updating cond	a. t. proportion correct (percent correct)	1.28 (90.99)	0.08 (4.85)	0.64[a]
N-back task	a. t. proportion correct (percent correct)	1.21 (86.29)	0.13 (8.64)	0.85[a]
Number letter task	RT difference	0.06	0.07	0.68[a]
Switching task	RT difference	0.05	0.06	0.45[a]
Global local task	RT difference	0.09	0.07	0.41[a]
Two choice RT task	RT	0.38	0.04	0.99[a]
Sternberg task	RT	0.91	0.22	0.98[a]
Posner task	RT	0.71	0.13	0.99[a]
Memory updating	Percent correct	63	20	0.88[b]
Binding	Percent correct	86	11	0.82[b]
Operation span	Percent correct	78	13	0.89[b]
Sentence span	Percent correct	84	11	0.87[b]
BIS-PC	Scales-scores	101.61	7.12	0.75[b]
BIS-PS	Scales-scores	101.14	7.15	0.49[b]
BIS-M	Scales-scores	98.59	7.16	0.58[b]
BIS-C	Scales-scores	98.15	6.97	0.45[b]

Note. Heterogeneous measurement scores; a. t. = arcsine-transformed; BIS-PC = processing capacity scale of the Berlin Intelligence Structure Test; BIS-PS = processing speed scale of the Berlin Intelligence Structure Test; BIS-M = memory scale of the Berlin Intelligence Structure Test; BIS-C = creativity scale of the Berlin Intelligence Structure Test

[a]Reliability estimates are based on Spearman–Brown corrected odd–even split correlations

[b]Reliability estimates are based on Cronbach's α; RT-values are displayed in seconds; proportion correct = arcsine-transformed proportion correct scores

Anderson, 2002). We assessed the statistical significance of model parameters with the two-sided critical ratio test.

Results

First, we tried to replicate the model of three distinct but interrelated factors of executive functions by Miyake et al. (2000) with heterogeneous measurement scores. Afterwards, we examined the factor structure of executive functions and its relation to higher-order cognitive abilities using the drift parameters of the drift–diffusion model as homogenous measurement scores. The descriptive statistics of the heterogeneous measurement scores are displayed in Table 1. We found large variations of reliability estimates for the heterogeneous measurement scores (see Table 1). The reliability estimates were excellent for the inhibition tasks if performance was measured by mean RTs and poor to acceptable if performance was measured by RT-differences scores. Reliabilities varied from moderate to good in the updating tasks, where performance was measured by arcsine-transformed proportion correct scores. Reliabilities were poor in the shifting tasks, where performance was measured by RT-difference scores. The correlations between the heterogeneous measurement scores are shown in Table S2 in the supplementary materials.

We specified the model of three distinct but interrelated factors of executive functions to replicate the model by Miyake et al. (2000) and compared how the factor structure of drift rates differed from the factor structure of heterogeneous measurement scores. In the original model by Miyake et al. (2000), they used RT-based difference scores between incongruent and congruent conditions to

Psychological Research (2024) 88:1092–1114

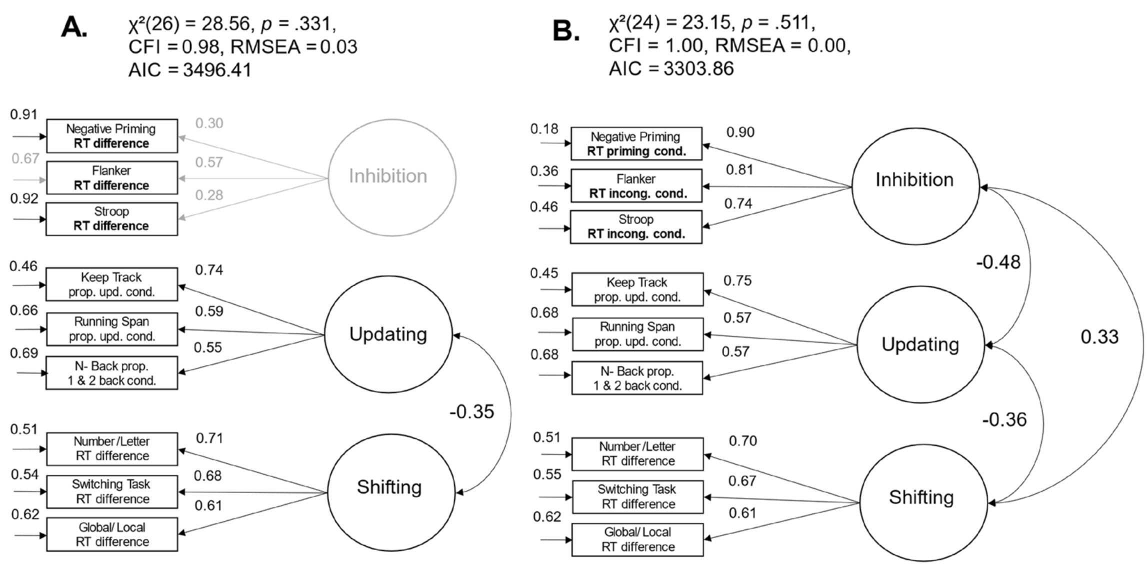


Fig. 2 Three-factor models of executive functions with heterogeneous measurement scores. *Note*. The standardized path weights, the unstandardized residual variances, and the correlation coefficients are shown next to the paths; non-significant estimators are grayed out

measure participants' abilities in inhibition, RT-based difference scores between shifting and repeat conditions to measure participants' shifting abilities, and arcsine-transformed proportion correct scores to measure participants' updating performance. When we used these measurement procedures in our own data, the model provided a good account of the data, $\chi^2(26) = 28.56$, $p = 0.331$, CFI = 0.98, RMSEA = 0.03, 95% CI [0.00, 0.08]; see Fig. 2A). However, we could not find significant variance in the latent inhibition factor with RT difference scores, $\sigma^2 = 0.09$, $p = 0.431$, 95% CI [– 0.13, 0.30]. Therefore, we decided to use the mean RT-scores of the conditions with greater processing demands, the inhibition conditions, to examine individual differences in inhibition. The corresponding model provided an excellent account of the data, $\chi^2(24) = 23.15$, $p = 0.511$, CFI = 1.00, RMSEA = 0.00, 95% CI [0.00, 0.07]; see Fig. 2B). The three latent executive function factors were moderately correlated. Participants who were less distracted by irrelevant information (shorter RTs in inhibition tasks) showed better updating abilities, $r = -0.48$, $p < 0.001$, 95% CI [– 0.66, – 0.30], and lower shifting costs, $r = 0.33$, $p = 0.004$, 95% CI [0.13, 0.53]. Moreover, participants with better updating abilities showed lower shifting costs, $r = -0.36$; $p = 0.010$, 95% CI [– 0.59, – 0.13].

We were (mostly) able to replicate the original model of three distinct but interrelated factors of executive functions. In the next step, we wanted to examine the factor structure of executive functions by using drift rates instead of RT-/accuracy-based performance scores.

Conducting the analyses based on the drift rate parameters, we first used drift rate difference scores to examine the reliability and factor structure of executive functions. However, the covariance matrix of the latent variables was not positive definite and the model did not converge. Furthermore, the Spearman–Brown corrected odd–even correlations indicated insufficient reliabilities or even inadmissible estimates for the drift rate difference scores (ranging from -0.12 to 0.66). In consequence, we can conclude that even for drift rates, the difference scores tended to be unreliable and did not prove to be useful indicators measuring individual differences in executive functions. This highlights the limited utility of difference scores in executive function research and underscores our strategy to examine drift rates of the conditions with greater processing demands and to disentangle the sources of variance at the latent level. Table 2 shows the descriptive statistics for drift rates from the conditions with greater processing demands of the nine executive function tasks and of the three processing speed tasks and (see Table S3 in the supplementary materials for the descriptive statistics of the other estimated drift–diffusion model parameters).

Overall, the reliabilities of drift rates were on average slightly smaller but comparable to RT- and accuracy-based performance measures. They showed a broad range from poor to good reliabilities. The small difference between the reliabilities of heterogeneous and homogeneous measurement scores are mainly driven by the RT average scores of the inhibition- and elementary cognitive tasks, which usually show very high reliabilities. In comparison to the reliabilities reported in Table 1, reliabilities were higher for updating and shifting tasks, but lower for inhibition tasks. The correlations between the drift rate parameters of each task are shown in Table S4 in the supplementary materials.

Table 2 Descriptive statistics of drift rates

	Measurement score	Mean	SD	Reliability
Negative priming task, priming cond	Drift parameter *v*	3.62	0.94	0.47
Flanker task, incong. cond	Drift parameter *v*	5.05	1.32	0.57
Stroop task, incong. cond	Drift parameter *v*	2.60	0.74	0.48
Keep track, updating cond	Drift parameter *v*	1.69	0.62	0.81
Running span, updating cond	Drift parameter *v*	1.78	0.58	0.67
N-back task	Drift parameter *v*	1.74	0.46	0.83
Number letter task, shifting cond	Drift parameter *v*	2.29	0.92	0.89
Switching task, shifting cond	Drift parameter *v*	2.16	0.85	0.90
Global local task, shifting cond	Drift parameter *v*	1.65	0.51	0.79
Two choice RT task	Drift parameter *v*	6.25	1.58	0.70
Sternberg task	Drift parameter *v*	2.35	0.71	0.45
Posner task	Drift parameter *v*	3.27	0.74	0.53

Note. Drift rates *v* as measurement scores; reliability estimates are based on Spearman–Brown corrected odd–even split correlations

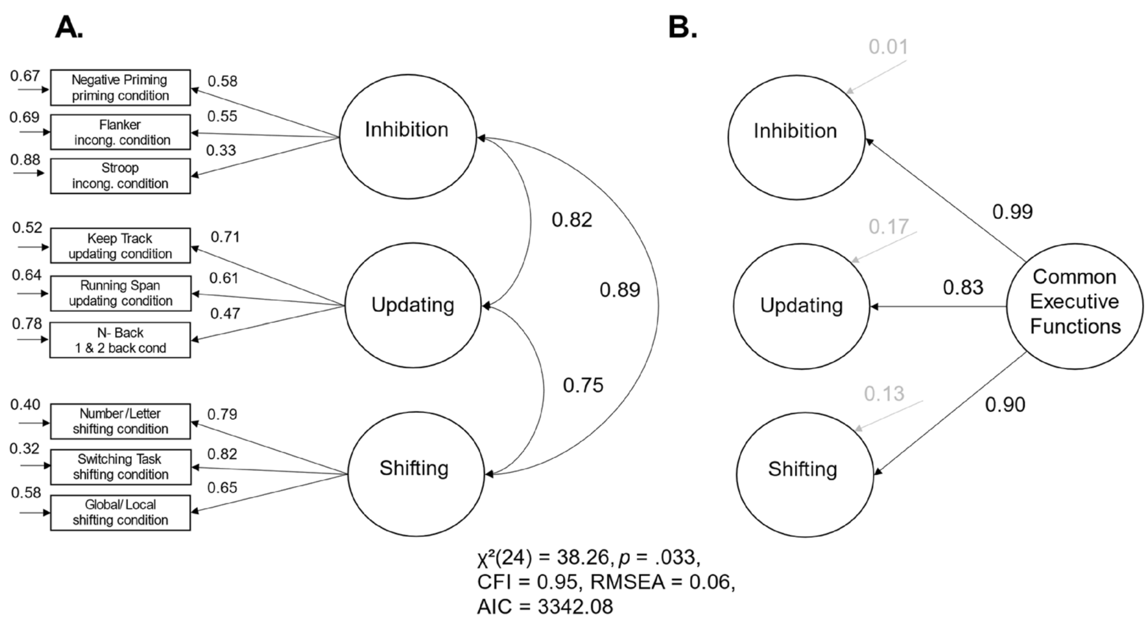


Fig. 3 Three-factor models of executive functions with drift rates as homogenous measurement scores. *Note*. The standardized path weights, the unstandardized residual variances, and the correlation coefficients are shown next to the paths; non-significant estimators are grayed out

Subsequently, we specified the model of three distinct, but interrelated factors based on drift rates instead of the heterogeneous measurement scores (see Fig. 3A). The model provided a good fit of the data, $\chi^2(24) = 38.26$, $p = 0.033$, CFI = 0.95, RMSEA = 0.06, 95% CI [0.00, 0.11]. The three latent executive function factors were highly correlated. Participants with higher drift rates in inhibition tasks showed higher drift rates in updating tasks ($r = 0.82$, $p < 0.001$, 95% CI [0.59, 1.06]) as well as higher drift rates in shifting tasks, $r = 0.89$, $p < 0.001$, 95% CI [0.70, 0.1.08]. Furthermore, participants with higher drift rates in updating tasks showed higher drift rates in shifting tasks, $r = 0.75$, $p < 0.001$, 95% CI [0.60, 0.90]. Taken together, we were also able to find the three latent factors of executive functions by using drift rates instead of heterogenous measurement scores. The positive manifold in the correlations between the three latent factors suggests a hierarchical factor structure with a higher-order factor of executive functions or a one-factor solution with a common factor of executive functions on the first level.

In consequence, we introduced a higher-order factor of executive functions (*common executive functions*) in our model with drift rates as manifest variables (see Fig. 3 B). The model fit was equivalent to the model just described, in which the latent first-order factors were correlated. However,

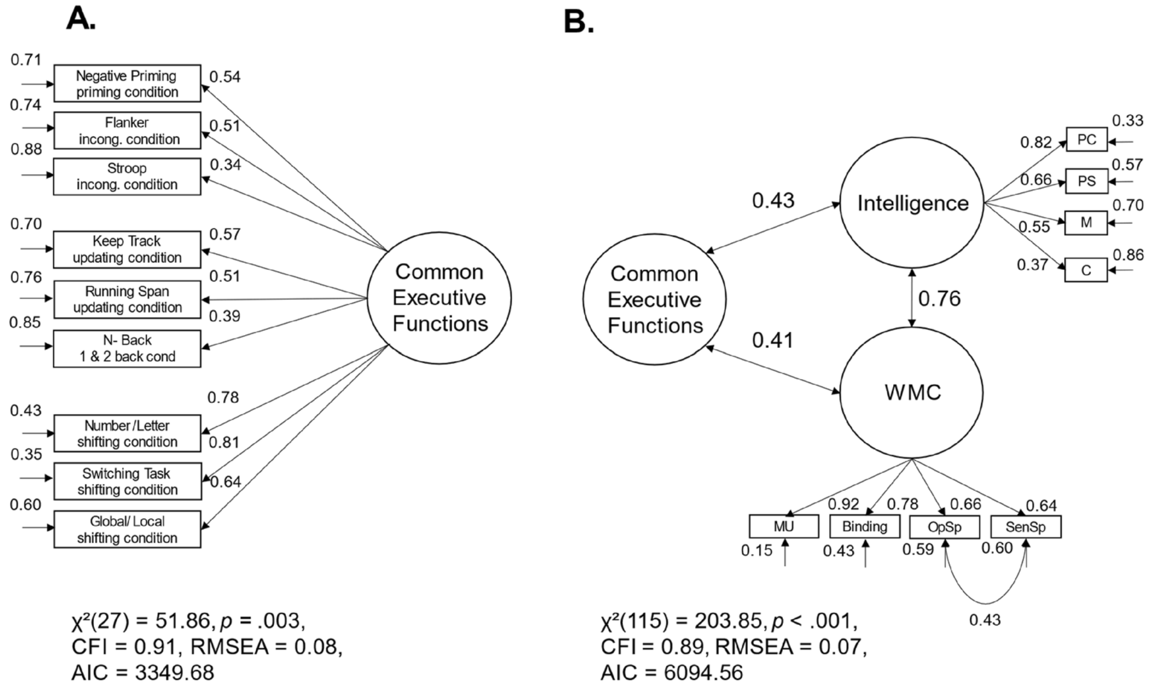


Fig. 4 Models with one common factor of executive functions with drift rates as homogenous measurement scores. *Note*. The standardized path weights, the unstandardized residual variances, and the correlation coefficients are shown next to the paths; MU = memory updating; BIS-PC = processing capacity scale of the Berlin Intelligence Structure Test; BIS-PS = processing speed scale of the Berlin Intelligence Structure Test; BIS-M = memory scale of the Berlin Intelligence Structure Test; BIS-C = creativity scale of the Berlin Intelligence Structure Test; WMC = working memory capacity

after introducing this second-order factor of executive functions, there was no remaining significant residual variance specific to each of the three executive functions (inhibition: residual variance $\sigma^2 = 0.01$, $p = 0.942$, 95% CI [− 0.14, 0.15]; updating: residual variance $\sigma^2 = 0.17$, $p = 0.061$, 95% CI [− 0.01, 0.34]; shifting: residual variance $\sigma^2 = 0.13$, $p = 0.133$, 95% CI [− 0.04, 0.30]). If we fixed the residual variances to zero, the model fit deteriorated only slightly, but not above the critical AIC difference proposed by (Burnham & Anderson, 2002), Δ AIC = 7.60, $\chi^2(27) = 51.86$, $p = 0.003$, CFI = 0.91, RMSEA = 0.08, 95% CI [0.04, 0.12]. Inhibition, updating, and shifting, as measured with v, were fully explained by the higher-order factor.

The non-significant residual variances of the three latent factors of executive functions on the first level suggested a one-factor structure of executive functions. Therefore, we specified a model with only one latent common executive functions factor (see Fig. 4A). The model fit was equivalent to the model just described, in which the residual variances of the first-order factors were set to zero. Our findings suggest that the drift parameters of the different executive function tasks represented individual differences in one common executive ability.

In the next step, we examined the correlations between the common factor of executive functions and higher-order cognitive abilities. Please note that in this model, we have not yet controlled for task-general speed of information uptake. We introduced Intelligence and WMC as latent factors into the model (see Fig. 4B). The model provided an almost acceptable fit of the data, $\chi^2(115) = 203.85$, $p < 0.001$, CFI = 0.89, RMSEA = 0.07, 95% CI [0.05, 0.09]. Intelligence and WMC factors were highly correlated, $r = 0.76$, $p < 0.001$, 95% CI [0.64, 0.88]. Individual differences in executive functions were moderately related to intelligence ($r = 0.43$, $p < 0.001$, 95% CI [0.25, 0.62]) and WMC, $r = 0.41$, $p = 0.001$, 95% CI [0.25, 0.58].[3]

So far, the common variance of the drift rates represented both task-general speed of information uptake as well as variance specific to executive functions. In the next step, we wanted to assess if the relationship between the variance specific to executive functions and intelligence as well as WMC pertained if we controlled for task-general speed of information uptake by introducing a latent *speed* factor to the model. This speed factor represented the covariance of

[3] If we introduced a higher-order cognitive abilities factor in our model instead of intelligence and WMC separately, the general executive functions factor was also moderately related to cognitive abilities, $r = .48$, $p < .001$, 95% CI [.31, .66].

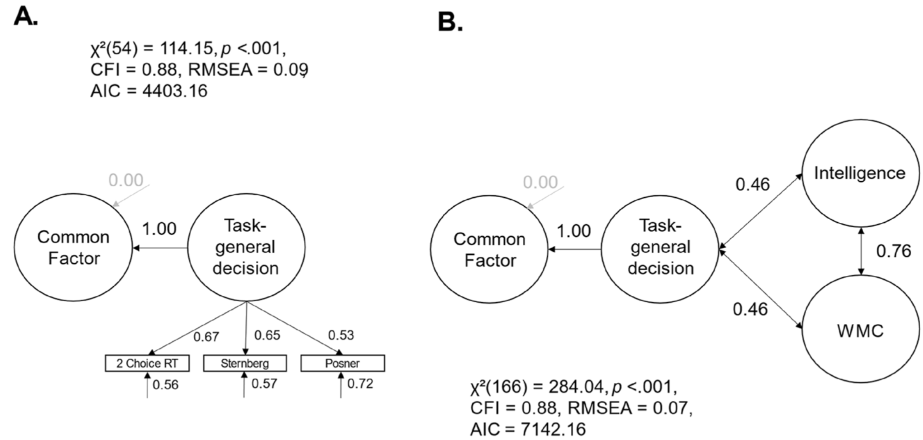


Fig. 5 Models with one common factor of executive controlled for task-general speed of information uptake and its correlations to higher-order cognitive abilities (drift rates as homogenous measurement scores). *Note.* The standardized path and regression weights, the unstandardized residual variances, and the correlation coefficients are shown next to the paths; non-significant estimators are grayed out.; WMC = working memory capacity

three elementary cognitive tasks. The common executive functions factor was regressed on the latent speed factor to account for individual differences in task-general speed of information uptake. The model provided an almost acceptable fit of the data, $\chi^2(53) = 114.14$, $p < 0.001$, CFI = 0.87, RMSEA = 0.09, 95% CI [0.06, 0.11]. However, there was no significant variance in the common executive functions factor independent of task-general speed of information uptake, $\sigma^2 = -0.00$, $p = 0.913$, 95% CI [− 0.07, 0.06]. Given the small negative and non-significant residual variance of the common executive functions factor, we followed the recommendations by Chen et al. (2012) and fixed this residual variance to zero. This hardly changed the model fit, Δ AIC = 1.99, $\chi^2(54) = 114.15$, $p < 0.001$, CFI = 0.88, RMSEA = 0.09, 95% CI [0.06, 0.11] (see Fig. 5A). In sum, the common factor of executive functions was completely explained by the task-general speed of information uptake, $\beta = 1.00$, $p < 0.001$. In this context, the shared variance of executive function tasks as measured with drift rates only represented individual differences in task-general speed of information uptake.

Again, we included WMC and intelligence as latent factors into the model to examine the relations between task-general speed of information uptake and higher-order cognitive abilities (see Fig. 5B). The model provided an almost acceptable fit of the data, $\chi^2(166) = 284.04$, $p < 0.001$, CFI = 0.88, RMSEA = 0.07, 95% CI [0.05, 0.09]. Participants' task-general speed of information uptake was moderately correlated with intelligence ($r = 0.46$, $p < 0.001$, 95% CI [0.29, 0.63]) and WMC ($r = 0.46$, $p < 0.001$, 95% CI [0.31, 0.62]).[4] Taken together, we found that executive functions measured by drift rates revealed one latent factor of common executive functions. This factor was completely explained by individual differences in task-general speed of information uptake. No variance specific to executive functions remained in our model, which could not be explained by the general speed of information uptake. Participants with a higher general speed of information uptake showed better performance in each of the nine executive function tasks and revealed higher intelligence test scores as well as greater WMC.

Discussion

The aim of the present study was to examine the factor structure of the three executive functions as described by Miyake et al. (2000) using the drift rate parameter of the drift–diffusion model as a homogenous measurement score instead of heterogeneous measurement scores.

Heterogenous measurement scores: a partial replication of the three-factor model

As a first step, we tried to replicate the original model of three distinct but interrelated factors with heterogeneous measurement scores, which was nearly identical to the model proposed in the seminal paper by Miyake et al. (2000). However, using RT difference scores, it was not possible to find significant variance for the factor of inhibition, that is, we did not find a coherent structure of the inhibition construct. This is in line with some recent research, suggesting that inhibition tasks often do not reveal a homogeneous structure of the underlying construct if one uses difference scores as performance measures from different tasks (Frischkorn & von Bastian, 2021; Hull et al., 2008; Krumm et al., 2009; Rey-Mermet et al., 2018; Stahl et al., 2014). When we used the mean RTs of the incongruent conditions of inhibition tasks, as done in several previous studies (see von Bastian et al., 2020), we found a coherent

[4] If we introduced a higher-order cognitive abilities factor in our model instead of intelligence and WMC separately, the correlation between higher-order cognitive abilities and task-general speed of information uptake was also high, $r = .53$, $p < .001$, 95% CI [.37, .69].

Psychological Research (2024) 88:1092–1114

latent inhibition factor and were able to replicate the three-factor structure of executive functions by Miyake et al. (2000) with substantial correlations between these factors. However, such condition-specific scores suffer from validity problems because they may also be affected by task-general processes.

The model of executive functions measured with drift rates

We used drift rates of the drift–diffusion model as homogenous measurement scores and structural equation models to separate task-general properties of drift rates from properties specific to executive functions. Our approach avoided the use of difference scores and their psychometric problems, thereby overcoming the two main issues of measuring executive functions—inconsistent use of RT- and accuracy-based scores and psychometric problems of difference scores. We found three latent factors of executive functions that loaded on one higher-level common executive functions factor. After introducing the higher-order factor, we observed no remaining variance specific to the three executive functions on the first-order latent level. We therefore specified a more parsimonious one-factorial solution, in which all tasks loaded on one latent first-order common executive functions factor. This model described the data only marginally worse than the more complex hierarchical model (Δ AIC = 7.60) and provided a good account of the data. In addition, given the absence of significant residual variances in the hierarchical model, the additional explanatory power of the more complex model seems highly questionable. In this context, the one-factor solution emerges as the more favorable model. However, future work should replicate our results with a larger sample size to get a better understanding of the nature of the common factor and the very small non-significant executive function specific variances as shown in Fig. 3B.

These findings, of a one-factor model, are in contrast to the three-factor model proposed by Miyake et al. (2000) and several other papers that found distinct factors of executive functions by using heterogeneous scoring methods (e.g., Friedman et al., 2006, 2008; Himi et al., 2019, 2021; Ito et al., 2015; Krumm et al., 2009; Schnitzspahn et al., 2013; Vaughan & Giovanello, 2010; Wongupparaj et al., 2015). Nevertheless, our finding of one common factor is consistent with previous results by Weigard et al. (2021), who also used drift rates and found that different cognitive control tasks loaded on only one common factor.

In the next step, we controlled for task-general processes included in drift rates (i.e., general speed of information uptake) and found that the latent common executive functions factor was fully accounted for by the task-general speed information uptake factor. In this model, no variance specific to executive functions remained. Hedge et al. (2022) found that inhibition tasks account for little common variance in inhibition processes but reflect consistent differences in task-general processing speed. They concluded that executive function processes should only be interpreted after controlling for task-general processes.Our results confirm the call by Hedge et al. (2022) to control the common variance among executive function tasks for task-general processes, suggesting that the observed common variance in the nine executive function tasks reflects nothing more than differences in the basic speed of information uptake. This is consistent with previous findings indicating that tasks supposedly measuring executive functions largely capture individual differences in the speed of information processing (Frischkorn & von Bastian, 2021; Frischkorn et al., 2019).

At this point we want to emphasize that drift rates are appropriate measures to separate task-general processes from domain-specific processes, and that these do not generally only yield one common speed of information uptake factor. In a recent paper, Lerche et al. (2020) demonstrated that drift rates can reflect different domain-general and domain-specific processes by showing that drift rates estimated from a battery of RT tasks differing in their complexity and in their content (figural, numeric, and verbal) reflected distinct factors of task-general as well as complexity- and content-specific variances. These results show that drift rates do not only measure the basic speed of information uptake, but that they may also reflect distinct processes. In consequence, our finding that drift rates in executive function tasks only represent task-general speed of information uptake is not a methodological artifact. Instead, it indicates that executive function tasks measure almost exclusively differences in basic speed of information uptake. Thus, executive function factors observed in previous studies likely only reflected individual differences in general processing speed.

Furthermore, we found that individual differences in general speed of information uptake were moderately correlated to intelligence and WMC. It is well known that information processing speed in elementary cognitive tasks is related to cognitive abilities (Doebler & Scheffler, 2016; Schubert et al., 2015, 2017, 2022b; Sheppard & Vernon, 2008). Several papers showing correlations between higher-order cognitive abilities and executive functions using heterogenous measurement scores (Benedek et al., 2014; Conway et al., 2021; Friedman et al., 2006, 2008; Wongupparaj et al., 2015) may have overestimated the relation between executive functions and higher-order cognitive abilities, because it is plausible that they largely estimated the relations between information processing speed and higher-order cognitive abilities.

Implications for future research on executive functions

From our findings we derive three possible consequences: First, there may be no individual differences in cognitive abilities that are specific for executive functions. Second, it may be necessary to think about the coherence of specific executive functions on a theoretical level. Third, many of the indicator scores and tasks used so far may be inappropriate to capture individual differences in executive functions.

Our first conclusion is contrary to experiences we make in daily life. Every day, we experience situations in which we have the feeling that we are using executive processes. We must ignore irrelevant or distracting information to navigate traffic safely, we must update our memory content when playing a memory card matching game, and, when multitasking, we must shift between different tasks. As already mentioned at the beginning of the introduction, we consider executive functions as abilities. If executive functions exist in the sense of abilities, we have to assume that people differ in these abilities. The word "ability" is defined "as the quality or state of being able" (Merriam-Webster, 2022, 13. July), which is characterized by variation, because qualities and states vary between individuals. It is also well known that an extremely low level of executive abilities is associated with unfavorable or pathological outcomes, as—for example—extremely low inhibition abilities are associated with attention-deficit/hyperactivity disorder (ADHD; e.g., Wodka et al., 2007). In consequence, the dual-pathway model of ADHD describes poor inhibitory control as a central aspect of ADHD symptoms (Sonuga-Barke, 2002). Taken together, it is hard to believe that people do not differ in their executive functions.

Perhaps we should reconsider the coherence of specific executive functions on the theoretical level. Von Bastian et al. (2020) showed in their review that executive function tasks yield on average only small correlations among each other (median $r = 0.16$). Specifically, inhibition tasks did usually show absent or very small correlations with each other (see von Bastian et al., 2020), which suggests that inhibition may be needed to be defined more precisely and possibly be split into distinct abilities. Rey-Mermet et al. (2018) already demonstrated that two distinct abilities of inhibition exist, the inhibition of prepotent responses and the inhibition of distractor interference. However, a subsequent study could not replicate the proposed two-factorial solution with adequate model fit (Gärtner & Strobel, 2021). In our study, the inhibition tasks showed only small to absent correlations (from $r = 08.$ to $r = 0.17$) when measured with RT-differences scores. This suggests that the executive processes contributing to performance in the inhibition tasks in our study may reflect distinct abilities, which could be one reason for why we did not find a coherent factor of inhibition when using RT-difference scores as indicator variables. Moreover, it is possible that the other executive functions also reflect a more differentiated pattern of the underlying abilities. Future research should reflect the divergence of executive functions on a theoretical level.

Alternatively, executive function tasks may be inappropriate to capture individual differences in executive functions. In our study, task-general speed of information uptake fully accounted for the shared variance between different executive function tasks. It seems that the classical executive function tasks capture to large parts task-general processes and no variance specific to executive functions. Therefore, we as a field should create new tasks or develop new cognitive mathematical models to better measure individual differences in these executive function abilities.

The development of new tasks to measure executive function abilities more validly

Recent studies have proposed developing and modifying executive function tasks to better capture individual differences in executive functions. Draheim et al. (2021) developed a battery of new and modified (Flanker and Stroop) inhibition tasks and compared them with different classical inhibition tasks. In the newly developed tasks, properties of the task (e.g., presentation time of the stimulus or the maximally allotted response time) adjusted dynamically as a function of participants' performance in previous trials. If the performance was good enough, the presentation times of stimuli or response deadlines were lowered, otherwise they were raised. In this adaptive staircase procedure, the authors used in some tasks the individually calibrated presentation times and in other tasks the individually calibrated response deadlines as dependent variable. Draheim et al. (2021) found substantial intercorrelations between most of these tasks and subsequently a coherent latent factor of attentional control, which is a construct closely related to executive functions. In addition, this latent factor correlated with WMC and intelligence and these correlations could not be explained by task-general processing speed. Further research reported additional evidence for the validity of this battery of novel executive function tasks by finding a common factor of executive processes independent of task-general processing speed (Burgoyne et al., 2022; Draheim et al., 2023). These modified tasks were highly reliable (all estimates ≥ 0.86) and fast to administer (see: Burgoyne et al., 2022).

However, there is a particular aspect that must be taken into account when discussing our findings with regard to the framework proposed by Draheim et al., (2021, 2023). It is possible that the discrepancy between both studies regarding the correlations between executive function abilities and processing speed stems from differences in the conceptualization of the mental speed factors and the specific tasks used to measure mental speed. Draheim and colleagues typically employed tasks in which participants compared patterns of stimuli and decided whether two patterns were equal or different. In contrast, our tasks required participants to make elementary decisions based on a currently presented stimulus. Both sets of tasks can be considered as measures of mental speed. However, the focus lies on different aspects of mental speed. The distinction between both concepts can be illustrated based on the Cattell-Horn-Carroll (CHC) model (Carroll, 1993). In the CHC model, Draheim and colleagues' tasks align with the processing speed factor (*Gs*), while our tasks align with the reaction and decision speed factor (*Gt*). Both of these factors belong to the CHC model's broader abilities in Stratum II and are considered to have separate contributions to general intelligence (e.g., Carroll, 1993). Because our battery of elementary cognitive tasks corresponds to the abilities represented in Carrol's *Gt* factor, the common factor of executive function tasks was completely explained by processing speed. All the variance shared among executive function tasks is essentially attributed to task-general information processing and decision-making abilities. In contrast, Draheim and colleagues' elementary cognitive tasks are measuring perceptual and clerical speed. These differences in the conceptualizations between Draheim et al. (2021, 2023) and our lab may account for the different empirical observations. It remains unclear to which extent the cognitive control factor by Draheim et al. (2021, 2023) diverges from a speed factor when measured with our processing speed tasks.

Our findings shed alarming light on classical executive function tasks, revealing that the shared variance among these tasks primarily represents task-general processing and decision-making abilities. Nevertheless, the work by Draheim et al. (2021, 2023) demonstrates that the development of novel measures and new tasks are promising approaches to make progress in the research of measuring individual differences in executive functions. Because it would be important to demonstrate that their tasks do not only measure processing speed as measured by the tasks included in the present study, it is obvious that more research is needed.

Cognitive mathematical modeling approaches to measure executive function abilities more validly

In addition to novel tasks and measurement scores, cognitive mathematical modeling approaches could also be a promising approach to validly measure individual differences in executive functions. A recent study validated the model parameters of the dual-stage two-phase model by Hübner et al. (2010)—a specific cognitive model to capture inhibition abilities in the Arrow Flanker Task—with inhibition-related electrophysiological correlates and found meaningful correlations between model parameters and event-related potential components (Schubert et al., 2022a). Jointly, the process parameters explained 37% variance in higher-order cognitive abilities (Schubert et al., 2022a). However, the authors did not control the model parameters for the influence of task-general processes. It therefore remains open to which degree the parameters reflected task-general processing efficiency. Nevertheless, the findings by Schubert et al., (2022a) demonstrate that cognitive mathematical models could be a fruitful way to capture individual differences in executive abilities. The use of cognitive mathematical model parameters and the development of specific cognitive mathematical models should be further promoted in the field of executive functions research. However, it is necessary to demonstrate that model parameters possess divergent validity to basic speed of information uptake.

Limitations

One major limitation of the diffusion modeling approach implemented in the present study is that the standard drift–diffusion model is not ideally suited to model RT distributions associated with incorrect responses in inhibition tasks. Because the drift rate is assumed to be constant over the course of a single trial, it is unable to account for the characteristic data pattern observed in conflict tasks, specifically the occurrence of faster errors in incongruent trials compared to correct responses (White et al., 2011). To address this limitation of the standard drift–diffusion model, models with time-varying drift rates like the diffusion model for conflict tasks (Ulrich et al., 2015), the dual-stage two-phase model (Hübner et al., 2010), and the shrinking spotlight model (White et al., 2011) have been developed. Since these models assume drift rates that change over time, they are more appropriate to account for the characteristic data pattern observed in conflict tasks. However, in the present study, it was not feasible to estimate these models with time-varying drift rates instead of the standard diffusion model, as we included not only conflict (inhibition) tasks but also updating and shifting tasks in our study. As a result, these models with time-varying drift rates could only be applied to a subset of our data and not to data from all nine executive function tasks. Nonetheless, we have confidence that our conclusions were not influenced by using the standard diffusion model, as our findings align with those of a previous study that used the diffusion model for conflict tasks to estimate the correlation of conflict-related model parameters across four different conflict tasks (Hedge et al., 2022). Consistent with our results, this study also found only very low and statistically insignificant correlations across tasks, although these correlations may have been underestimated due to the low reliabilities of control-related model parameters.

Another limitation pertaining to the estimation of the diffusion model in the present study is that v reflects both RTs and accuracies jointly, since the RT distributions of both correct and incorrect responses are used for its estimation. However, this was not the case for all tasks in our study. For the Two Choice Reaction Time task, the drift rate parameter mainly reflected RT variance because accuracy rates were near ceiling (i.e., there was virtually no distribution of incorrect responses, mean correct responses 99.47%, $SD = 0.98\%$). That is why the parameter recovery revealed no correlation between predicted and observed accuracies ($r = 0.05$). In comparison, the quantiles of the RT-distribution of the Two Choice Reaction Time task were recovered with high precision (range from $r = 0.98$ to $r = 0.99$; see also Fig. S4 in the supplementary materials). Nevertheless, the drift rates of the Two Choice Reaction Time task were substantially correlated with the drift rates of both other elementary cognitive tasks, $r = 0.35$ to $r = 0.37$. These manifest correlations were comparable to the correlation between the drift rates in those two other tasks ($r = 0.39$) and indicate that the drift rates of the Two Choice Reaction Time task showed convergent validity to the drift rates of the two other elementary cognitive tasks. For the other 11 tasks, we observed correlations between $r = 0.47$ to $r = 0.99$ for RTs and accuracies between predicted and observed scores. All in all, we can be relatively certain that our models yielded valid parameter estimates.

Finally, we examined a sample of $N = 148$ participants, which is a sufficient sample size as we needed a minimum sample size of $N = 95$ to test the hypothesis of close fit. However, given the uncertainty of correlations, examining larger groups of individuals would strengthen the robustness of our correlational findings (Kretzschmar & Gignac, 2019; Schönbrodt & Perugini, 2013). Therefore, future research should try to replicate the absolute magnitude of correlations in our study as their estimations had a relatively large degree of uncertainty.

Conclusion

In our present study, we examined the factor structure of the three executive functions by Miyake et al. (2000). We used a cognitive mathematical modeling approach to overcome the problems associated with the inconsistent use of accuracy vs. RT-based scores as indicator variables and the use of manifest difference scores, which can sometimes cause psychometric problems. Applying the drift–diffusion model, we found a one-factorial structure of executive function tasks. However, in this analysis, we used only the drift rates from the conditions with greater processing demands. Because drift rates in these conditions were affected by both task-general and executive function processes, the latent common executive functions factor reflected individual differences in both types of processes. After controlling for individual differences in these task-general processes, we observed no unique variance specific to executive functions. This indicates that the covariance between different executive function tasks can be fully accounted for by individual differences in the general speed of information uptake, which was moderately related to higher-order cognitive abilities. Applying this drift–diffusion model account thus shed alarming light on tasks supposedly measuring executive functions. We observed no variance specific to executive functions that was independent of the general speed of information uptake. Thus, the development or modification of executive function tasks is necessary to capture individual differences in executive functions reliably and validly, assuming that such differences exist.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00426-023-01924-7.

Acknowledgements We want to thank our research assistants—Johanna Hein, Florian Kaulhausen, Jan Göttmann, and Larissa Kunoff—who supported us with the recruitment of participants, data collection, project organization, and data management. In addition, the first author (C.L.) would like to express a very special thanks to Birgit Zwickert-Biniasch and Jörg Biniasch, who provided for him a home office working space for a few months during the COVID-19 pandemic. This research was supported by the German Research Foundation (DFG) [grant numbers SCHU 3266/1-1 and SCHU 3266/2-1]. We declare no conflicts of interest. The preprocessed data and scripts supporting the findings of this study are available in the Open Science Framework repository at https://osf.io/6c4pu/ and the raw data and the materials are available in the Open Science Framework repository at https://osf.io/4pvz3/.

Authors' contributions CL: conceptualization, methodology, software, data curation, writing—original draft preparation, visualization, investigation, project administration, formal analyses. GF: conceptualization, methodology, supervision, validation, writing—reviewing and editing. DH: supervision, validation, writing—reviewing and editing, resources. KS: validation, writing- reviewing and editing, project administration. A-LS**:** conceptualization, methodology, supervision, validation, writing—reviewing and editing, funding acquisition, project administration, formal analyses.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was supported by the German Research Foundation (DFG) [grant numbers SCHU 3266/1-1 and SCHU 3266/2-1].

Data availability We provide access to the preprocessed data and the statistical analysis code used for this paper via the Open Science Framework (https://osf.io/6c4pu/). In addition, we provide access to the raw data and to the materials via the Open Science Framework (https://osf.io/4pvz3/; except for the materials of the BIS, which are commercially licensed). Neither the study nor the analyses were preregistered.

Declarations

Conflict of interests We declare no conflicts of interest.

Ethical approval The study was approved by the ethics committee of the faculty of behavioral and cultural studies of Heidelberg University (approval number: Löf2019/1-3). All procedures were conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013). At the beginning of the first study session, participants signed an informed consent. We have not preregistered this study.

References

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*. https://doi.org/10.1016/j.intell.2020.101440

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. https://doi.org/10.1016/j.intell.2014.05.007

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*(2), 563–570. https://doi.org/10.3758/BRM.42.2.563

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2022). *Nature and Measurement of Attention Control*. PsyArXiv. https://doi.org/10.31234/osf.io/7y5fp

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2. Aufl.). Springer-Verlag. https://doi.org/10.1007/b97636

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge University Press.

Chen, D. D., Chan, J. S. Y., Pei, L., & Yan, J. H. (2012). Transfer Effects of Manipulating Temporal Constraints on Learning a Two-Choice Reaction Time Task with Low Stimulus-Response Compatibility. *Perceptual and Motor Skills, 115*(2), 415–426. https://doi.org/10.2466/22.23.25.PMS.115.5.415-426

Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J.-P. (2021). Individual Differences in Attention and Intelligence: A United Cognitive/Psychometric Approach. *Journal of Intelligence, 9*(3), 34. https://doi.org/10.3390/jintelligence9030034

Cronbach, L. J., & Furby, L. (1970). How we should measure „change": Or should we? *Psychological Bulletin, 74*(1), 68–80. https://doi.org/10.1037/h0029382

Doebler, P., & Scheffler, B. (2016). The relationship of choice reaction time variability and intelligence: A meta-analysis. *Learning and Individual Differences, 52*, 157–166. https://doi.org/10.1016/j.lindif.2015.02.009

Donders, F. C. (1869). On the speed of mental processes. *Acta Psychologica, 30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin, 145*(5), 508–535. https://doi.org/10.1037/bul0000192

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General, 150*(2), 242–275. https://doi.org/10.1037/xge0000783

Draheim, C., Tshukara, J. S., & Engle, R. W. (2023). Replication and extension of the toolbox approach to measuring attention control. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02140-2

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience, 14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science, 17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in Origin. *Journal of Experimental Psychology. General, 137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology, 47*(5), 1410–1430. https://doi.org/10.1037/a0023750

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

Application. *Journal of Intelligence*. https://doi.org/10.3390/jintelligence6030034

Frischkorn, G. T., & von Bastian, C. C. (2021). In Search of the Executive Cognitive Processes Proposed by Process-Overlap Theory. *Journal of Intelligence*. https://doi.org/10.3390/jintelligence9030043

Frischkorn, G., Schubert, A.-L., & Hagemann, D. (2019). Processing speed, working memory, and executive functions: Independent or inter-related predictors of general intelligence. *Intelligence, 75*, 95–110. https://doi.org/10.1016/j.intell.2019.05.003

Gärtner, A., & Strobel, A. (2021). Individual Differences in Inhibitory Control: A latent Variable Analysis. *Journal of Cognition*. https://doi.org/10.5334/joc.150

Gevins, A., Smith, M. E., Le, J., Leong, H., Bennett, J., Martin, N., McEvoy, L., Du, R., & Whitfield, S. (1996). High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroencephalography and Clinical Neurophysiology, 98*(4), 327–348. https://doi.org/10.1016/0013-4694(96)00288-X

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition, 69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Harrell, Jr., F. E. (2019). Package 'hmisc'. *CRAN2018, 2019*, 235–236.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2022). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(10), 1448–1469. https://doi.org/10.1037/xlm0001028

Himi, S. A., Bühner, M., Schwaighofer, M., Klapetek, A., & Hilbert, S. (2019). Multitasking behavior and its related constructs: Executive functions, working memory capacity, relational integration, and divided attention. *Cognition, 189*, 275–298. https://doi.org/10.1016/j.cognition.2019.04.010

Himi, S. A., Bühner, M., & Hilbert, S. (2021). Advancing the Understanding of the Factor Structure of Executive Functioning. *Journal of Intelligence, 9*(1), 16. https://doi.org/10.3390/jintelligence9010016

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), 1–55. https://doi.org/10.1080/10705519909540118

Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review, 117*(3), 759–784. https://doi.org/10.1037/a0019471

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology, 22*(4), 508–522. https://doi.org/10.1037/0894-4105.22.4.508

Ishihara, S. (2000). *The series of plates designed as a test for colour deficiency: 38 plates edition.* Kanehara.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology, 108*(2), 187–218. https://doi.org/10.1037/a0038557

Jäger, A. O., Süß, H., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe. https://madoc.bib.uni-mannheim.de/14578/

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Hrsg.), *Variation in Working Memory* (S. 21–48). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0002

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Kelley, K. (2007). Methods for the Behavioral, Educational, and Social Sciences: An R package. *Behavior Research Methods, 39*(4), 979–984. https://doi.org/10.3758/BF03192993

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology, 63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.

Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry, 27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality, 80*, 17–22. https://doi.org/10.1016/j.jrp.2019.03.007

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37*(4), 347–364. https://doi.org/10.1016/j.intell.2009.02.003

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

Lerche, V., von Krause, M., Voss, A., Frischkorn, G. T., Schubert, A.-L., & Hagemann, D. (2020). Diffusion modeling and intelligence: Drift rates show both domain-general and domain-specific relations with intelligence. *Journal of Experimental Psychology: General, 149*(12), 2207–2249. https://doi.org/10.1037/xge0000774

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571–585. https://doi.org/10.3758/BRM.42.2.571

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Merriam-Webster. (n.d.). Ability. In *Merriam-Webster.com dictionary*. Retrieved July 13, 2022, from https://www.merriam-webster.com/dictionary/ability

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin and Review, 20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

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. https://doi.org/10.1016/S0160-2896(99)80058-0

Pedersen, T. L. (2020). Patchwork: The composer of plots. *R Package Version, 1*(1), 182.

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review, 74*(5), 392–409. https://doi.org/10.1037/h0024913

R. Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. URL https://www.R-project.org/

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

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence, 36*(1), 10–17. https://doi.org/10.1016/j.intell.2006.12.002

Revelle, W. (2020). psych: Procedures for Psychological, Psychometric, and Personality Research. In *Northwestern University, Evanston, Illinois. R package version 2.0.12,*. https://CRAN.R-project.org/package=psych.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 501–526. https://doi.org/10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General, 148*(8), 1335–1372. https://doi.org/10.1037/xge0000593

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1–36.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin and Review, 26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Sadus, K., Schubert, A.-L., Löffler, C., & Hagemann, D. (2023). An explorative multiverse study for extracting differences in P3 latencies between young and old adults. *Psychophysiology*. https://doi.org/10.1111/psyp.14459

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating: When flanker meets the n-back. *Psychophysiology, 52*(10), 1293–1304. https://doi.org/10.1111/psyp.12500

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*(3), 414–429. https://doi.org/10.1037/0096-3445.136.3.414

Schnitzspahn, K. M., Stahl, C., Zeintl, M., Kaller, C. P., & Kliegel, M. (2013). The role of shifting, updating, and inhibition in prospective memory performance in young and older adults. *Developmental Psychology, 49*(8), 1544–1553. https://doi.org/10.1037/a0030579

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

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. https://doi.org/10.1016/j.intell.2015.05.002

Schubert, A.-L., Frischkorn, G., Hagemann, D., & Voss, A. (2016). Trait characteristics of diffusion model parameters. *Journal of Intelligence, 4*(3), 7. https://doi.org/10.3390/jintelligence4030007

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

Schubert, A.-L., Löffler, C., & Hagemann, D. (2022a). A neurocognitive psychometrics account of individual differences in attentional control. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001184

Schubert, A.-L., Löffler, C., Hagemann, D., & Sadus, K. (2022b). How robust is the relationship between neural processing speed and cognitive abilities? *Psychophysiology, Advance Online Publication*. https://doi.org/10.1111/psyp.14165

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

Sonuga-Barke, E. J. S. (2002). Psychological heterogeneity in AD/HD—a dual pathway model of behaviour and cognition. *Behavioural Brain Research, 130*(1–2), 29–36. https://doi.org/10.1016/S0166-4328(01)00432-6

Spearman, C. (1904). „General Intelligence, " Objectively Determined and Measured. *The American Journal of Psychology, 15*(2), 201–292. https://doi.org/10.2307/1412107

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General, 143*(2), 850–886. https://doi.org/10.1037/a0033981

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*(4), 421–457.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance, 13*(1), 89–103. https://doi.org/10.1037/0096-1523.13.1.89

Tipper, S. P., & Cranston, M. (1985). Selective Attention and Priming: Inhibitory and Facilitatory Effects of Ignored Primes. *The Quarterly Journal of Experimental Psychology Section A, 37*(4), 591–611. https://doi.org/10.1080/14640748508400921

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology, 78*, 148–174. https://doi.org/10.1016/j.cogpsych.2015.02.005

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging, 25*(2), 343–355. https://doi.org/10.1037/a0017729

von Bastian, C. C., Blais, C., Brewer, G., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., Wiemers, E. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical,* *methodological, and analytical considerations*. PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

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

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

Weigard, A., Clark, D. A., & Sripada, C. (2021). Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition*. https://doi.org/10.1016/j.cognition.2021.104818

White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology, 63*(4), 210–238. https://doi.org/10.1016/j.cogpsych.2011.08.001

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software, 4*(43), 1686.

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

Wodka, E. L., Mahone, E. M., Blankner, J. G., Larson, J. C. G., Fotedar, S., Denckla, M. B., & Mostofsky, S. H. (2007). Evidence that response inhibition is a primary deficit in ADHD. *Journal of Clinical and Experimental Neuropsychology, 29*(4), 345–356. https://doi.org/10.1080/13803390600678046

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

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA, 310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Supplementary materials

Table S1 *Presentation times of the memory-set stimuli, probe stimuli, and intervals between stimuli and trials in the different tasks*

Task name	Process	Fixation cross	Inter-stimulus interval	Memory presentation time	Cue presentation time	Probe presentation time	Inter-trial interval
Negative priming task	inhibition	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Flanker task	inhibition	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Stroop task	inhibition	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Keep track task	updating	400 - 600 ms	400 - 600 ms	1000 ms	800 - 1200 ms	1000 - 3000 ms	1000 - 1500 ms
Running span task	updating	400 - 600 ms	400 - 600 ms	1000 ms	800 - 1200 ms	1000 - 3000 ms	1000 - 1500 ms
N-Back task	updating	---	400 - 600 ms	---	---	1500 ms	---
Number-letter task	shifting	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Switching task	shifting	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Global-local task	shifting	400 - 600 ms	400 - 600 ms	---	---	1000 - 3000 ms	1000 - 1500 ms
Two-choice-RT task	speed	1000 - 1500 ms	---	---	---	1000 - 3000 ms	1000 - 1500 ms
Sternberg task	speed	1000 - 1500 ms	400 - 1000 ms	1000 ms	1800 - 2200 ms	1000 - 3000 ms	1000 - 1500 ms
Posner task	speed	1000 - 1500 ms	---	---	---	1000 - 3000 ms	1000 - 1500 ms

Note. Probe stimuli were presented until participants responded; If participants' response was faster than 1000 ms the stimulus remained till 1000 ms were reached; The stimulus disappeared after 3000 ms if the participants did not respond (after 1500 ms in the N-Back task); ms = milliseconds.

Table S2 *Correlations between all variables (heterogeneous measurement scores)*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	Negative Pr. task, RT difference																						
2	Flanker task, RT difference	**.17**																					
3	Stroop task, RT difference	.08	**.15**																				
4	Negative Pr. task, RT priming cond.	**.43**	**.20**	**.17**																			
5	Flanker task, RT incong. cond.	**.22**	**.42**	**.23**	**.72**																		
6	Stroop task, RT incong. cond.	**.25**	**.20**	**.60**	**.67**	**.59**																	
7	Keep Track, prop. updating cond.	-.05	-.08	.09	**-.34**	**-.20**	**-.21**																
8	Running Span, prop. updating cond.	-.01	-.08	.04	**-.20**	-.10	-.11	**.43**															
9	N-Back task prop	-.09	**-.17**	.02	**-.30**	**-.24**	**-.22**	**.35**	**.32**														
10	Number Letter task, RT difference	.01	.01	.09	.15	**.25**	**.23**	-.11	-.09	**-.17**													
11	Switching task, RT difference	.13	.12	**.18**	**.22**	**.19**	**.24**	**-.21**	.01	**-.28**	**.45**												
12	Global Local task, RT difference	.00	.09	.15	**.17**	**.23**	**.25**	-.12	-.07	-.13	**.44**	**.41**											
13	Two Choice RT task, RT	**.17**	.11	.16	**.71**	**.66**	**.52**	**-.25**	-.08	**-.21**	.09	.12	**.22**										
14	Sternberg task, RT	.14	**.19**	.10	**.60**	**.64**	**.41**	**-.26**	-.14	**-.33**	.14	**.21**	**.22**	**.53**									
15	Posner task, RT	**.18**	.06	.15	**.74**	**.62**	**.54**	**-.42**	**-.27**	**-.29**	**.17**	**.22**	.12	**.70**	**.59**								
16	Memory updating	-.08	-.16	.03	**-.30**	**-.38**	**-.20**	**.25**	**.46**	**.42**	-.15	-.05	-.17	**-.24**	**-.33**	**-.38**							
17	Binding	**-.18**	-.16	.10	**-.34**	**-.38**	**-.22**	**.25**	**.32**	**.35**	-.07	-.02	-.10	**-.22**	**-.40**	**-.31**	**.70**						
18	Operation span	.03	.01	.09	-.12	**-.24**	-.06	.09	**.28**	**.26**	-.03	.00	-.07	**-.19**	**-.29**	**-.18**	**.58**	**.50**					
19	Sentence span	.03	.03	.04	-.06	-.12	-.05	.14	**.33**	**.24**	-.12	.02	-.09	-.14	-.14	-.16	**.61**	**.35**	**.66**				
20	BIS-PC	-.04	-.02	-.08	**-.26**	**-.25**	**-.23**	**.23**	**.31**	**.43**	-.05	-.13	-.05	-.16	**-.31**	**-.36**	**.60**	**.53**	**.38**	**.41**			
21	BIS-PS	-.08	-.09	-.02	**-.36**	**-.39**	**-.28**	**.33**	**.25**	**.41**	-.16	-.13	**-.20**	**-.34**	**-.35**	**-.38**	**.42**	**.44**	**.20**	**.22**	**.50**		
22	BIS-M	.08	-.13	-.02	-.13	**-.21**	**-.25**	**.20**	**.26**	**.34**	-.15	**-.22**	**-.24**	-.10	-.12	**-.20**	**.34**	**.34**	**.23**	**.27**	**.43**	**.43**	
23	BIS-C	-.01	-.07	-.13	**-.21**	-.11	-.08	.05	.08	.12	.04	-.02	.08	-.03	**-.17**	**-.20**	.13	**.19**	.09	.13	**.35**	**.34**	.13

Note. Heterogeneous measurement scores; BIS-PC = processing capacity scale of the Berlin Intelligence Structure Test; BIS-PS = processing speed scale of the Berlin Intelligence Structure Test; BIS-M = memory scale of the Berlin Intelligence Structure Test; BIS-C = creativity scale of the Berlin Intelligence Structure Test; 1-3 RT difference scores; 4-6 mean RT of the incongruent condition; 7-9 arcsine transformed proportion correct scores; 10-12 RT difference scores; 13-15 mean RT; 16-19 percentage correct; 20-23 scale scores; Significant correlations ($p < .05$) are presented in bold.

Table S3 *Descriptive statistics of the drift-diffusion model parameters*

	a mean	a *SD*	v mean	v *SD*	t_0 mean	t_0 *SD*	st_0 mean	st_0 *SD*
Negative priming task, priming cond.	1.28	0.29	3.62	0.94	0.42	0.07	0.13	0.08
Flanker task, incong. cond.	1.05	0.29	5.05	1.32	0.40	0.05	0.12	0.05
Stroop task, incong. cond.	1.44	0.37	2.60	0.74	0.55	0.12	0.28	0.15
Keep track, updating cond.	1.72	0.34	1.69	0.62	0.41	0.17	0.27	0.24
Running span, updating cond.	1.49	0.40	1.78	0.58	0.53	0.12	0.18	0.18
N-Back task	1.47	0.22	1.74	0.46	0.36	0.11	0.25	0.17
Number letter task, shifting cond.	1.67	0.43	2.29	0.92	0.39	0.15	0.18	0.26
Switching task, shifting cond.	1.74	0.44	2.16	0.85	0.37	0.11	0.23	0.26
Global local task, shifting cond.	2.01	0.37	1.65	0.51	0.62	0.25	0.44	0.40
Two choice RT task	0.95	0.24	6.25	1.58	0.30	0.04	0.08	0.05
Sternberg task	1.52	0.41	2.35	0.71	0.61	0.16	0.24	0.16
Posner task	1.38	0.35	3.27	0.74	0.49	0.07	0.19	0.09

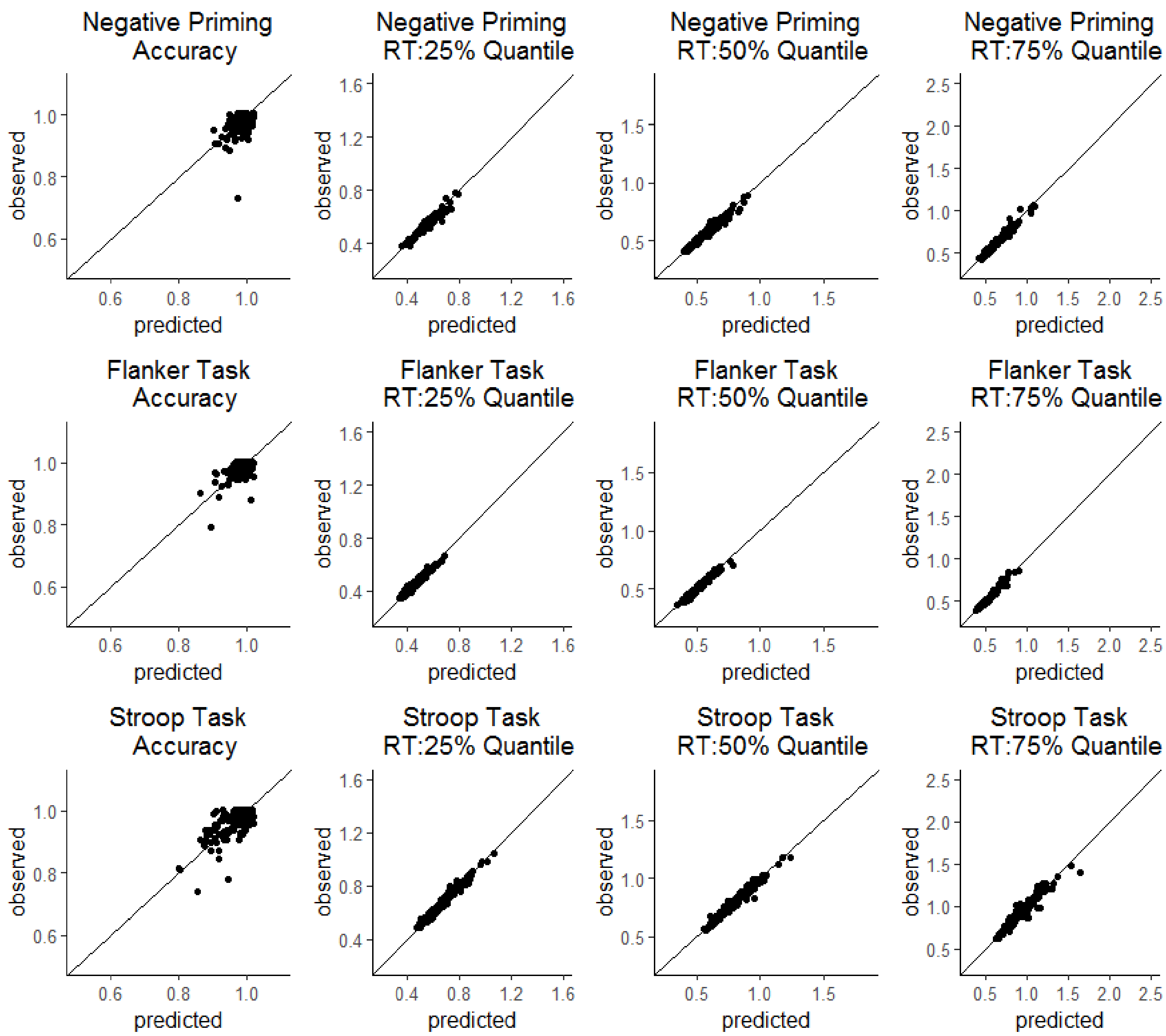
Note. All estimated drift-diffusion model parameters are displayed in this table; a = boundary separation parameter; v = drift parameter; t_0 = non-decision time parameter; st_0 = inter-trial variability of non-decisional components.

Table S4 *Correlations between all variables (drift rates v as homogenous measurement scores)*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	Negative Pr. task, priming cond.																			
2	Flanker task, incong. cond.	**.32**																		
3	Stroop task, incong. cond.	.05	**.30**																	
4	Keep Track, updating cond.	**.34**	**.29**	.07																
5	Running span, updating cond.	**.36**	**.24**	.04	**.35**															
6	N-Back task	**.27**	.15	-.01	**.33**	**.26**														
7	Number Letter task, shifting cond.	**.39**	**.34**	**.21**	**.38**	**.24**	**.24**													
8	Switching task, shifting cond.	**.42**	**.35**	**.34**	**.42**	**.35**	**.26**	**.60**												
9	Global Local task, shifting cond.	**.18**	**.37**	**.33**	.17	**.26**	**.26**	**.47**	**.53**											
10	Two Choice RT task	**.34**	**.44**	**.21**	**.38**	**.36**	**.21**	**.45**	**.57**	**.29**										
11	Sternberg task	**.30**	**.34**	.13	**.49**	**.53**	**.41**	**.44**	**.38**	**.35**	**.37**									
12	Posner task	**.34**	**.34**	.14	**.39**	**.36**	**.25**	**.32**	**.31**	.15	**.35**	**.39**								
13	Memory updating	**.17**	.11	.02	**.25**	**.46**	**.38**	.16	**.31**	**.25**	**.24**	**.43**	**.23**							
14	Binding	**.19**	.03	-.12	**.32**	**.29**	**.27**	.03	**.22**	.09	**.19**	**.24**	**.23**	**.70**						
15	Operation span	.02	-.02	.01	.14	**.32**	**.24**	.02	**.24**	.02	.10	**.29**	.03	**.58**	**.50**					
16	Sentence span	-.02	.03	.04	.05	**.23**	.17	.15	**.25**	.15	.09	**.24**	.05	**.61**	**.35**	**.66**				
17	BIS-PC	.14	**.17**	.14	**.22**	**.23**	**.31**	.07	**.23**	**.18**	.16	**.27**	**.27**	**.60**	**.53**	**.38**	**.41**			
18	BIS-PS	.16	.15	.05	**.28**	.17	**.44**	.12	**.32**	**.20**	**.21**	**.25**	**.21**	**.42**	**.44**	**.20**	**.22**	**.50**		
19	BIS-M	.14	**.19**	.16	.14	.14	**.20**	.11	**.32**	**.22**	.12	.16	.14	**.34**	**.34**	**.23**	**.27**	**.43**	**.43**	
20	BIS-C	.12	.01	.05	.03	.00	.11	-.01	.08	-.01	.07	.11	.05	.13	**.19**	.09	.13	**.35**	**.34**	.13

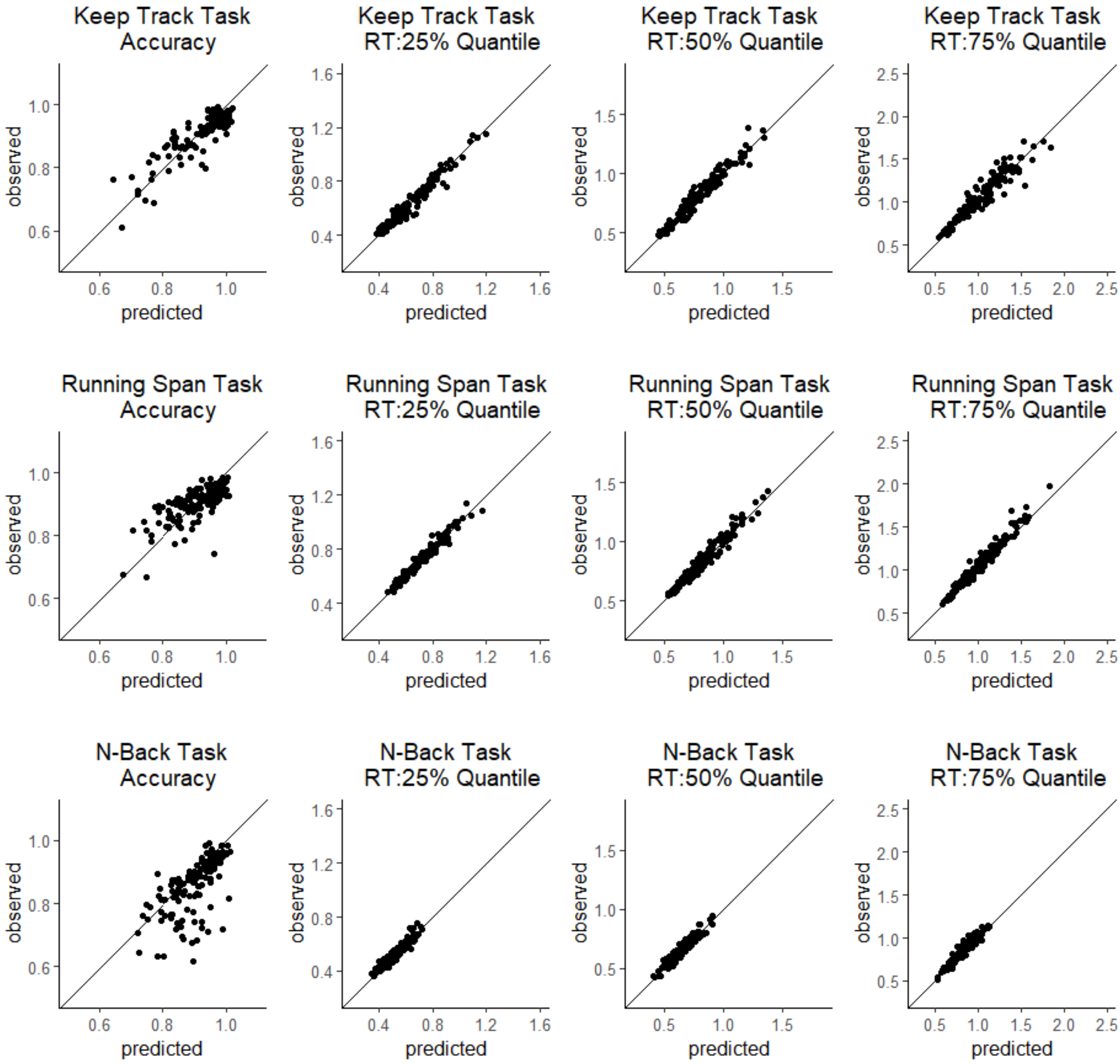
Note. Drift rates *v* as homogenous measurement scores; BIS-PC = processing capacity scale of the Berlin Intelligence Structure Test; BIS-PS = processing speed scale of the Berlin Intelligence Structure Test; BIS-M = memory scale of the Berlin Intelligence Structure Test; BIS-C = creativity scale of the Berlin Intelligence Structure Test; 1-12 drift rates *v*; 13-16 percentage correct; 17-20 scale scores; Significant correlations ($p < .05$) are presented in bold.

Figure S1 *QQ-plots for the assessment of model fit based on the comparison of statistics (accuracy rate, 25%, 50%, and 75% quantile) of the observed and predicted data for inhibition tasks*



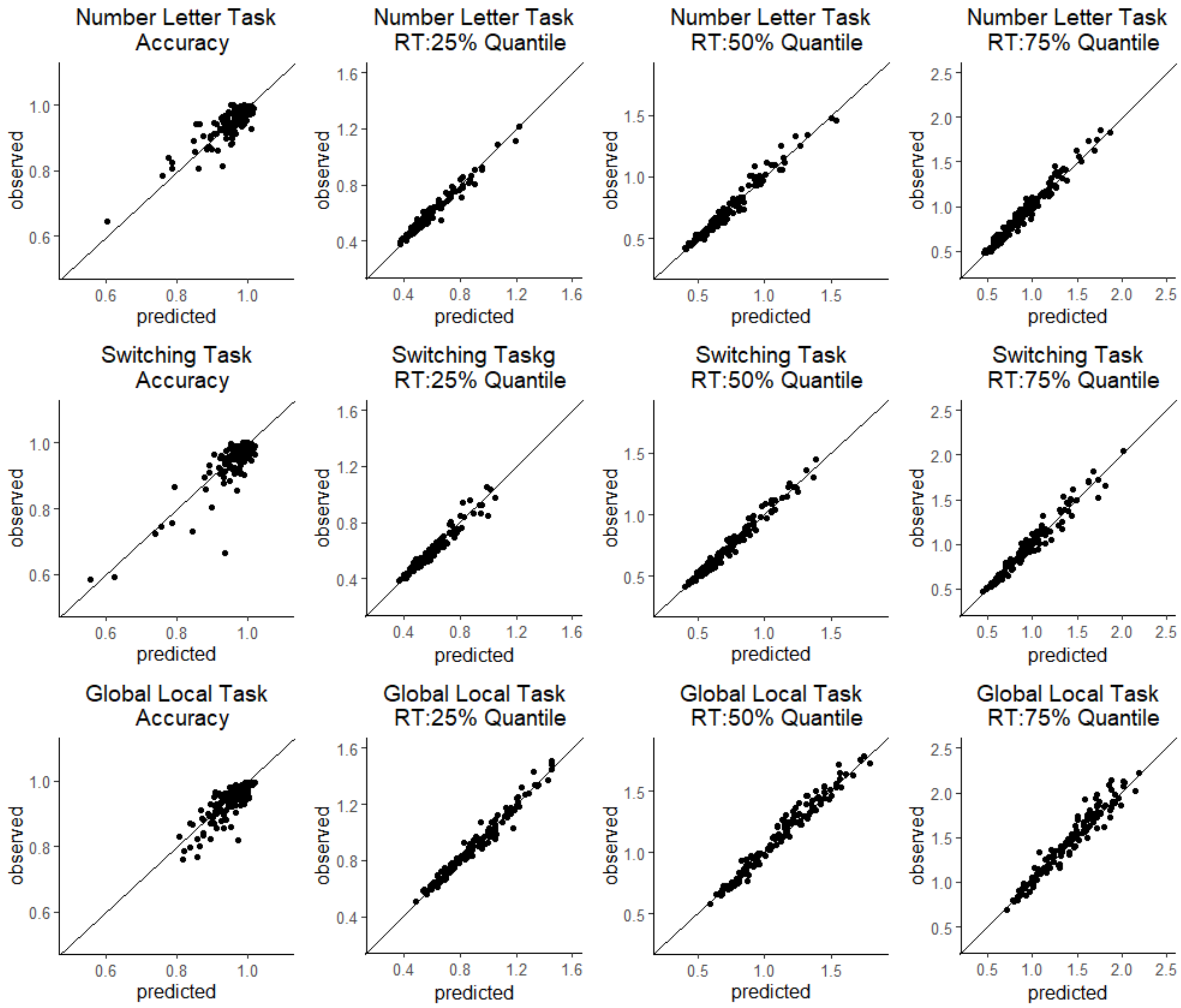
Note. Each data point represents one participant. The diagonal lines indicate perfect model fit.

Figure S2 *QQ-plots for the assessment of model fit based on the comparison of statistics (accuracy rate, 25%, 50%, and 75% quantile) of the observed and predicted data for updating tasks*



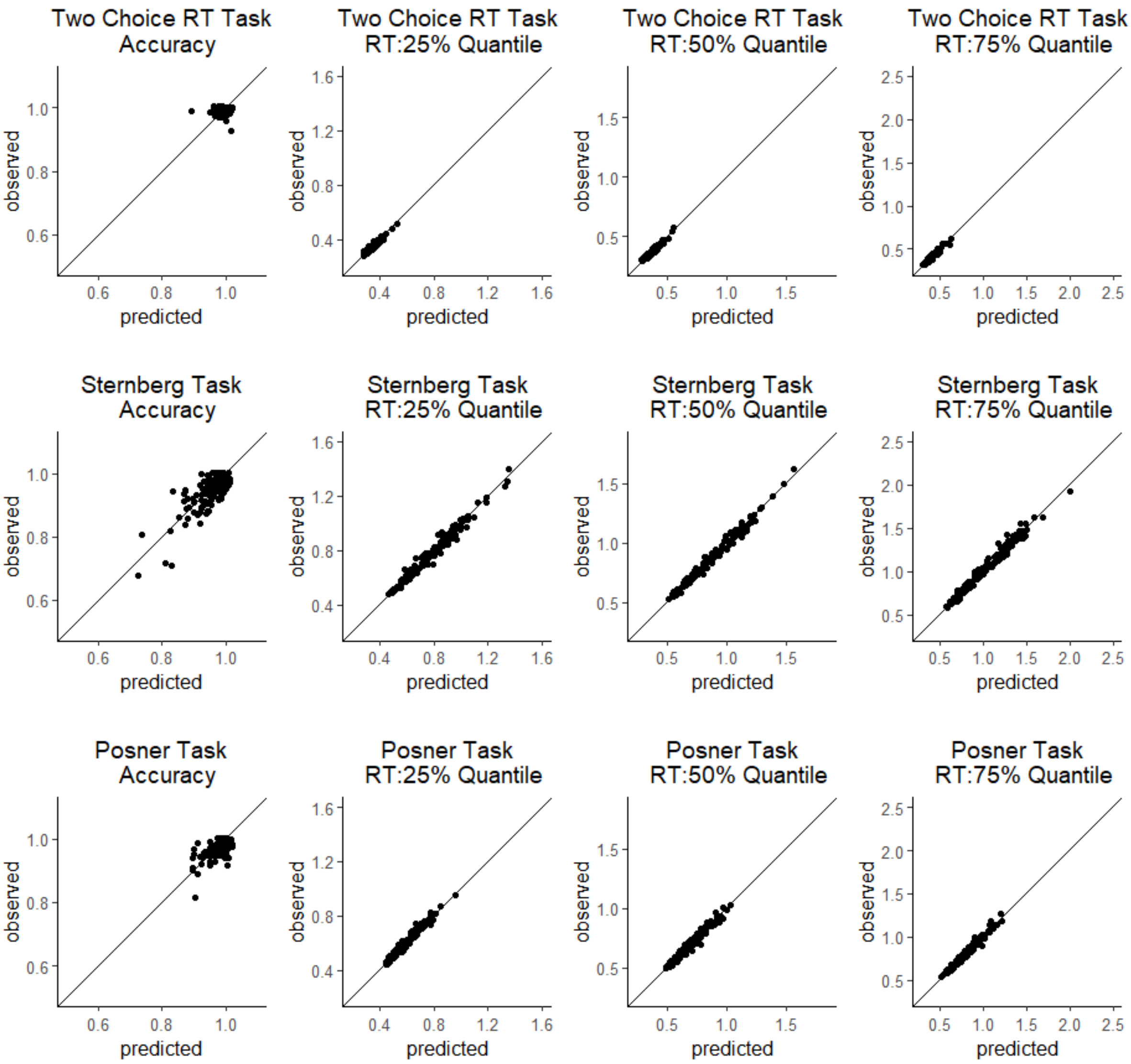
Note. Each data point represents one participant. The diagonal lines indicate perfect model fit.

Figure S3 *QQ-plots for the assessment of model fit based on the comparison of statistics (accuracy rate, 25%, 50%, and 75% quantile) of the observed and predicted data for shifting tasks*



Note. Each data point represents one participant. The diagonal lines indicate perfect model fit.

Figure S4 *QQ-plots for the assessment of model fit based on the comparison of statistics (accuracy rate, 25%, 50%, and 75% quantile) of the observed and predicted data for elementary cognitive tasks (information processing speed)*



Note. Each data point represents one participant. The diagonal lines indicate perfect model fit.

Appendix Manuscript III

The factor structure of executive functions measured with electrophysiological correlates: An event-related potential analysis

Christoph Löffler [ab], Kathrin Sadus [a], Gidon T. Frischkorn [c], Dirk Hagemann [a] & Anna-Lena Schubert [b]

[a] Institute of Psychology, Heidelberg University, Heidelberg, Germany

[b] Department of Psychology, University of Mainz, Mainz, Germany

[c] Department of Psychology, University of Zurich, Switzerland

Author Note

This work was supported by German Research Foundation (DFG) [grant number SCHU 3266/1-1]. We declare no conflicts of interest. The preprocessed ERP data and the scripts supporting our findings are available in the Open Science Framework repository at https://osf.io/a6zr9/ (Löffler et al., 2024). The raw data and the materials of this study are available in the Open Science Framework repository at https://osf.io/4pvz3/ (Löffler & Schubert, 2024). The results of this study were already presented at the Psychology and Brain 2022 in Freiburg, Germany, and at the Psychonomic Society's 64th Annual Meeting 2023 in San Francisco, USA. This manuscript is available as a preprint via the preprint server PsyArXiv with the DOI 10.31234/osf.io/kfqt4 under the following link: https://osf.io/preprints/psyarxiv/kfqt4

Correspondence concerning this article should be addressed to Christoph Löffler, Johannes Gutenberg University Mainz, Department of Psychology, Binger Straße 14-16, D-55122 Mainz, Germany, Phone: +49(0)6131-3982077,

E-mail: chloeffl@students.uni-mainz.de

Abstract

The three-factor model of executive functions is widely employed in cognitive control research. However, recent studies have revealed psychometric problems with commonly used difference scores in behavioral measures of executive functions. Examining behavioral scores, several studies were unable to find a coherent factor structure for executive functions or identify significant individual differences in specific executive function abilities. These findings have raised questions about the utility of established measurement scores for executive functions. Our study sought to reassess the three-factor model proposed by Miyake et al. (2000), employing event-related potentials (ERPs) from electroencephalography (EEG) as a means to directly probe underlying cognitive processes, leveraging the EEG's high temporal resolution. We conducted an analysis of the factor structure of the three executive functions (updating, shifting, and inhibition) in a sample of 148 participants. We employed Bayesian structural equation models to examine the relationships between the mean amplitudes of the N2 and P3 components, obtained from a battery of nine executive function tasks. Our results indicate that amplitudes of the ERP components measured in executive function tasks almost exclusively represent variance related to general processes rather than executive function-specific variance. Notably, no task demonstrated variance uniquely attributable to individual differences in executive function processes added through experimental manipulations. These results cast doubt on the validity of current executive function tasks in accurately reflecting individual differences in these processes.

Keywords: executive functions, EEG, event-related potentials, Bayesian structure equation modeling, cognitive abilities

The factor structure of executive functions measured with electrophysiological correlates: An event-related potential analysis

Our everyday life is full of different types of information. This information includes relevant content as well as distracting noise. It is therefore essential that we filter the relevant parts included in the information from our environment and adjust our ongoing actions based on the incoming information. For this, we use basic cognitive processes that help us to keep our attentional focus on the relevant information, to plan the next steps, to update new relevant information in our mind, and to shift our attentional focus between different ongoing actions. In cognitive psychology, such top-down regulated processes would be considered as individual abilities under the umbrella term of executive functions (EFs), also known as cognitive control, attentional control, executive attention, or cognitive control (Rey-Mermet, Gade, Souza, et al., 2019; von Bastian et al., 2020). A wide range of abilities are summarized under the term of EFs. In the present study, we focus on three commonly separated EFs, namely shifting, inhibition, and updating (Miyake et al., 2000). Shifting is the ability to shift between different tasks or mind sets, inhibition is one's ability to focus attention on the actual task and current task-goals while ignoring irrelevant information, and updating is one's ability to monitor the actual memory content and store new information in memory (Friedman et al., 2008; Friedman & Miyake, 2017; Miyake et al., 2000; Miyake & Friedman, 2012; Rey-Mermet et al., 2018). Miyake and colleagues (2000) have shown that while these three EFs are correlated with each other, there are also substantial parts of variance specific to the each of the three underlying EFs. These findings led to the "unity and diversity" framework of EFs, which means that the three EFs shifting, inhibition, and updating are considered separate but related processes. Subsequent research has provided further support for this description of unity and diversity in EFs and also found that EFs are related to individual differences in higher-order cognitive abilities such as intelligence and working memory (Burgoyne et al., 2023; Draheim et al., 2021, 2023; Friedman et al., 2006, 2008, 2011; Friedman & Miyake, 2017). Some theoretical accounts even claim that individual differences in EF abilities give rise to individual differences in intelligence and working memory capacity (WMC; Kane et al., 2008; Kovacs & Conway, 2016; see for an detailed overview: Mashburn et al., 2023).

Measuring executive functions and the problem of behavior-based scores

Despite the extensive attention that EFs have received in cognitive psychological research in the recent two decades, a growing body of research identified psychometric problems of measures of EFs that cast doubt on the three-factor structure of EFs (Hedge et al., 2018; Rey-Mermet et al., 2018; Rouder et al., 2023; Rouder & Haaf, 2019). Usually, a person's performance in an EF task (e.g., their inhibition ability) is measured by calculating the difference in their mean performances between two conditions. In general, EF tasks (e.g., an Arrow Flanker task; Eriksen & Eriksen, 1974) have at least two conditions, one baseline condition and one condition with greater processing demands. The baseline condition is designed in such a way that it does not require EF abilities at all or at least only to a negligible extent. In contrast, the condition with greater processing demands works almost identically to the baseline condition except that *one* manipulation has been added to the design. This manipulation requires a specific executive function to make the correct decision. Because this additional processing demand makes the task more difficult, participants typically show worse performances (i.e., lower accuracies and longer RTs) in conditions with greater processing demands in contrast to the baseline conditions. The performance increment between the two conditions is thought to represent individual's performance in the underlying EF (Donders, 1869).

While these difference scores are very useful for exploring general cognitive processes on the experimental level, it is important to note that they revealed psychometric problems in studies investigating individual differences in EFs (Hedge et al., 2018; von Bastian et al., 2020). This is due to the design of the tasks. In typical EF tasks with two conditions, individuals' performances in the two conditions are highly correlated. For example, when reanalyzing data across two sessions from the Flanker task and the Stroop task reported by Hedge et al. (2018), strong correlations between both conditions in each of the sessions were found ($\bar{r}$ =. 91). These very high correlations indicate that the different conditions in these inhibition tasks measure the same thing. When two conditions within a single task exhibit correlations exceeding r = .90, only a small percentage of the variability between the conditions remains unexplained. Hence, most of the cognitive processes involved in solving an EF task are needed in both conditions, as indicated by the very high correlations between performances in both conditions. However, researchers are primarily interested in the remaining unexplained variance, which is considered specific to EFs and represents individual differences in the experimental effect. For this reason, researchers calculate difference scores.

However, such difference scores capture not only EF-specific variance but also error variance. Consequently, the small percentage of unexplained variance between two conditions from our example represents not only individual differences in EFs but also error variances. Classical test theory claims that error variables are independent of each other (Lord et al., 1968; Novick, 1966). In consequence, difference scores of EF tasks include on the one hand the very small amount of EF-specific variance (Rouder & Haaf, 2019) and on the other hand the variances of the error variables of *both* conditions (see Schubert et al., 2022, for a more extensive treatment of this issue). This disproportionate influence of error variances results in low reliabilities of difference scores used to measure EFs as demonstrated in a number of recent studies (Hedge et al., 2018; Rouder et al., 2023; Rouder & Haaf, 2019).

In addition or as a consequence of the problem of unsatisfactory reliabilities, various scores measuring EF abilities showed insufficient convergent validity to other tasks used to measure the same EF (Frischkorn et al., 2019; Hedge et al., 2018; Hull et al., 2008; Karr et al., 2018; Klauer et al., 2010; Krumm et al., 2009; Rey-Mermet et al., 2018; Rey-Mermet, Gade, Souza, et al., 2019; Rouder et al., 2023; Rouder & Haaf, 2019; Stahl et al., 2014; von Bastian et al., 2020). For example, Rey-Mermet et al. (2018) and Rey-Mermet, Gade, Souza, et al. (2019) could not find a coherent pattern of correlations for both RT-based variables and accuracy-based indicator scores measured with a battery of 11 inhibition tasks. Further evidence of insufficient relations between measurements of inhibition has been provided in several recent studies (Krumm et al., 2009; Rouder et al., 2023; Rouder & Haaf, 2019; Stahl et al., 2014). A systematic review by Karr et al. (2018) reanalyzed nine datasets of EF data in adult populations and specified different types of factor models to find out which model describes the data best. Therefore, they specified and compared one-factor models, two-factor models, three-factor models, and bi-factor models. Their analysis revealed that none of these theoretically plausible models described the data clearly better than the others, and none of the models described the data sufficiently good (Karr et al., 2018). These findings suggest that performance measures derived from EF tasks show a lack of validity. Now, there is a chance that the reliability problem with difference scores is also responsible for the validity problem of EF tasks. This is because square root of the reliability of one variable provides the upper limit for possible correlations with any other variable. Alternatively, it is possible that both psychometric problems are unrelated to each other.

One possible solution to address the challenge of low reliabilities is the utilization of structural equation models (SEMs). SEMs allow for the separation of various sources of variance inherent in manifest measurement scores into one or more latent factors in addition to error variances. The latent factors represent the shared variance among the underlying manifest indicator variables and exhibit perfect reliabilities, since they are already separated from the statistically independent error variances present in the measured variables.

However, even if SEMs solve the problem of low reliability, the validity problem of EF tasks remained persistent. In our own recent research (Löffler et al., 2024) [1], we used a battery of nine EF tasks and modeled participants' individual performances with the drift parameters of the drift diffusion model (Ratcliff, 1978), allowing us to control for speed-accuracy tradeoffs. On the latent level, using SEM, we tried to separate individual differences in EF abilities from general, EF-unrelated processes reflected in the drift parameters. Our results shed alarming light on EF tasks, as no variance specific to EFs remained after controlling for general, EF-unrelated processes included in the drift parameters (Löffler et al., 2024). Despite using a state-of-the-art modeling approach and combining computational models of cognition with SEMs, our findings revealed that EF tasks do not appear to measure anything else than general, EF-unrelated processes on a behavioral level.

Measuring executive functions with electrophysiological process parameters

In the present study, we used electroencephalography (EEG) as an electrophysiological technique that offers several advantages compared to behavioral measurements and which is therefore a valuable alternative for the examination of EFs. The EEG is a non-invasive method that allows the measurement of neural brain activity at the scalp with high temporal resolution (Berger, 1929). Various metrics can be extracted from EEG data, either based on the frequency spectrum or in the time domain. Analyses in the time domain often employ event-related potentials (ERPs), which reflect the electrophysiological activity recorded on a specific scalp position after task-related events have occurred. For example, following the presentation of a stimulus event, electrophysiological activity can be measured across electrodes placed on the scalp and then averaged across many trials. The resulting ERPs are commonly divided into components with either positive or negative directions of the voltage values at specific recording locations and time windows. These ERPs are sensitive to task-

[1] Note that the study by Löffler et al. (2024) is based on the same dataset as the present study.

specific manipulations across different conditions (Gaillard, 1988). In addition, there is evidence that parameters from ERP components from specific scalp positions and time windows (e.g. the frontal N2 component) are associated with specific cognitive processes, such as the N2 is mainly associated with response inhibition (for a brief review see: Luck, 2014). This makes ERPs suitable markers for assessing individual differences in neurocognitive activity during EF tasks. In the present study, we therefore used these ERPs as dependent variables.

Typically, ERP components are quantified based on either their amplitude voltage values or on the latencies of their peaks. For our investigation, we opted to use two ERP parameters, the fronto-central N2 and the parieto-central P3 amplitude, to quantify individual differences in the electrophysiological activity during EF tasks. It is crucial to note that these ERP measures often exhibit only low-to-moderate reliabilities, which needs to be taken into account in individual differences research (Cassidy et al., 2012; Nebe et al., 2023; Schubert et al., 2023). Accordingly, we addressed the issue of unreliable ERP components by integrating analyses of the electrophysiological measures with a structural equation modeling approach.

The fronto-central N2 component

The fronto-central N2 component corresponds to the second negative peak of the ERP waveform that is recorded at fronto-central electrode sites following the presentation of a visual stimulus. This component usually peaks between 200 and 350 milliseconds after the visual stimulus was displayed. However, the specific timing of the peak varies depending on individual differences and task characteristics (see for an overview of the N2: Folstein & Van Petten, 2008). Several theoretical frameworks and decades of empirical research have interpreted the N2 as an electrophysiological parameter reflecting cognitive control processes and there is a large body of research showing how the amplitude of the N2 is modulated by cognitive control demands across different EF tasks (Folstein & Van Petten, 2008; Luck, 2014). For example, in typical inhibition tasks such as the Arrow-Flanker task (Eriksen & Eriksen, 1974), the N2 component has consistently exhibited a larger amplitude in the incongruent trials compared to congruent and neutral trials (e.g., Bartholow et al., 2005; Heil et al., 2000; Yeung et al., 2004). This enhanced N2 amplitude in conflict trials could be interpreted as indicative of increased activation of prefrontal cognitive control functions, given that the frontal N2 is predominantly generated in the anterior cingulate cortex (Nieuwenhuis et al., 2003; Yeung et al., 2004), a region in the brain known to play a crucial role in EF processes (Cameron et al., 1999). In addition, typical shifting tasks have been associated with observable effects on the N2 component, characterized by an increased amplitude in shifting trials compared to repeat trials (e.g., Gajewski et al., 2010). Similarly, the N2 component has been linked to updating-related processes in typical updating tasks. For instance, previous research has demonstrated that N2 amplitudes decrease with higher updating demands, as observed in an N-Back task (Gevins et al., 1996). Specifically, the 3-back condition exhibited smaller N2 amplitudes compared to the 1-back condition (Gevins et al., 1996; Salmi et al., 2019). Therefore, based on previous experimental findings, we selected the mean amplitude of the N2 component as an ERP parameter that may also capture individual differences in EF processes across various EF tasks.

The parieto-central P3 component

The parieto-central P3 component describes the third positive peak that occurs after a relevant stimulus is presented. This ERP component typically peaks between 300 and 650 milliseconds after the stimulus was presented (for an overview of the P3 see Polich, 2007). However, as with any ERP component, the specific timing of the peak varies depending on individual differences and task characteristics.

Donchin (1981) as well as Donchin and Coles (1988) interpreted the P3 component as a process parameter that reflects context updating, that is, the mental updating of broader environmental representations. Specifically, the P3 reflects changes in the environment that are relevant to the ongoing task. In the following years, other researchers provided different interpretations and suggested that the P3 component specifically reflects the updating of working memory. (Luck, 1998, 2014; Polich, 2007; Polich & Kok, 1995; Vogel et al., 1998; Vogel & Luck, 2002). Furthermore, Polich (2007) described the P3 as a process parameter representing cognitive demands and the allocation of attentional resources. According to Polich's (2007) perspective, the P3 amplitude can be considered as an electrophysiological process parameter linked to executive demands or cognitive control. To be more specific, inhibition, shifting, and updating tasks demand higher cognitive resources in the conditions with greater processing demands compared to those conditions with lower processing demands. The mean amplitude of the P3 component is thought to reflect these increased updating, attention allocation, and inhibition demands. Consequently, this ERP component emerges as a promising parameter to capture individual differences in EFs.

Empirically, several EF tasks have demonstrated effects on the mean amplitude of the P3 component between the condition with greater and the condition with lower processing demands. For instance, Pratt et al. (2011) observed increased P3 amplitudes in incongruent trials compared to neutral or congruent trials in a classical Flanker task. Additionally, in shifting tasks, numerous studies have reported larger P3 amplitudes in shifting conditions compared to repeat trials (e.g., Gajewski & Falkenstein, 2011). In updating tasks, particularly in the N-Back task, a consistent finding has been that P3 amplitudes decrease with increasing updating demands (Dong et al., 2015; Scharinger et al., 2015; Watter et al., 2001). Collectively, these findings indicate that heightened EF demands in various EF tasks elicit effects on the amplitudes of the P3 component. As a result, we consider the mean amplitude of the P3 as a possibly suitable ERP component representing individual differences in EF abilities across a broad range of EF tasks.

The present study

The primary goal of this study was to investigate the factor structure of EFs using electrophysiological measures rather than behavioral assessments. We decided to use ERP components as process parameters, because they may provide a more accurate representation of the underlying cognitive processes due to their higher temporal resolution. To accomplish this, we analyzed EEG data collected from 148 participants who completed a battery of nine EF tasks and three simple speeded binary choice tasks. Specifically, we focused on participants' mean amplitudes of two ERP components: The fronto-central N2 and the parieto-central P3, both known to be associated with EF processes. Our aim was to assess whether EFs, as measured by N2 and P3 amplitudes, exhibited the classical three-factor structure previously identified in studies using behavioral measures (e.g., Friedman et al., 2008; Friedman & Miyake, 2017; Miyake et al., 2000). Additionally, we examined the divergent validity of these factors to general, EF-unrelated properties of ERP components. Therefore, we isolated the variance specific to EF on the latent level by controlling for general, EF-unrelated processes measured with a battery of simple speeded decision tasks (Löffler et al., 2024). Lastly, we explored correlations between the EF-specific variance in the N2 and P3 components with intelligence and WMC. This exploration was motivated by accounts positing that differences in EF contribute to individual differences in higher-order cognitive abilities (e.g., Kane et al., 2008; Kovacs & Conway, 2016; Mashburn et al., 2023).

Methods

This study was approved by the ethics committee of the faculty of behavioral and cultural studies of Heidelberg University (reference number: Löf 2019/1-3). All procedures were conducted in accordance with the Declaration of Helsinki.

Participants

One hundred and fifty-one people from the general population participated in our study. We recruited our sample with advertisements in local newspapers, with flyers, and via a pool of people generally interested in participating in studies. After three persons withdrew from further study attendance, 148 people remained in the final sample (♀ 96, ♂ 51, one diverse). Participants' age ranged for 18 to 60 years (M_{age} = 31.52, SD_{age} = 13.91). All participants were fluent in German. Thirteen participants reported being left-handed. Seventy-eight participants wore glasses; therefore, all participants had normal or corrected-to-normal vision. Four participants had a middle-school degree, six had a qualification for university entrance for applied science, 81 participants had qualification for university entrance, 18 participants had a university degree in applied science, 37 participants had a university degree, and two participants had a PhD. This paper is part of a larger study that includes several projects. Consequently, the sample size was not specifically planned for this project. While our sample size is above average for a study with electrophysiological recordings, it is only moderate for SEM analyses. To mitigate any issues related to the moderate sample size, we utilized a Bayesian approach to estimate our structural equation models, aiming to benefit from shrinkage (see the "Structural Equation Modeling" section in this method part for detailed information). As compensation for participating in our study, participants received 75 € and personal feedback about their intelligence and WMC test results.

Materials

All tasks were programmed in MATLAB 2018b (The MathWorks Inc., Natick, Massachusetts) with the software package Psychtoolbox version 3.0.13 (Kleiner et al., 2007). In each of the following EF tasks and ECTs, the relevant stimuli were presented centered in the middle of the screen in front of a black background. If the task did not require the stimulus-presentation in a specific color, the stimuli were shown in gray. The RGB color codes, which we used for the shifting-, updating-, inhibition-, and ECTs can be found in Table S1 in the supplementary materials. In the EFs and ECTs, we instructed participants to respond as quickly and as accurately as possible. Table 1 shows the presentation times of the EF tasks and ECTs. At the beginning of each task, participants completed practice trials with feedback, followed by the experimental block without feedback.

Table 1 *Presentation times of the experimental tasks*

Task	Process	Fixation cross	ISI	MPT	CPT	PPT	ITI
OE-LM	shifting	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
GL	shifting	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
NL	shifting	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
FL	inhibition	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
NP	inhibition	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
Stroop	inhibition	400 - 600	400 - 600	---	---	1000 - 3000	1000 - 1500
NB	updating	---	400 - 600	---	---	1500	---
KT	updating	400 - 600	400 - 600	1000	800 - 1200	1000 - 3000	1000 - 1500
RS	updating	400 - 600	400 - 600	1000	800 - 1200	1000 - 3000	1000 - 1500
CRT	speed	1000 - 1500	---	---	---	1000 - 3000	1000 - 1500
Sternberg	speed	1000 - 1500	400 - 1000	1000	1800 - 2200	1000 - 3000	1000 - 1500
Posner	speed	1000 - 1500	---	---	---	1000 - 3000	1000 - 1500

Note. All values in the table represent milliseconds; ISI = Inter-stimulus interval; MPT = Memory item presentation time; CPT = Cue stimulus presentation time; PPT = Probe stimulus presentation time; ITI = Inter-trial interval; OE-LM = Odd/even-less/more task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task; CRT = Two Choice Reaction Time task; Posner = Posner task; Sternberg = Sternberg Memory task; Probe stimuli were presented until participants responded. If participants responded faster than 1000 milliseconds the stimulus remained until 1000 milliseconds were elapsed. The stimulus disappeared after 3000 milliseconds if the participants did not respond.

Shifting tasks

In the following three shifting tasks, the fixation-cross and the stimuli were presented in the same color.

Odd/Even-Less/More task. We adopted this task from Sudevan and Taylor (1987). In each trial, participants saw a number between one and nine (except five). Depending on the color of the stimuli (fixation cross and probe stimulus), participants had to decide whether the number was less than or greater than five (red; les/more response-set) or whether the number was an odd or an even number (green; odd/even response-set). Participants responded by pressing one of two keys on the keyboard. In 50% of the trials the stimulus appeared in the same color as one trial before (repeat condition) and in 50% of the trials the color changed between two successive trials (shifting condition). We pseudo-randomized the trials such that none of the response-sets and none of the conditions could appear more than three times in a row and that none of the digits appeared twice in a row. Participants first completed 10 practice trials for each response-set (repeat trials only) followed by 20 practice trials in which the response-sets could shift. After the practice block, participants completed 384 experimental trials.

Global/Local task. We adopted this task from Miyake et al. (2000). In this task, participants saw Navon figures (Navon, 1977) centered in the middle of the screen. A Navon figure describes a geometric figure with a large shape (global figure) composed of small geometric figures (local figures). The figure-set included the following four geometrical shapes: Circle, triangle, square, and cross. Within each trial, all the small figures occurred in the same shape, but this shape was different from the global figure. Depending on the color of the stimuli (fixation cross and probe stimulus), participants had either to identify the shape of the large figure (red; global response-set) or the shape of the small figures (green; local response-set). Participants responded by pressing one of four keys on the keyboard. In 50% of the trials the stimulus appeared in the same color as one trial before (repeat condition) and in 50% of the trials the color changed between two successive trials (shifting condition). We pseudo-randomized the trials such that none of the response-sets and none of the conditions could appear more than three times in a row as well as none of the shapes of the global figures appeared twice in a row. Participants first completed 10 practice trials for each response-set (repeat trials only) followed by 20 practice trials in which the response-sets could shift. After the practice block, participants completed 384 experimental trials.

Number/Letter task. We adopted this task from Rogers and Monsell (1995). Participants simultaneously saw a pair of stimuli consisting of a number and a letter in the middle of the screen. Depending on the color the stimuli (fixation cross and probe stimulus), participants either had to decide whether the number was less than or greater than five (red; number response-set) or whether the letter was a vowel or a consonant (green; letter response-set). Participants responded by pressing one of two keys on the keyboard. In 50% of the trials the stimulus appeared in the same color as one trial before (repeat condition) and in 50% of the trials the color changed between two successive trials (shifting condition). We pseudo-randomized the trials such that none of the response-sets and none of the conditions could appear more than three times in a row as well as none of the stimuli appeared twice in a row. Participants first completed 10 practice trials for each response-set (repeat trials only) followed by 20 practice trials in which the response-sets could shift. After the practice block, participants completed 256 experimental trials.

Inhibition tasks

Arrow-Flanker task. We used a standard Arrow-Flanker task (Eriksen & Eriksen, 1974) to measure inhibition. In each trial, participants saw a centrally presented arrow, which pointed either to the left or to the right side, flanked by four arrows, two on each side. Compared to the central arrow (target stimulus), the flanker arrows could either point to the same direction (congruent condition) or to the opposite direction (incongruent condition). Participants were instructed to indicate whether the target stimulus pointed to the left or to the right side, while ignoring the spatial orientation of the flanker arrows. Both conditions occurred equally often. Participants responded by pressing one of two keys on the keyboard. We pseudo-randomized the trials such that none of the conditions and none of the orientation of the target arrow appeared more than three times in a row. Participants completed 20 practice trials followed by 200 experimental trials.

Negative-Priming task. We used the Negative-Priming task developed by Tipper and Cranston (1985). In each trial, four horizontal lines appeared next to each other in the middle of the vertical axes of the screen. Subsequently, an X and an O appeared simultaneously on two of these four lines. By pressing one of four keys on the keyboard, participants had to indicate on which of the four lines the O appeared. This target stimulus occurred in half of the trials on the same position where the X appeared one trial before. Therefore, in this case, this position was negatively primed by the distractor. To redirect the attention to this negatively primed position, participants must allocate more attentional resources to overcome the remaining inhibition to this position. We pseudo-randomized the trials such that none of the conditions (negatively primed vs. not negatively primed) appeared more than three times in a row and none of the stimuli appeared more than three times in a row on the same position. Participants completed 20 practice trials followed by 192 experimental trials.

Stroop task. In this task, which was adapted from Stroop (1935), participants saw one of four color-words presented in one of four colors in the middle of the screen. The color-set consisted of the colors green, red, yellow, and blue. Participants had to indicate the color of the word while ignoring the meaning of the word by pressing one of four keys on the keyboard. The color and the meaning of the word could either be equal (congruent condition; 50% of the trials) or it could be different (incongruent condition; 50% of the trials). We pseudo-randomized the trials such that none of the conditions appeared more than three times in a row and none of the words or colors appeared twice in a row. Participants completed 20 practice trials followed by 192 experimental trials.

Updating tasks

N-Back task. This task was adopted from the verbal working memory conditions of the tasks by Gevins et al. (1996). It contained three blocks with different updating steps. The first block was a 0-back condition, which started with the presentation of a target letter followed by 96 trials. Participants had to decide whether the presented letter is the target letter or not. For each participant, the target and non-target letters were randomly drawn from a set of four letters. Before the first block, participants completed 20 practice trials. We did not include the data of the 0-back condition in our analyses. The second block was a 1-back condition. The stimulus-set in this condition contained four letters. In each trial, participants saw one letter and had to indicate whether this letter was the same as one trial before by pressing one of two keys on the keyboard. The third block was a 2-back condition. Again, the stimulus-set contained four letters. In each trial, participants saw one letter and had to indicate whether this letter was the same as two trials before by pressing one of two keys on the keyboard. The practice parts of the second and third block contained 30 trials each, followed by 96 experimental trials. Fifty percent of the trials were matching-trials, which means that the presented stimulus was equal to the target-letter (respectively the stimulus one or two trials before). We pseudo-randomized the trials such that none of the matching-conditions and none of the letters appeared more than three times in a row.

Keep-Track task. We adopted the Keep-Track task from Miyake et al. (2000). Participants completed two blocks with different memory demands (set size one and set size three). The stimulus-set was divided in four categories (geometrical-figures, colors, numbers, letters). Each of these categories contained six different stimuli. At the beginning of each trial in the first block, participants were given one of these categories as target category for the current trial. Then they saw a sequence of seven stimuli, which contained stimuli from each of the four categories. After this sequence, an additional probe stimulus from the target category followed, and participants had to indicate whether this probe was the last stimulus presented from that target category by pressing one of two keys on the keyboard. Half of the trials were match-trials, which means that the probe was identical to the last stimulus in the target category. Moreover, in half of the trials, the stimulus from the target category was updated, which means that two stimuli from the target category were shown within one sequence. We pseudo-randomized the trials such that none of the conditions (updating and matching) and none of the target categories appeared more than three times in a row. In the second block, participants were given three target categories instead of only one, again followed by the sequence of seven stimuli and one probe, equal to the procedure of the first block. In both blocks, participants completed 10 practice trials followed by 96 experimental trials.

Running-Span task. We adopted the Running-Span task from Broadway and Engle (2010). This task had two blocks with different memory-set sizes (three memory-stimuli in the first and five memory-stimuli in the second block). The updating steps ranged from zero to three. In the first block, participants saw a sequence of three letters followed by zero to three updating letters. After this sequence, a probe stimulus appeared, and participants had to decide whether this probe was part of the last three presented letters. In the second block, participants saw a sequence of five letters followed by zero to three updating letters. Again, after this sequence, a probe stimulus appeared, and participants had to decide whether this probe was part of the last five presented letters. They indicated their choice by pressing one of two keys on the keyboard. Within each block, half of the trials had zero updating steps and the other half of the trials contained one to three updating steps with equal frequency. Additionally, half of the trials were match-trials (i.e., the probe stimulus was part of the updated stimuli in the previous sequence). We pseudo-randomized the trials such that none of the updating steps, none of the matching conditions, and none of the probe stimuli appeared more than three times in a row. In each of the two blocks, participants completed 10 practice trials followed by 120 experimental trials.

Elementary cognitive tasks (ECTs)

Two Choice Reaction Time task. In the Two Choice Reaction Time task, participants saw a fixation-cross in the middle of the screen, which was surrounded by one quadratic frame on its left and one quadratic frame on its right side. In each trial, a plus sign appeared in one of these frames and participants had to indicate whether this plus appeared in the left or in the right frame by pressing one of two keys on the keyboard (Chen et al., 2012). The plus sign appeared equally often in both frames. We pseudo-randomized the trials such that none of the stimulus presentation sides appeared more than three times in a row. Participants completed 20 practice trials followed by 100 experimental trials.

Sternberg Memory task. In each trial of the Sternberg Memory task, developed by Sternberg (1969), participants saw a sequence of five distinct digits randomly drawn between zero and nine. This sequence represented the memory-set of the corresponding trial. After this sequence, a question mark appeared as a cue followed by a probe stimulus. Participants had to decide whether this probe was part of the previously presented memory-set by pressing one of two keys on the keyboard. In 50% of the trials the probe was part of the memory-set (matching condition). We pseudo-randomized the trials such that none of the conditions appeared more than three times in a row and none of the probe stimuli appeared twice in a row. Participants completed 20 practice trials followed by 100 experimental trials.

Posner task. In this task, developed by Posner and Mitchell (1967), participants had to decide whether a pair of two simultaneously presented letters are semantically identical. The stimulus set consisted of the following letters: A, B, F, H, Q, a, b, f, h, q. Participants responded by pressing one of two keys on the keyboard. For example, if participants saw an "AA" or an "Aa", they had to respond that the two letters had identical names. Instead, if participants saw an "AB" or an "Ab", they had to response that the two letters had different names. Each condition appeared in 50% of the trials. We pseudo-randomized the trials that none of the conditions appeared more than three times in a row. Participants completed 20 practice trials followed by 120 experimental trials.

Fluid Intelligence

We used the short version of the Berlin-Intelligence-Structure Test (BIS; Jäger et al., 1997) to measure participants' fluid intelligence. This instrument is useful for the assessment of a broad range of reasoning abilities in only one hour. The test allows to measure four operation-related (processing capacity [PC], processing speed [PS], memory [M], creativity [C]) and three context-related (verbal, numerical, figural) components of intelligence with a battery of 15 tasks in total. To measure participants intelligence, we aggregated their normalized *z*-score values of the four operation-related scales. In our sample, participants had a mean IQ of 95.86 (*SD* = 15.90).

Working memory capacity (WMC)

Participants' WMC was assessed with the Memory-Updating task, the Operation- and the Sentence-Span task, and the Spatial Short-Term Memory task from the working memory test battery by Lewandowsky et al. (2010). In addition, all except five participants completed the Letter-Location Binding task (Wilhelm et al., 2013). Due to an error in the program, we could not use the data of the Spatial Short-Term Memory task. As dependent variable, we used participants' proportion of correctly solved items for each task.

Procedure

Data collection took place within one year and participants came to our laboratory for three measurement occasions at three-month intervals. At the beginning of the first measurement occasion, participants signed an informed consent. We screened participants for color blindness using a set of Ishihara-plates (Ishihara, 2000). In session one and two, we prepared participants for EEG recording and seated them in a dimly lit cabin. They then completed several tasks in the following order. At the first occasion: Sternberg Memory task, Arrow Flanker task, Global Local task, N-Back task, Odd/Even-Less/More task, and Stroop task. Moreover, they also completed a questionnaire that asked for gender, age, education, and related information. At the second occasion: Running-Span task, Two Choice Reaction Time task, Number Letter task, Negative Priming task, Keep-Track task, and Posner task. In total, these two measurement occasions lasted approximately 3.5 hours each. At the third measurement occasion, participants completed the intelligence test, followed by the working memory test battery, and the letter binding task. Additionally, we administrated two further short tests measuring higher-order cognitive abilities, a mind-wandering questionnaire, and a pretzel task (these data are not reported in this paper). The third measurement occasion lasted approximately two hours. We administered all tasks in the same order for all participants to reduce between-subjects error variance (Goodhew & Edwards, 2019).

EEG Recording

We measured participants' EEG with 32 equidistant Ag-AgCl electrodes (see Figure S1 in the supplementary materials for detailed information of the electrode montage). We used the FpZ as ground electrode and the Cz as online reference. Offline, we re-referenced the data to the average of all electrodes. The impedances of all electrodes were kept under 5 kΩ. We recorded the EEG signal with a sampling rate of 1,000 Hz and a high-pass filter of 0.1 Hz. Due to recording errors, we had to discard the EEG data from two participants in session one and four participants in session two.

Data Analysis

For statistical analyses, we used the open-source software R – version 4.2.1 (R. Core Team, 2022) and the following packages: We preprocessed the data with "tidyverse" (Wickham et al., 2019) and calculated descriptive statistics with "psych" (Revelle, 2020) and the correlations with "Hmisc" (Harrell, 2019). Finally, we estimated the SEM parameters with a Bayesian approach using the package "blavaan" (Merkle & Rosseel, 2018).

EEG data preprocessing and intraindividual outlier analysis

Before preprocessing the EEG data, we conducted intra-individual outlier analyses for all tasks and removed any trials with RTs faster than 150 ms or with logarithmized RTs deviating more than 3 *SD*s from the intra-individual mean. On average, 0.69% (range: from 0.33% to 1.06%) of the trials were detected as intra-individual outliers.

EEG data were preprocessed using the open-source toolbox EEGLAB (version 2022.1; Delorme & Makeig, 2004) in MATLAB 2022a (The Math Works Inc., Natick, Massachusetts). The steps and criterion in the preprocessing pipelines were identical for each of the EF tasks and ECTs. First, we created a separate dataset for the independent-component analyses (ICA) and down-sampled this to 200 Hz. Afterwards, we filtered the continuous EEG and ICA datasets within each task with a second order infinite impulse response Butterworth band-pass filter (EEG dataset: 0.1-30 Hz; ICA dataset: 1-30 Hz). Based on the probability (threshold 5 *SD*) and kurtosis (threshold 10 *SD*) of the channel data, we subsequentially detected and discarded bad channels. Later, we interpolated these channels. After the bad-channel detection, we re-referenced the data to the average reference. Then, we segmented the continuous EEG data into segments 1200 ms long, starting 200 ms before stimulus onset. For the detection of artifact-contaminated segments, we used an iterative automatic procedure with following thresholds: 1000 μV to detect large fluctuations, 5 *SD* of probability for improbable data detection, and 5% maximum of the number of segments that could be rejected in each iterative step. Afterwards, we conducted an independent component analysis (ICA), using the infomax algorithm (Bell & Sejnowski, 1995). The resulting decomposition matrix was added to the EEG dataset to identify independent components containing artifacts with the ICLabel algorithm (Pion-Tonachini et al., 2019). On average, 14.07 (SD = 4.37) ICs were excluded. In the last preprocessing step, we conducted a further automated identification of artifact-containing segments with the same criterions as before.

Event-related potentials

We quantified and analyzed the ERPs with ERPLAB (version 9; Lopez-Calderon & Luck, 2014), an open source toolbox for MATLAB. We calculated the ERPs time-locked on different types of stimuli for each EF separately. Shifting ERPs were time-locked to the colored fixation cross at the beginning of each trial. Updating ERPs were time-locked to the stimuli, which had to be updated in the sequence of the memory-set in the Keep-Track and in the Running-Span task, and to each stimulus in the N-Back task. The ERPs for the inhibition- and ECTs were time-locked to the probe stimuli. We used the preceding 200 ms before stimulus onset as baseline interval. ERP waveforms were calculated in the 1000 ms after stimulus onset. By visual inspection of the grand averages, we specified separate time windows for each of the different ERP components in each task (see Table 2 for the specific time windows). Within these time windows, we extracted participants' mean amplitudes for the corresponding ERP component. For each of the different ERP components, we used the same electrode positions in all tasks. We measured the N2 at a fronto-central electrode, and the P3 at a parieto-central electrode (see Figure S1 in the supplementary materials).

Table 2 *Time windows for mean amplitudes of the ERP components*

Task	Process	ERP time-locked to:	N2	P3
OE-LM	shifting	fixation cross	320-380	330-580
GL	shifting	fixation cross	330-385	310-480
NL	shifting	fixation cross	320-380	340-590
FL	inhibition	probe	280-360	320-500
NP	inhibition	probe	250-350	320-540
Stroop	inhibition	probe	300-370	310-640
NB	updating	probe/ updating stimulus	250-350	310-460
KT	updating	updating stimulus	290-375	330-660
RS	updating	updating stimulus	260-310	335-395
CRT	speed	probe	300-370	270-480
Posner	speed	probe	290-410	340-670
Sternberg	speed	probe	300-390	350-670

Note. All values in the table represent milliseconds; OE-LM = Odd/Even-Less/More task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task; CRT = Two Choice Reaction Time task; Posner = Posner task; Sternberg = Sternberg Memory task.

Structural equation modeling

Bayesian modeling approach: Before we estimated the parameters of the SEMs, we conducted inter-individual outlier analyses. Separately for each task, we excluded participants' data if they showed accuracies below 70%[2]. In this step of the inter-individual outlier detection, we removed on average 2.20% of participants from each task (range: from 0% to 6.9%). Furthermore, we removed data if the mean amplitudes of the ERP components deviated more than 3 *SD* from the mean (average of removed data: 0.91 %; range: from 0% to 2.11%). Afterwards, we *z*-standardized all variables for subsequent analyses.

We estimated the SEMs using Bayesian estimation procedures, because Bayesian SEMs provide more adequate parameter estimates, particularly in smaller sample sizes, compared to conventional frequentist estimation methods (McNeish, 2016). In each model, the parameters were sampled with three Markov chain Monte Carlo (MCMC) chains. Each chain comprised 1000 warm-up samples and 3000 samples after the warm-ups. Model convergence was evaluated based on the Gelman-Rubin convergence statistic $\hat{R}$, which compares the estimated between-chains and within-chain variances for each model parameter (Gelman & Rubin, 1992). Negligible differences between these variances were indicated by $\hat{R}$ values close to 1. Goodness-of-fit was assessed using the Bayesian versions of the comparative fit index (BCFI) and the root mean square error of approximation (BRMSEA). Analogous to the interpretation of the comparative fit indices for frequentist-estimated models, BCFI values > .95 and BRMSEA ≤ .06, indicated a good model fit and BCFI values > .90 and BRMSEA ≤ .08 indicated an acceptable model fit (Garnier-Villarreal & Jorgensen, 2020; Hoofs et al., 2018). Because we *z*-standardized all variables, we fixed the intercepts of

[2] Although guessing probability is at 50% for all tasks, except for the Stroop task, the Global/Local task, and the Negative Priming task with a guessing probability of 25%, these tasks are generally very easy with subjects making very little mistakes (Mean accuracy > 90%). Therefore, we decided to have a more conservative threshold for exclusion at 70% instead of a threshold based on the actual guessing probability.

all manifest variables to zero, both in our candidate models as well as in the corresponding baseline models specified for calculating the BCFI.

For the Bayesian SEMs, we used the following priors: For regression and factor loading parameters, we used normal priors with $\mu = 0$ and $\sigma = 10$, for correlation parameters beta priors with $\alpha = 1$ and $\beta = 1$, and for variance parameters gamma priors with a shape of 1 and a rate of scale of 0.5. Our choice of the beta and gamma priors assures that correlations could not exceed the range between -1 and 1, and variances could not take on negative values.

Model compositions: To examine the factor structure of EFs measured with ERPs and simultaneously account for the mean amplitudes of the N2 and the P3, we employed joint models across ERP components and tasks. Initially, we attempted a classical six-factor model (three EF factors for each ERP component) with manifest difference scores to assess whether these scores could effectively capture individual differences in EFs with ERPs. Afterwards, we avoided manifest difference scores and instead used the ERPs derived from the conditions with greater processing demands. Again, we designed a six-factor joint model, incorporating one N2 and one P3 factor for each of the three EFs. In accordance with the three-factor model of EF (Miyake et al., 2000), we specified the model with correlations between these latent factors.

Subsequently, to assess the common variance across shifting, updating, and inhibition, we formulated a hierarchical factor model introducing higher-order factors across the three EFs for the N2 and the P3, respectively. Furthermore, to isolate the EF-specific parts within these higher-order factors, we controlled these higher-order factors for the general, EF-unrelated properties of the ERP components, which were measured with a battery of simple speeded decision tasks (ECTs).

In the final step, we examined the specificity of the observed EF-specific variance. To achieve this, we employed the same hierarchical model, including indicators from both conditions – those with greater processing demands and those with lower processing demands. All resulting latent factors were controlled for general, EF-unrelated properties of the ERP components. Furthermore, we introduced higher-order factors over the remaining credible residuals of the latent first-level factors on both sides (greater and lower processing demands). For instance, a higher-order factor was introduced that loaded onto the residuals of the latent incongruent P3 factor and the latent congruent P3 factor in inhibition tasks not accounted for by the general P3 factor. Moreover, we assessed the relationships of the latent factors with higher-order cognitive abilities, more precisely to WMC and intelligence.

Openness and transparency

We provide access to the preprocessed data and code underlying this paper via the Open Science Framework repository at the following link: https://osf.io/a6zr9/ (Löffler et al., 2024). Furthermore, we provide access to the raw data and materials through the Open Science Framework repository: https://osf.io/4pvz3/ (Löffler & Schubert, 2024). Please note that the materials for the BIS are commercially licensed and are, therefore, not included. Neither the study nor the analyses were preregistered. We declare no conflicts of interest.

Results

Table 3 shows the descriptive statistics for all relevant variables in our study. Notably, the ERP measures exhibited excellent reliability. Table 4 displays the manifest correlations between the mean amplitudes of the ERP components.

Table 3 *Descriptive statistics of mean amplitudes of ERP components in each task*

Task	Dependent variable/ ERP component	Mean	*SD*	Reliability
Odd/Even-Less/More	N2	-1.25	1.71	.87[a]
Global/Local	N2	-1.4	1.58	.90[a]
Number/Letter	N2	-1.26	1.53	.86[a]
Odd/Even-Less/More	P3	2.64	1.69	.92[a]
Global/Local	P3	2.73	1.81	.91[a]
Number/Letter	P3	2.59	1.76	.89[a]
Arrow-Flanker	N2	-0.66	2.38	.94[a]
Negative-Priming	N2	-1.78	2.3	.94[a]
Stroop	N2	-2.65	2.75	.93[a]
Arrow-Flanker	P3	3.95	2.66	.96[a]
Negative-Priming	P3	4.01	2.61	.94[a]
Stroop	P3	3.46	2.57	.94[a]
N-Back	N2	0.58	1.76	.85[a]
Keep-Track	N2	-0.37	1.44	.84[a]
Runing-Span	N2	-0.1	1.41	.92[a]
N-Back	P3	3.14	2.07	.89[a]
Keep-Track	P3	2.93	1.64	.86[a]
Runing-Span	P3	2.53	1.33	.89[a]
Posner	N2	-1.59	2.16	.94
Sternberg	N2	0.18	2.17	.93
Choice RT	N2	-0.22	2.49	.95
Posner	P3	5.89	2.52	.96[a]
Sternberg	P3	2.76	2.38	.96[a]
Choice RT	P3	4.98	2.15	.94[a]
MU	percent correct	63	20	.88[b]
Binding	percent correct	86	11	.82[b]
Operation span	percent correct	78	13	.89[b]
Sentence span	percent correct	84	11	.87[b]
BIS-PC	scales-scores	101.61	7.12	.75[b]
BIS-PS	scales-scores	101.14	7.15	.49[b]
BIS-M	scales-scores	98.59	7.16	.58[b]
BIS-C	scales-scores	98.15	6.97	.45[b]

Note: The mean amplitudes of the N2 and P3 components are given in µV; MU = memory updating; BIS-PC = processing capacity scale of the Berlin-Intelligence-Structure Test; BIS-PS = processing speed scale of the Berlin-Intelligence-Structure Test; BIS-M = memory scale of the Berlin-Intelligence-Structure Test; BIS-C = creativity scale of the Berlin-Intelligence-Structure Test; [a] = Reliability scores were estimated with Spearman-Brown corrected correlations; [b] = Reliability scores were estimated with Cronbach's α.

Table 4 *Manifest correlations between the mean amplitudes of the ERP components in each task*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	Odd/Even-Less/More N2																							
2	Global/Local N2	**.66**																						
3	Number/Letter N2	**.67**	**.67**																					
4	Odd/Even-Less/More P3	**-.30**	**-.34**	**-.20**																				
5	Global/Local P3	**-.22**	**-.43**	**-.22**	**.68**																			
6	Number/Letter P3	-.16	**-.25**	-.16	**.72**	**.57**																		
7	Arrow-Flanker N2	**.30**	**.42**	**.40**	-.09	-.12	-.06																	
8	Negative-Priming N2	**.40**	**.35**	**.43**	-.09	-.02	-.03	**.66**																
9	Stroop N2	**.45**	**.45**	**.48**	**-.19**	**-.25**	-.07	**.46**	**.51**															
10	Arrow-Flanker P3	**-.26**	**-.35**	**-.37**	**.26**	**.28**	**.22**	**-.36**	**-.26**	**-.34**														
11	Negative-Priming P3	-.11	**-.22**	-.17	**.25**	**.23**	**.32**	**-.25**	**-.26**	**-.19**	**.64**													
12	Stroop P3	**-.38**	**-.32**	**-.38**	**.30**	**.23**	.16	**-.22**	**-.25**	**-.52**	**.57**	**.48**												
13	N-Back N2	**.26**	**.29**	**.32**	.11	.04	.03	**.48**	**.39**	**.34**	**-.22**	-.07	-.14											
14	Keep-Track N2	**.32**	**.38**	**.39**	-.06	-.10	-.05	**.37**	**.50**	**.41**	**-.29**	**-.23**	**-.27**	**.41**										
15	Runing-Span N2	**.34**	**.34**	**.39**	-.08	-.10	-.05	**.37**	**.46**	**.37**	**-.29**	-.16	**-.24**	**.47**	**.56**									
16	N-Back P3	**-.27**	**-.35**	**-.34**	**.32**	**.40**	**.28**	**-.21**	-.17	**-.33**	**.61**	**.43**	**.56**	-.15	**-.32**	**-.24**								
17	Keep-Track P3	**-.24**	**-.30**	**-.25**	**.43**	**.44**	**.55**	-.15	-.08	**-.21**	**.46**	**.47**	**.38**	-.06	**-.20**	-.16	**.55**							
18	Runing-Span P3	-.12	**-.27**	**-.23**	**.28**	**.37**	**.35**	**-.19**	-.14	**-.22**	**.42**	**.41**	**.35**	-.02	-.13	**-.18**	**.48**	**.53**						
19	Posner N2	**.42**	**.39**	**.48**	**-.33**	**-.30**	**-.24**	**.48**	**.55**	**.61**	**-.26**	**-.32**	**-.35**	**.37**	**.51**	**.44**	**-.30**	**-.33**	**-.23**					
20	Sternberg N2	.06	**.17**	**.23**	.00	-.04	.08	**.49**	**.37**	**.27**	**-.20**	-.13	**-.22**	**.36**	**.22**	**.26**	-.07	-.12	-.06	**.41**				
21	Choice RT N2	.15	**.24**	**.29**	-.12	**-.35**	-.10	**.38**	**.34**	**.29**	**-.23**	-.15	-.07	.16	.11	**.18**	**-.24**	-.16	**-.23**	**.42**	**.22**			
22	Posner P3	**-.32**	**-.32**	**-.38**	**.39**	**.35**	**.49**	**-.24**	**-.26**	**-.31**	**.52**	**.48**	**.51**	-.14	**-.25**	-.03	**.56**	**.59**	**.41**	**-.40**	-.17	-.15		
23	Sternberg P3	**-.25**	**-.32**	**-.29**	**.25**	**.29**	**.18**	**-.24**	**-.19**	**-.36**	**.51**	**.40**	**.45**	-.11	-.11	-.15	**.59**	**.49**	**.25**	**-.26**	**-.22**	-.14	**.47**	
24	Choice RT P3	**-.23**	**-.29**	**-.33**	**.25**	**.28**	**.31**	**-.26**	**-.23**	**-.27**	**.56**	**.58**	**.39**	-.07	-.11	**-.20**	**.36**	**.40**	**.34**	**-.28**	-.15	**-.22**	**.50**	**.45**

Note: Bold printed values represent significant correlation coefficients ($p < .05$).

Experimental effects

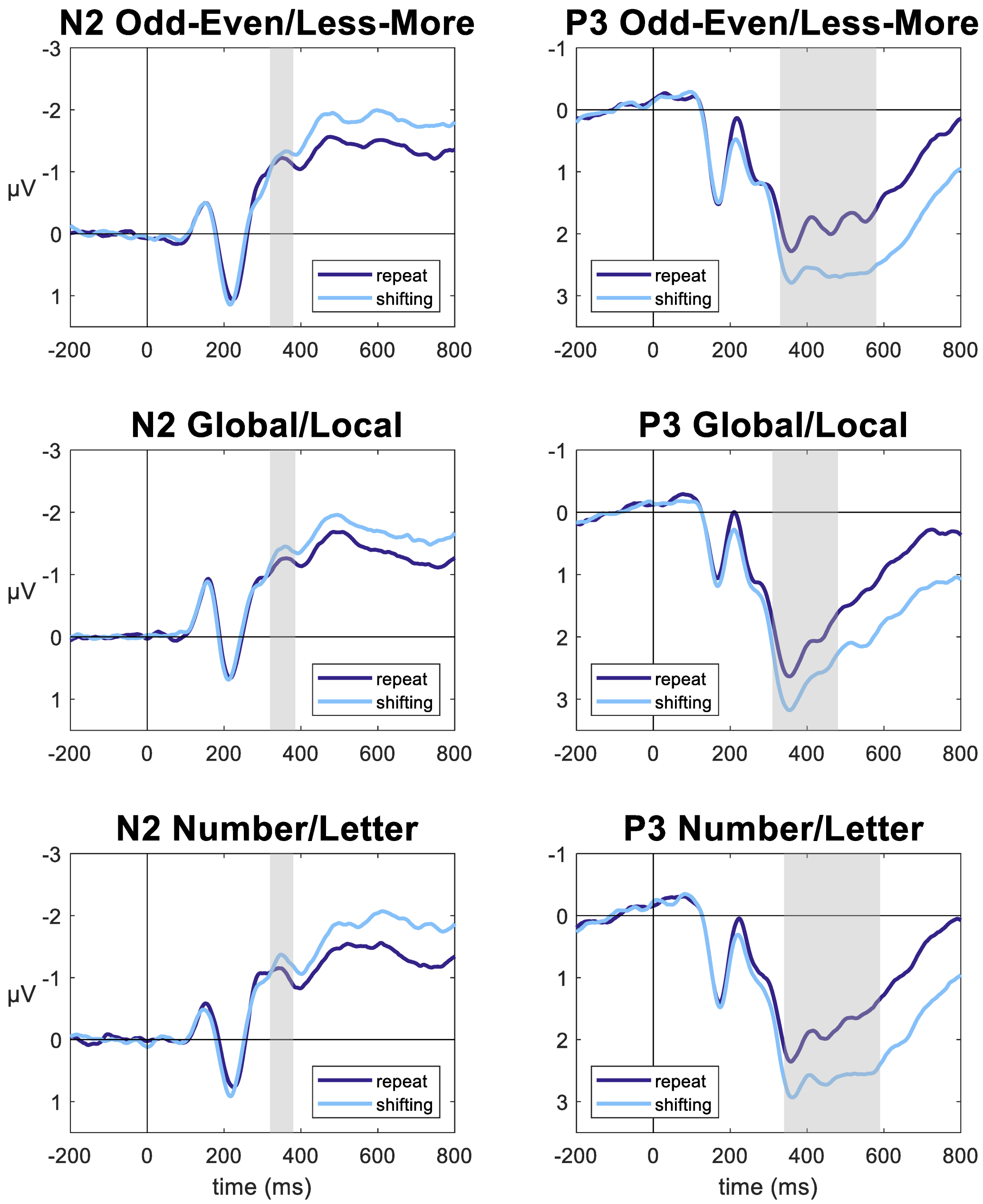
Prior to investigating the experimental effects on the ERP measures, we conducted an initial analysis to examine differences between task conditions at the behavioral level. These analyses utilized scores employed by Miyake et al. (2000), including condition-dependent mean reaction time scores for the shifting and inhibition tasks and arcsine-transformed proportion correct scores for the updating tasks. On the behavioral level, effect sizes revealed medium to strong effects, with Cohen’s d ranging from 0.72 for the Running-Span task to 1.85 for the Stroop task (see Table S2 in the supplementary materials for detailed information on the experimental effects at the behavioral level). These findings suggest significant variations between task conditions. More details about the behavioral effects and findings can be found in a previous publication of the data (Löffler et al., 2024). Based on these results, it is evident that the tasks are suitable for further investigation at the electrophysiological level. For the sake of completeness, we would like to note that Figure S2 in the supplementary materials shows the ERP waves from the fronto-central N2 and the parieto-central P3 of the ECTs.

Shifting tasks

N2. We found a mixed pattern of effects for the differences between the repeat and shifting conditions in each of the three tasks when examining the mean amplitudes of the N2 component (see Figure 1). In the Odd/Even-Less/More task, we did not observe any effect between the two conditions, with $t(139) = 1.33$, $p = .185$, Cohen’s $d = 0.11$, 95% CI [-0.05, 0.28]. However, in the Global/Local task and in the Number/Letter task, participants showed larger N2 amplitudes in the shifting than in the repeat conditions, with $t(143) = 2.78$, $p = .009$, Cohen’s $d = 0.23$, 95% CI [0.07, 0.40], and $t(132) = 2.18$, $p = .031$, Cohen’s $d = 0.19$, 95% CI [0.02, 0.36], respectively.

P3. We observed moderate to strong effects in the differences between the repeat and shifting conditions in each of the three tasks for the mean amplitudes of the P3 component (see Figure 1). Participants showed larger P3 amplitudes in shifting compared to repeat trials in the Odd/Even-Less/More task $t(138) = -10.91$, $p < .001$, Cohen's $d = -0.93$, 95% CI [-1.13, -0.73], in the Global/Local task, $t(142) = -7.78$, $p <$ 001, Cohen's $d = -0.65$, 95% CI [-0.83, -0.47], and in the Number/Letter task $t(130) = -8.59$, $p < .001$, Cohen's $d = -0.75$, 95% CI [-0.95, -0.56].

Figure 1 *The ERP wave forms of the fronto-central N2 and parieto-central P3 in the shifting tasks*



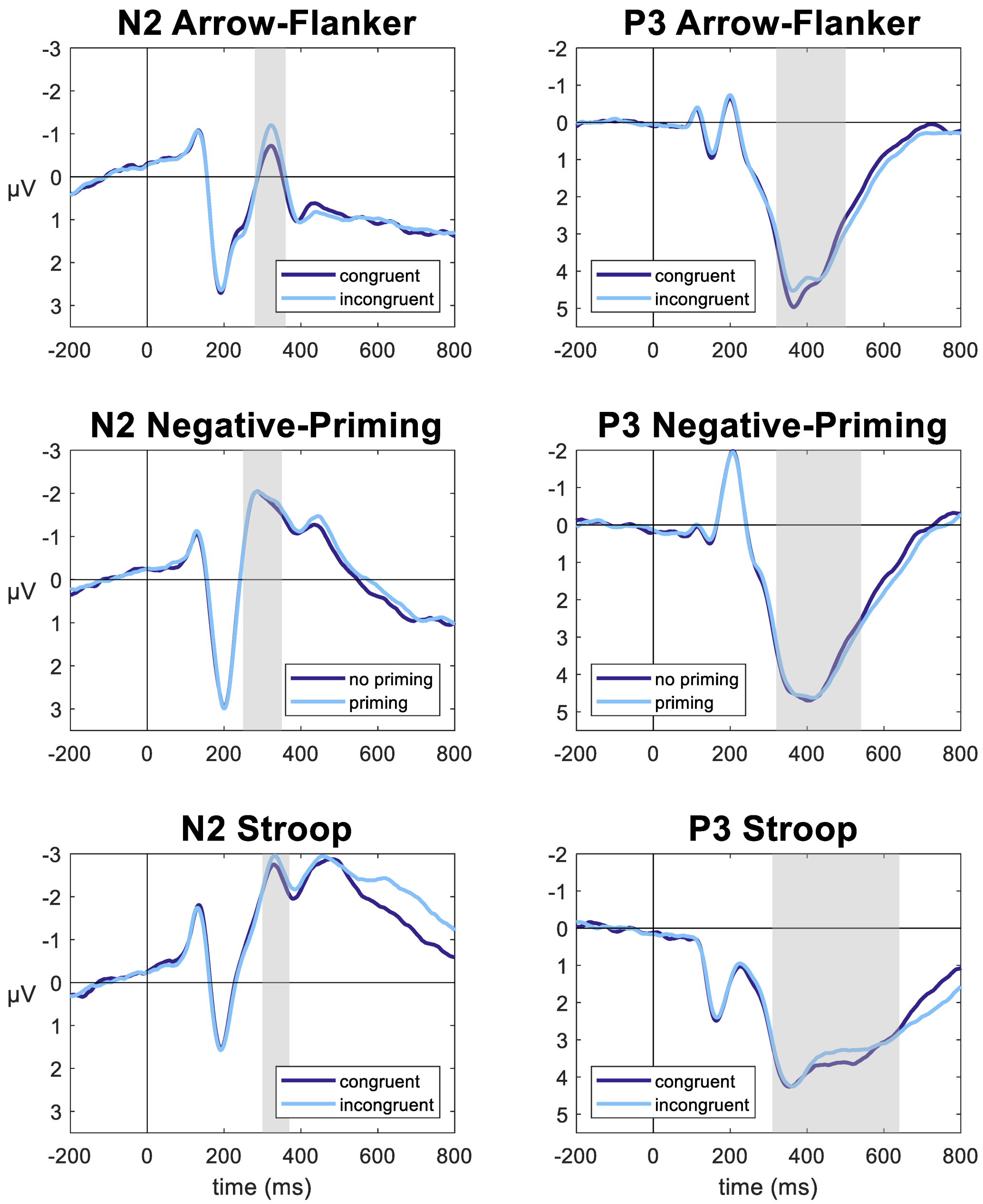
Note. The ERPs are locked to the colored cues (fixation cross) in the corresponding tasks. The gray shaded areas indicate the time intervals in which the mean amplitudes were measured.

Inhibition tasks

N2. Regarding the mean amplitudes of the N2 component in inhibition tasks, we observed an inconsistent pattern of effects between the congruent (no-priming) and incongruent (priming) conditions across all three tasks (see Figure 2). Participants showed a larger N2 mean amplitude in the incongruent compared to the congruent condition in the Arrow-Flanker task, $t(142) = 4.34$, $p < .001$, Cohen's $d = 0.36$, 95% CI [0.19, 0.53], as well as in the Stroop task, $t(142) = 2.21$, $p = .028$, Cohen's $d = 0.18$, 95% CI [0.02, 0.35]. However, we did not find any significant differences in the N2 mean amplitudes between the two conditions in the Negative-Priming task, with $t(128) = 0.67$, $p = .505$, Cohen's $d = 0.06$, 95% CI [-0.11, 0.23].

P3. We did not observe any effects on the mean amplitudes of the P3 component in the inhibition tasks in any of the three tasks (see Figure 2). In detail, we did not find any differences in mean P3 amplitudes in the Arrow-Flanker task, $t(142) = 1.38$, $p = .170$, Cohen's $d = 0.12$, 95% CI [-0.05, 0.28], in the Negative-Priming task, $t(128) = -0.67$, $p = .506$, Cohen's $d = -0.06$, 95% CI [-0.23, 0.11], or in the Stroop task, $t(144) = 1.91$, $p = .058$, Cohen's $d = 0.16$, 95% CI [-0.01, 0.23].

Figure 2 *The ERP wave forms of the fronto-central N2 and parieto-central P3 in the inhibition tasks*



Note. The ERPs are locked to the probe stimuli in the corresponding tasks. The gray shaded areas indicate the time intervals in which the mean amplitudes were measured.

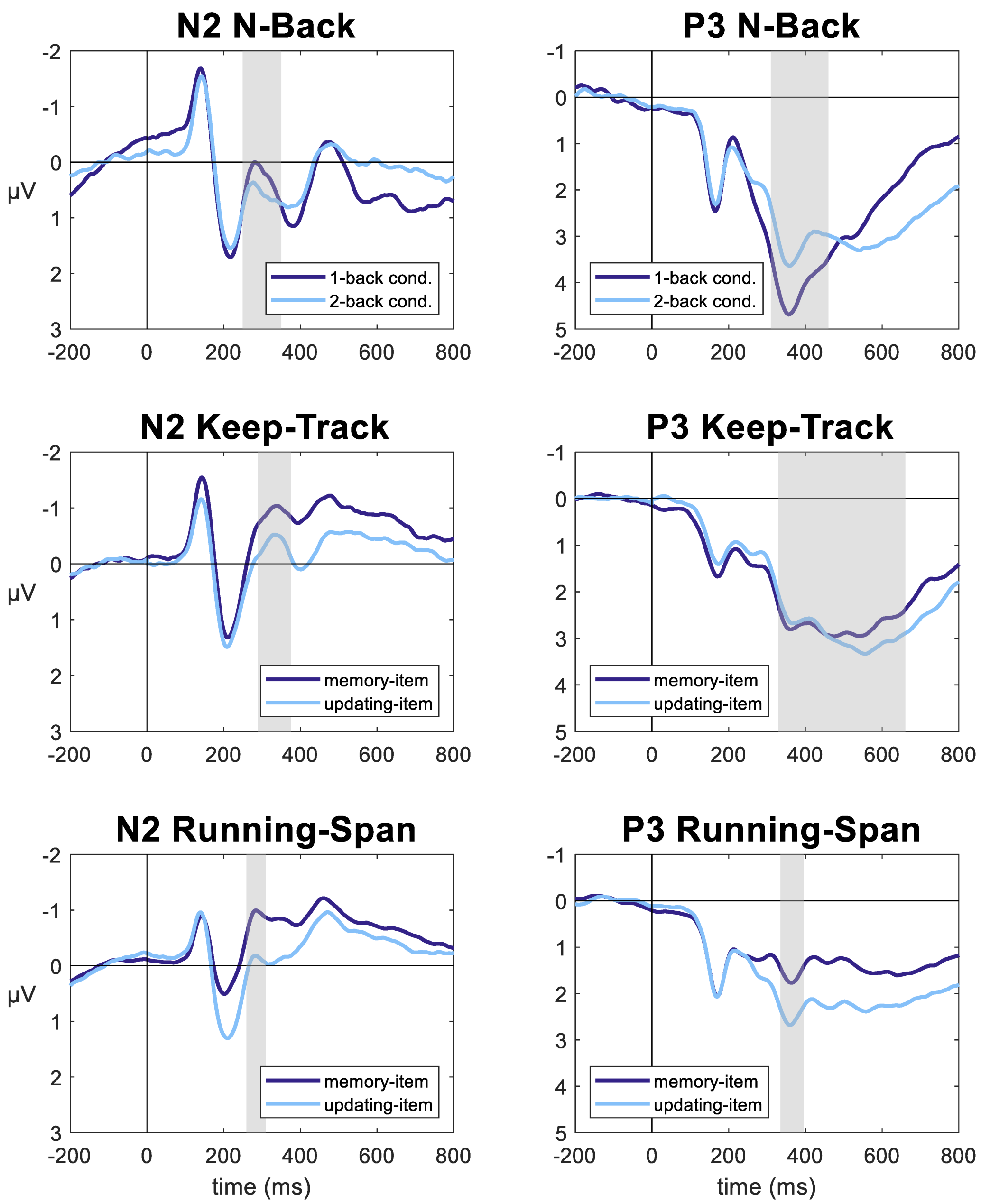
Updating tasks

N2. We found small to moderate effects for the differences between the relevant conditions in the updating tasks (see Figure 3). Participants showed smaller N2 amplitudes in the 2-back condition compared to the 1-back condition in the N-Back task, $t(133) = -2.31$, $p = .022$, Cohen's $d = -0.20$, 95% CI [-0.37, -0.03]. Also, in the Keep-Track task and the Running-Span task, they showed smaller N2 amplitudes when memory item was updated in participants' memory compared to items on positions in the memory-set, which were encoded for the first time, $t(123) = -6.72$, 95%, $p < .001$, Cohen's $d = -0.60$, 95% CI [-0.80, -0.41] and, $t(132) = -9.95$, $p < .001$, Cohen's $d = -0.86$, 95% CI [-1.07, -0.67], respectively. Taken together, the condition with greater processing demands showed less negative amplitudes in the N2 component compared to the condition with lower processing demands.

P3. We observed a mixed pattern of effects for the mean amplitudes of the P3 component in the updating tasks (see Figure 3). Firstly, in the N-Back task, we observed a significant difference in P3 amplitude between the 1-back and 2-back conditions. Specifically, the P3 amplitude was smaller in the 2-back condition compared to the 1-back condition, $t(132) = 7.52$, $p < .001$, Cohen's $d = 0.65$, 95% CI [0.47, 0.84]. In contrast, the Running-Span task showed an opposite trend. Here, participants exhibited a larger P3 amplitude when an item was updated in the memory set, as opposed to when an item was encoded for the first time, $t(131) = -8.79$, $p < .001$, Cohen's $d = -0.76$, 95% CI [-0.96, -0.57]. However, a detailed examination of the grand average ERP data, particularly in the parieto-central region, suggests that this P3 effect might have originated from a difference already present in the N2 time window. Finally, in the Keep-Track task, no significant differences were found in the P3 amplitudes between updating and memory items, $t(122) = -1.68$, 95% CI , $p =.096$, Cohen's $d = -0.15$, 95% CI [-0.33, 0.03]. Overall, these findings indicated that the effects on P3 amplitudes vary depending on the specific updating task involved. This inconsistency underscores the complexity of the cognitive processes involved in these tasks.

In summary, our analysis revealed a wide range of effects in the EF tasks. While shifting tasks revealed moderate to strong effects on the P3 amplitude, differences in the N2 component were either absent or minimal. Inhibition tasks exhibited small effects on the N2 amplitudes for two out of three tasks but no effects on P3 amplitudes. Lastly, updating tasks demonstrated small to strong effects on N2 amplitudes in each of the three tasks and largely inconsistent effects on P3 amplitudes.

Figure 3 *The ERP wave forms of the fronto-central N2 and parieto-central P3 in the updating tasks*



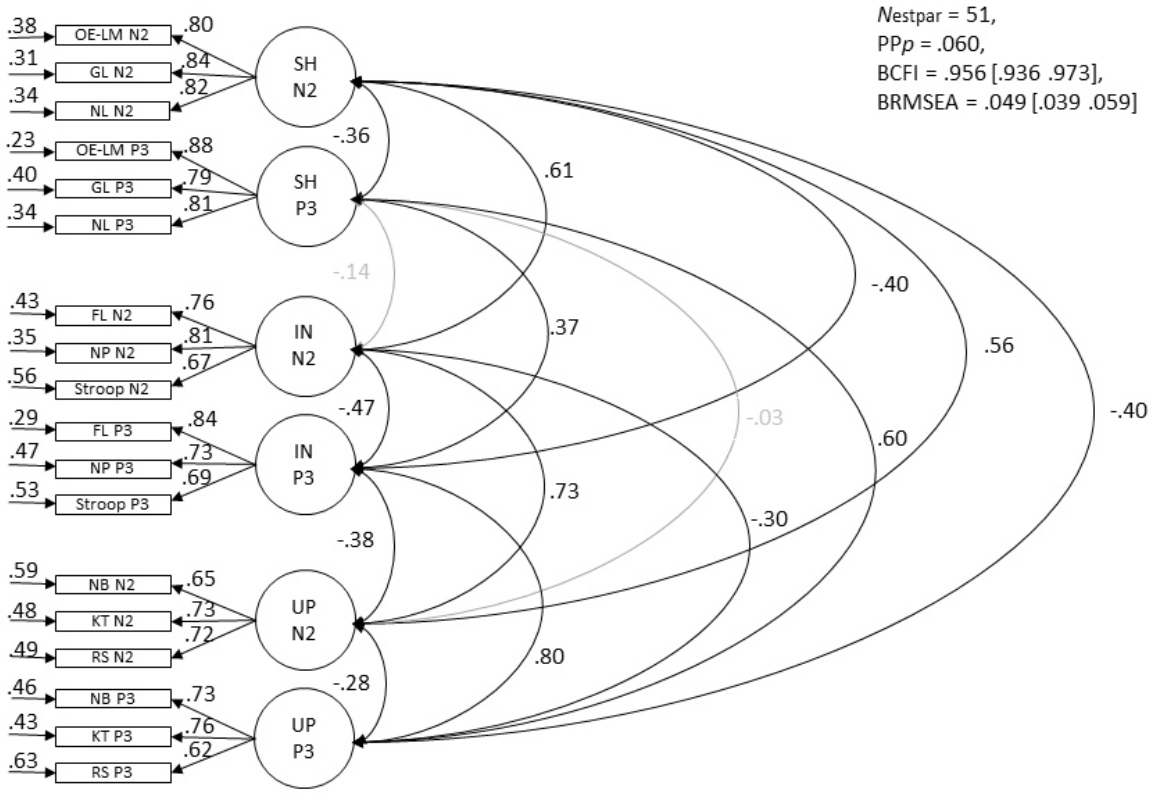
Note. The ERPs are locked to the probe stimuli in the N-Back task and to the stimuli in the sequences of the memory-sets in the Keep-Track and Running-Span task. The gray shaded areas indicate the time intervals in which the mean amplitudes were measured.

The factor structure of EFs

To assess the three-factor structure of EFs with mean amplitudes of the N2 and P3 components simultaneously, we used a joint modeling approach. Our first model incorporated six factors – three for each ERP component – and used mean amplitude difference scores as indicators. This model demonstrated a mediocre fit to the data, PP*p* (51) = .330, BCFI = .693, 95% CI [.380, 1.00], BRMSEA = .030, 95% CI [.007, .050]. The relatively low BCFI value implies that the baseline model, which treats all manifest variables as uncorrelated, was already adequate in explaining the data. This inference is further supported by the BRMSEA of .053 (95% CI [.048, .057] of the baseline model, which suggests that the N2 and P3 difference scores across tasks may not be credibly related. Additionally, the latent factors in our model mostly did not show credible variances, with the exceptions of shifting P3, σ^2 = .050, 95% PI [.001, .283], and updating P3, σ^2 = .063, 95% PI [.001, .246]. This result indicates that our model might not be the best representation of the data structure. Moreover, the reliability of the difference scores was largely unsatisfactory. A substantial majority, 15 out of 18 difference scores, exhibited Spearman-Brown corrected odd-even correlations below .50, as detailed in Table S3 in the supplementary materials. These findings suggest that the manifest difference scores of the mean N2 and P3 amplitudes are not adequate for measuring individual differences in EFs.

We therefore abandoned the difference scores in the following models and rather used the mean N2 and P3 amplitudes of the conditions with greater processing demands as manifest variables. To account for general, EF-unrelated processes, some of these models included control variables that are based on simple decision tasks (ECTs). In a first step, we analyzed the correlation structure of the mean N2 and P3 amplitudes in conditions with greater processing demands. This model is illustrated in Figure 4. It demonstrated a good fit to the data PP*p* (51) = .060, BCFI = .956, 95% CI [.936, .973], BRMSEA = .049, 95% CI [.039, .059]. Crucially, this model identified distinct factors for the N2 and P3 components within each of the three EFs. These latent factors showed moderate to strong correlations (see Figure 4). However, an exception was noted: the shifting P3 factor did not show a credible relationship with the inhibition N2 and updating N2 amplitudes.

Figure 4 *Three factors of executive functions with N2 and P3 amplitudes and correlated factors*

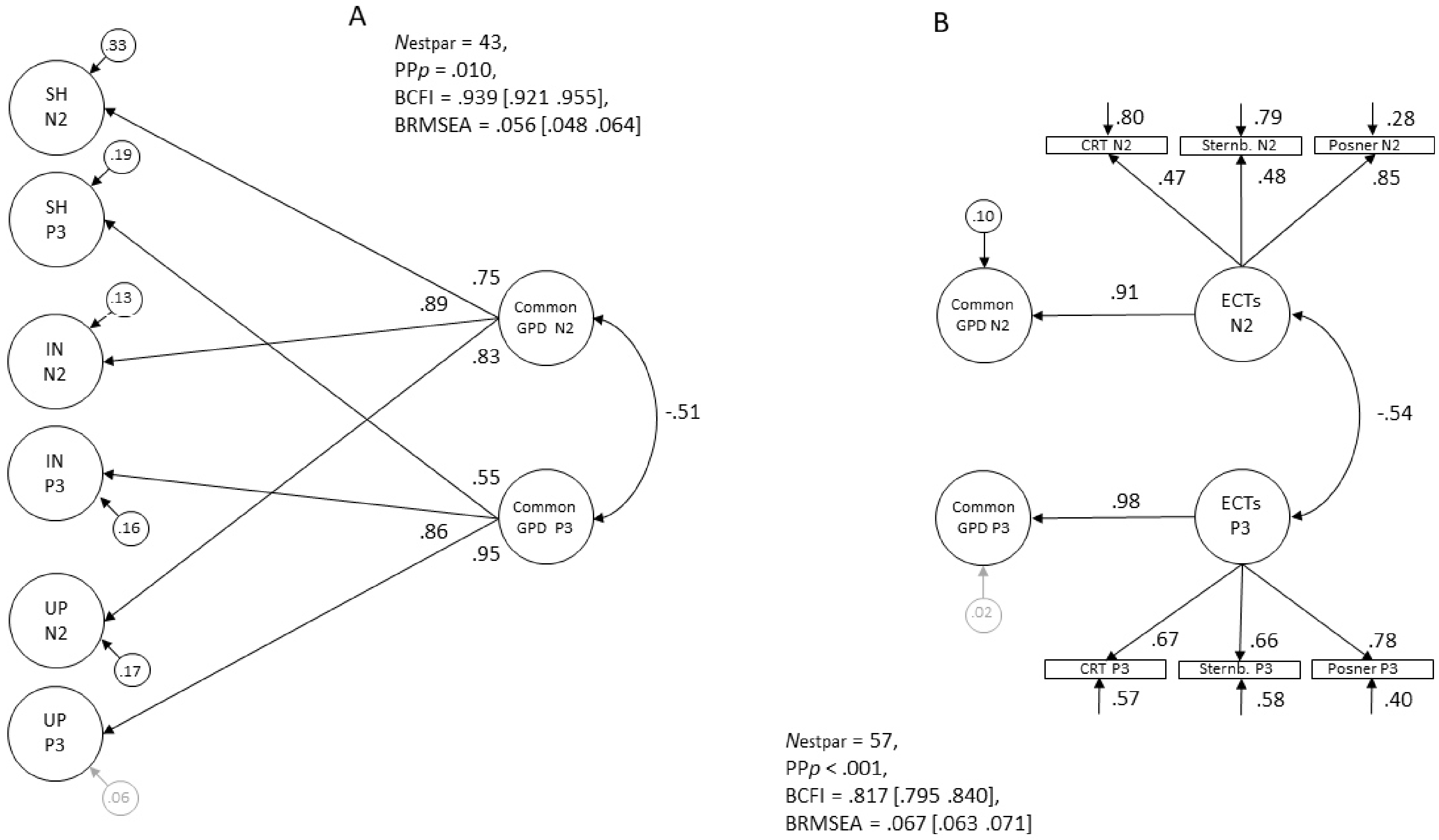


Note. The figure displays standardized path weights, unstandardized residual variances, and correlation coefficients alongside the corresponding paths. Any non-credible estimates are grayed out. OE-LM = odd/even-less/more task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task; *N*estpar = Number of estimated parameters; PP*p* = Posterior predictive *p* value; BCFI = Bayesian comparative fit index; BRMSEA = Bayesian root mean square error of approximation.

Given the correlations we found within the N2 and P3 ERP components, we refined our model to include two higher-order latent factors. This is a common factor for *greater processing demands* (GPD) for the N2 component and for the P3 component respectively, as shown in Figure 5A. There is one common factor that GPD for the N2 component, and there is one analog factor for the P3 component, as shown in Figure 5A. This revised model also fit the data well, evidenced by PP*p* (43) = .010, BCFI = .939, 95% CI [.921, .955], BRMSEA = .056, 95% CI [.048, .064]. A key finding in this model was the strong correlation between these two higher-order GPD factors, with a correlation coefficient of *r* = -.51, 95% CI [-.67, -.34]. This suggests a credible relationship between the GPD factors across the N2 and P3 components. Additionally, we observed credible residual variances for all lower-level latent factors, except for the P3 updating factor (see Figure 5A). These residual variances may indicate individual differences in specific EFs as captured by the ERP components. Prior to delving deeper into these individual differences, we aimed to discern the extent to which the common GPD factors for N2 and P3 represented shared variance in EFs and variance attributable to general, non-EF related decision processes.

To explore the influence of general, EF-unrelated processes on the N2 and P3 components, we developed a further model. This model incorporated two new factors representing these general processes, specifically within the N2 and P3 components. For this purpose, we used a battery of three simple decision tasks (ETCs) to construct these factors, referred to as ECTs N2 and ECTs P3 (illustrated in Figure 5B). This model showed a largely acceptable fit of the data, PP*p* (57) < .001, BCFI = .817, 95% CI [.795, .840], BRMSEA = .067, 95% CI [.063, .071]. A notable finding was that the common GPD P3 factor was completely accounted for by the EF-unrelated ECTs P3 factor, leaving no credible residual variance, $\sigma^2 = .002$, 95% PI [.000, .082]. Similarly, the common N2 factor was almost entirely explained by the EF-unrelated ECTs N2 factor. However, this factor did show credible residual variance, $\sigma^2 = .010$, 95% PI [.003, .239], suggesting that it might capture some EF-specific processes related to the N2 amplitudes across different EF tasks. Furthermore, the two general, EF-unrelated ERP factors correlated strongly negatively with each other $r = -.54$, 95% CI [-.68, -.39]. At the first latent level, all residuals exhibited credible variances, with the exceptions of the updating P3 ($\sigma^2 = .056$, 95% PI [.000, .165]) and the inhibition N2, $\sigma^2 = .062$, 95% PI [.000, .191].

Figure 5 *The hierarchical structure of the three factors of executive functions measured with N2 and P3 amplitudes*



Note. The figure displays standardized path weights, unstandardized residual variances, and correlation coefficients alongside the corresponding paths. Any non- credible estimates are grayed out. Note that the path coefficients linking the latent “ETCs” factors to the latent “Common” factors reflect regression weights, not factor loadings. CRT = Two Choice Reaction Time task; Posner = Posner task; Sternberg = Sternberg Memory task; GPD = Greater processing demands; *N*estpar = Number of estimated parameters; PP*p* = Posterior predictive *p* value; BCFI = Bayesian comparative fit index; BRMSEA = Bayesian root mean square error of approximation.

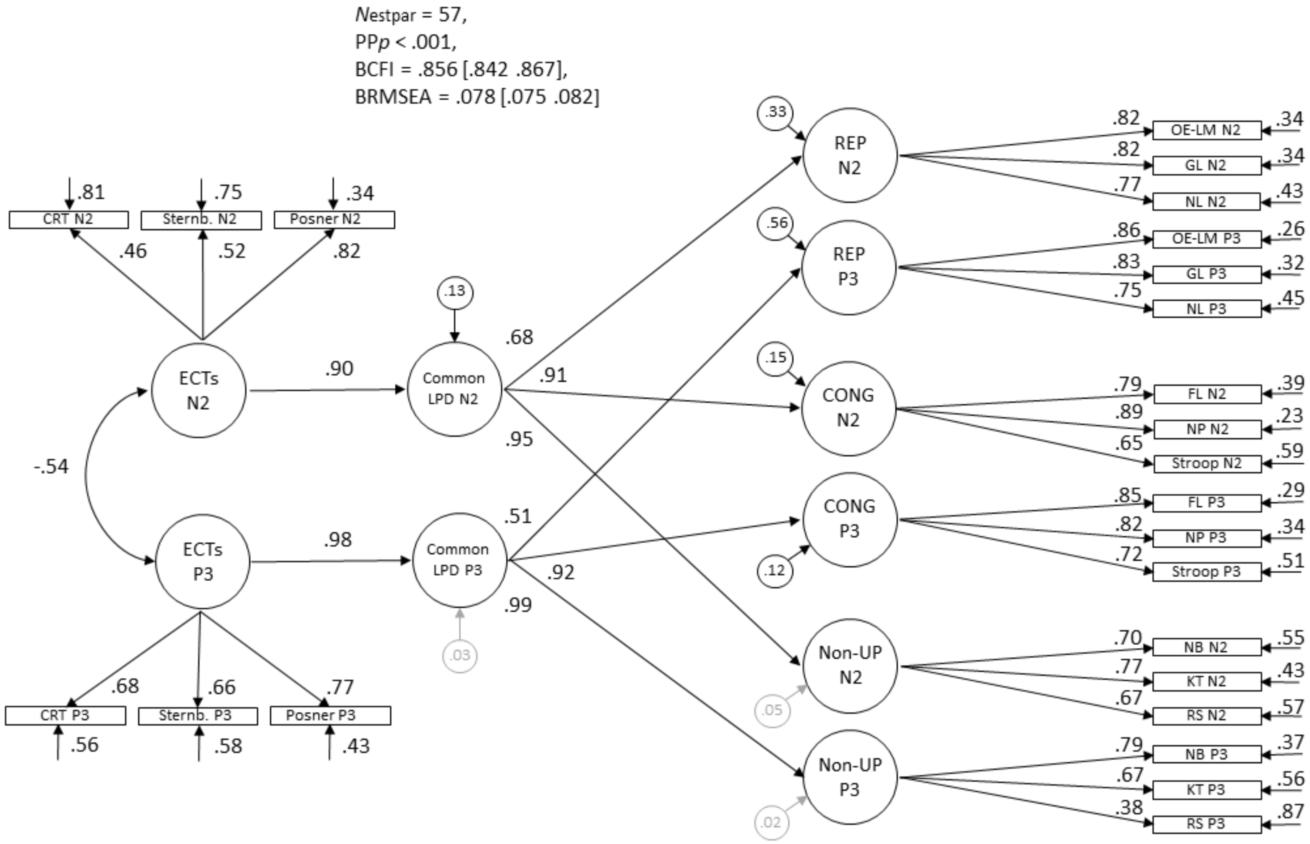
Testing the first-level residuals for specificity in EF processes

After we found credible residuals at the first latent level that were not attributable to general, EF-unrelated properties of ERP components, we sought to determine whether these residuals were specific to EF processes. Our objective was to discern if these residuals reflected variances specific to EFs, or if they were present in both lower and greater processing demands conditions of EF tasks. To this end, we adapted the previously described model (see Figure 5B) to include the lower processing demands conditions of the EF tasks, replacing the greater processing demands conditions. Again, this model included higher-order common N2 and P3 factors across EFs, which were controlled for the general processes reflected in N2 and P3 components measured in ECTs (see Figure 6). The fit of this new model to the data was largely acceptable, PPp (57) < .001, BCFI = .856, 95% CI [.842, .867], BRMSEA = .078, 95% CI [.075, .082].

Findings were similar to the model with greater processing demands: In this model, the common P3 factor across the conditions with lower processing demands was fully explained by general processes represented by the P3 ECT factor (σ^2 = .030, 95% PI [.000, .106]). The common N2 factor, however, again showed credible residual variance after controlling for the ECTs N2 factor, σ^2 = .129, 95% PI [.004, .283], mirroring observations from the previous greater processing demands model. At the first latent level, all residuals showed credible variances except for the non-updating N2 (σ^2 = .046, 95% PI [.000, .145]) and the non-updating P3 (σ^2 = .016, 95% PI [.000, .077]) factors.

The similarity in the structures between these two models raises an important question: do the residual variances in both models represent the same or different processes? If we follow the concept of additive processes in EF tasks, the residuals of the greater processing demand conditions should be independent of the lower processing demands conditions if they are specific to EFs. Conversely, if these residuals show strong correlations with their counterparts (e.g., the shifting P3 residuals with the repeat P3 residuals), it implies that these variances represent properties common to all tasks within a specific EF (e.g., specific task characteristics or heightened attentional control demands), but they may not be indicative of experimentally manipulated EF abilities.

Figure 6 *The three factors of the conditions with less processing demands controlled for general, EF-unrelated properties of the N2 and P3 components*



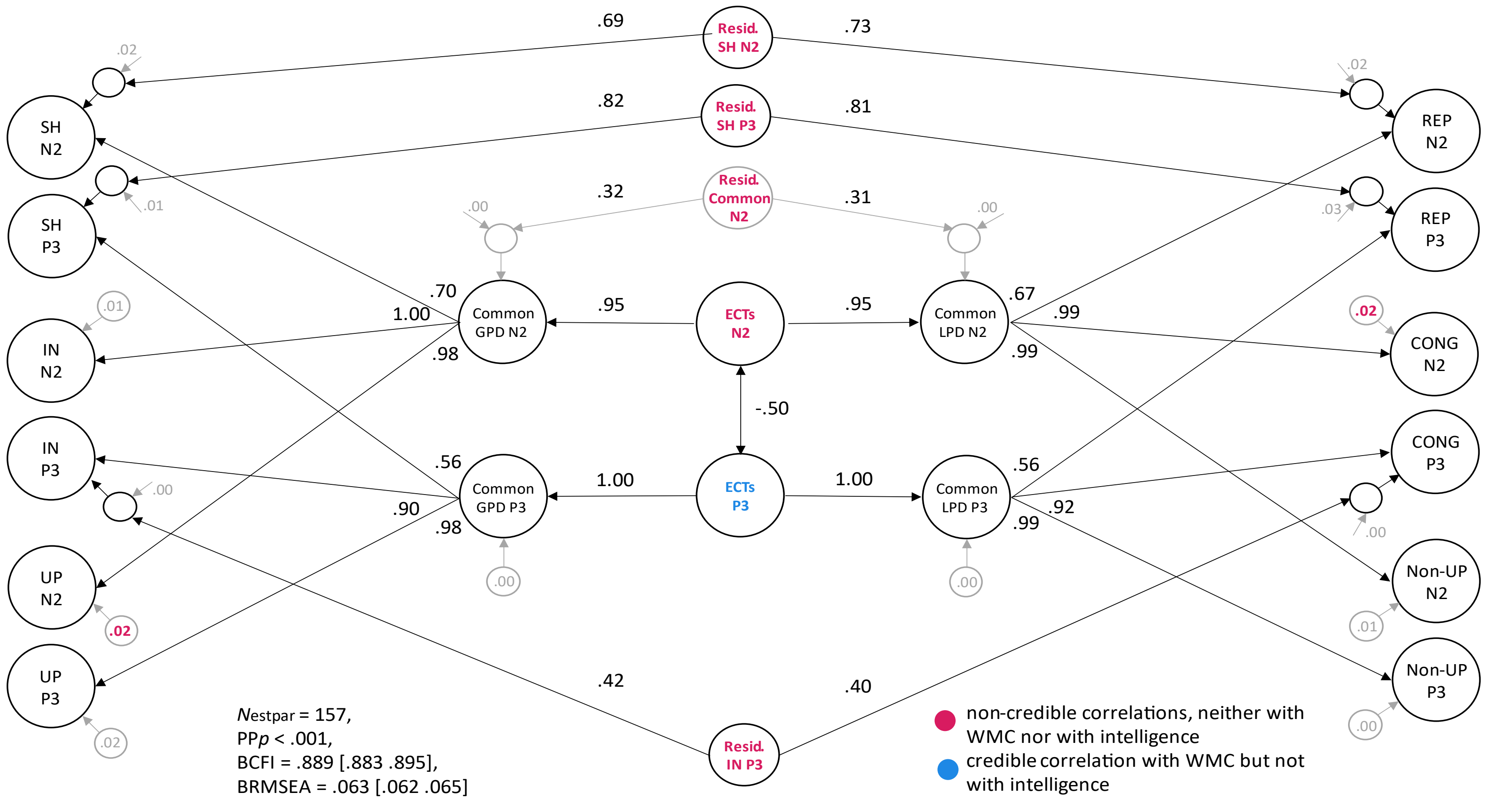
Note. The figure displays standardized path weights, unstandardized residual variances, and correlation coefficients alongside the corresponding paths. Any non-credible estimates are grayed out. Note that the path coefficients linking the latent "ETCs" factors to the latent "Common" factors reflect regression weights, not factor loadings. OE-LM = odd/even-less/more task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task; CRT = Two Choice Reaction Time task; Posner = Posner task; Sternberg = Sternberg Memory task. LPD = lower processing demands; *N*estpar = Number of estimated parameters; PP*p* = Posterior predictive *p* value; BCFI = Bayesian comparative fit index; BRMSEA = Bayesian root mean square error of approximation.

In consequence, we specified a model including the conditions with both greater and lower processing demands as indicator scores. This model (illustrated in Figure 7) maintained the hierarchical factor structure previously described for both conditions. We controlled the common factors for the general properties of the N2 and P3 components, as represented by the ECT factors. Additionally, we integrated common factors for credible residuals identified in our earlier models (shown in Figures 5 and 6). Specifically, we included common factors for the N2 and P3 residuals in both conditions of shifting tasks, for the P3 residuals in both conditions of inhibition tasks, and for the higher-order common N2 residuals in the conditions with lower and greater processing demands. A key aspect of our analysis was examining the relationship between these credible latent variances and higher-order cognitive abilities, namely intelligence and WMC. This examination was crucial to determine whether these residuals were specific to executive functions, as EF-specific abilities are anticipated to correlate with such higher-order cognitive abilities.

The model provided a largely acceptable account of the data, PPp (157) < .001, BCFI = .889, 95% CI [.883, .895], BRMSEA = .063, 95% CI [.062, .065]. The residuals of the N2 and P3 in shifting tasks each loaded onto a common factor across conditions with lower and greater processing demands, effectively leaving no variance specific to either the conditions with greater or lower processing demands at the first latent level. Similarly, the residuals of the P3 in inhibition tasks were accounted for by a common factor, again leaving no credible variance specific to either the conditions with greater or lower processing demands at the first latent level. For the common N2 factor across different EFs, we found no variance shared between lower and higher processing demands, and no specific variance for either condition. Furthermore, the residuals of the updating N2 in the greater processing demand conditions and the inhibition N2 in the lower processing demand conditions were no longer credible. This could be due to either the reduced power from estimating a large number of parameters compared to the previous models or the inclusion of task-specific residual correlations not necessary in the earlier models.

Interestingly, none of the latent factors showed a credible relationship with participants' intelligence or WMC. This includes even the common residual factors representing ERP amplitudes measured under both lower and higher processing demands. The only exception was the general, EF-unrelated P3 amplitude (measured in the speeded decision task battery), which exhibited a positive correlation with WMC (r = .32, 95% CI [.15, .48]). This finding suggests that individuals with higher P3 amplitudes tend to have greater WMC.

Figure 7 *The latent factors of the conditions with greater and lower processing demands controlled for general, EF-unrelated properties of the N2 and P3 components and the relations to intelligence and WMC*



Note. The figure displays standardized path weights, unstandardized residual variances, and correlation coefficients alongside the corresponding paths. Any non-credible estimates are grayed out. Note that the path coefficients linking the latent "ECTs" factors to the latent "Common" factors reflect regression weights, not factor loadings. The latent variables with colored names were correlated with intelligence and WMC; GPD = greater processing demands; LPD = Lower processing demands; *N*estpar = Number of estimated parameters; PP*p* = Posterior predictive *p* value; BCFI = Bayesian comparative fit index; BRMSEA = Bayesian root mean square error of approximation.

Discussion

In this study, we aimed to investigate the factor structure of EFs by employing ERP mean amplitudes as electrophysiological correlates measuring EF abilities. Specifically, we used the mean amplitudes of the fronto-central N2 and parieto-central P3 components, as previous studies have found experimental effects in EF tasks on these components (e.g., Bartholow et al., 2005; Dong et al., 2015; Gajewski et al., 2010; Heil et al., 2000; Pratt et al., 2011; Salmi et al., 2019; Scharinger et al., 2015; Watter et al., 2001; Yeung et al., 2004). Moreover, we used Bayesian SEMs to disentangle various sources of variance and to isolate EF-specific parts of ERP components from variance associated with participants' general, EF-unrelated processes. Our results revealed that the variances in N2 and P3 amplitudes in EF tasks could be decomposed into components specific to the tasks, general cognitive processing, and components specific to EF. Further analysis, however, indicated that these EF-specific components were present in experimental conditions both with and without induced cognitive control demands. Moreover, we found that these components were strongly correlated across conditions with varying levels of control demands. Ultimately, our findings raise questions about the validity of using N2 and P3 amplitudes as indicators of individual differences in EFs, especially when based on the additive framework commonly employed in typical EF tasks.

Experimental effects on ERP amplitudes

Shifting tasks

On the experimental level, we observed a mixed pattern of experimental effects. Specifically, participants showed larger negative N2 amplitudes and more positive P3 amplitudes in shifting trials compared to repeat trials in each of the shifting tasks, except for the N2 in the Odd-Even/Less-More task. These results align with our hypotheses and are consistent with previous research findings (e.g., Gajewski et al., 2010; Gajewski &

Falkenstein, 2011). Shifting trials require more of participants' cognitive control processes and contain enhanced task demands, reflected by enhanced N2 amplitudes (Folstein & Van Petten, 2008; Luck, 2014). Furthermore, the observed effects on the P3 amplitudes during shifting trials are in line with the interpretation of the P3 component as a process parameter reflecting context updating processes in response to relevant environmental changes (Donchin, 1981; Donchin & Coles, 1988). During shifting trials, participants have to update their response sets as the ongoing tasks changes for the next trial, leading to increased P3 amplitudes compared to repeat trials.

Inhibition tasks

In the inhibition tasks, our experimental results partially align with previous empirical findings and the theoretical interpretations of the N2 and P3 components. Consistent with previous research (e.g., Bartholow et al., 2005; Heil et al., 2000; Yeung et al., 2004), we observed larger N2 amplitudes in incongruent trials across most tasks, excluding the Negative-Priming task. These findings provide further evidence for the interpretation of the fronto-central N2 component as an ERP parameter reflecting attentional control processes (Folstein & Van Petten, 2008; Luck, 2014). The increased N2 amplitudes in incongruent trials suggest greater engagement of participants' attentional control processes in these challenging conditions in comparison to the congruent conditions.

However, our findings diverge when considering the P3 component. In our study, the P3 components in the three inhibition tasks did not show significant differences in mean amplitudes between the conditions with lower and greater processing demands. These results do not align with the findings of Groom & Cragg (2015) as well as of Pratt et al. (2011), who observed larger P3 amplitudes in incongruent compared to congruent trials in the Flanker task. According to Polich (2007), the P3 component reflects the allocation of attentional resources, which, in our interpretation, includes the allocation of executive processes.

However, Polich's (2007) perspective may consider a different understanding of attentional resources, one that does not conceive these attentional resources as reflecting the ability to ignore irrelevant distractors and maintain current task goals, which is at the core of the three inhibition tasks used in the present study. This discrepancy might explain why we observed a limited influence of inhibitory processes on the mean amplitude of the P3 component. In conclusion, our study suggests that the N2 component might be more effective in capturing inhibitory processes, whereas the P3 component showed a relative insensitivity to these processes.

Updating tasks

In the updating tasks, the experimental results for the N2 component were consistent with both previous findings and our hypotheses. Across all updating tasks, we observed a consistent pattern of smaller N2 amplitudes for stimuli with higher updating demands. These results align with the findings of Gevins et al. (1996) and Salmi et al. (2019), who reported smaller N2 amplitudes with higher n-back levels in the N-Back task. Our findings provide further support for the N2 as an ERP parameter reflecting cognitive control and executive function processes (Folstein & Van Petten, 2008; Luck, 2014).

The results for the P3 component, however, presented a more complex picture. In the N-Back task, we observed decreased P3 amplitudes with higher updating demands, in line with findings from previous research (Dong et al., 2015; Scharinger et al., 2015; Watter et al., 2001). However, we observed the opposite effect in the Running-Span task, where the P3 was larger for updated stimuli compared to stimuli requiring only memorization. Upon examining the grand average ERP morphology at the parieto-central position, we speculate that this P3 effect could partly be a continuation of the N2 effect. Further complicating the interpretation, the Keep-Track task showed no discernible differences in P3 amplitudes between more and less process-demanding stimuli. These inconsistent findings on the P3 amplitudes in updating tasks were unexpected, as the P3 component is often interpreted as an ERP process parameter reflecting working memory updating (Luck, 1998, 2014; Polich, 2007; Polich & Kok, 1995; Vogel et al., 1998; Vogel & Luck, 2002). One possible explanation for these findings could be the differences in the designs of the updating tasks, as the N-Back task differs significantly from the Running-Span and Keep-Track tasks, and it is well-known that task design modulates ERPs (e.g., Luck, 2014).

The factor structure of individual differences in ERP components

To study individual differences in ERP amplitudes, we used Bayesian SEMs to separate different sources of variance and examine the factor structure of EFs.

Problems measuring EFs using difference scores

One objective of our study was to investigate whether manifest difference scores would exhibit sufficient interrelations to support the three-factor structure of EF (Miyake et al., 2000). However, this model did not provide an adequate fit to the data, and we found that most ERP measures did not correlate across tasks. This aligns with findings from previous research, which have demonstrated that manifest difference scores in EF tasks based on behavioral measures often suffer from unreliability due to the large influence of unsystematic errors (Hedge et al., 2018; Rouder et al., 2023; Rouder & Haaf, 2019; Schubert et al., 2022).

No EF-specific variance in the conditions with greater processing demands

Because the N2 and P3 difference scores were not meaningfully correlated across tasks, we instead used the N2 and P3 amplitudes in the conditions with greater processing demands and controlled them for N2 and P3 amplitudes recorded in tasks with minimal attentional control demands (three simple decision tasks). This model initially confirmed the expected three-factor structure on the first latent level within each of the two ERP components, with moderate to strong intercorrelations within the ERP components, leading to a hierarchical structure. This structure included higher-order factors for both the N2 and P3 components, along with credible residual variances at the first latent level. These findings initially aligned with the classical unity/diversity framework of EFs with a higher-order factor for the N2 and the P3 components, respectively, along with some remaining credible residual variances at the first latent level. It appeared that EF abilities share common mechanisms contributing to individual differences in EF (the hierarchical factors). Additionally, each EF contains specific abilities, which are independent from the other EFs (the residual variances). These findings initially aligned with the classical unity/diversity framework of EFs with a higher-order factor for the N2 and the P3 components (e.g., Friedman et al., 2008; Friedman & Miyake, 2017; Miyake et al., 2000).

However, when we introduced factors representing general, EF-unrelated processes (the latent ECTs factors), these accounted for nearly all the variance in the hierarchical factors. The higher-order P3 factor was fully explained, and the higher-order N2 factor showed only a small residual variance. Further analysis, including a model with lower processing demands conditions, revealed a similar pattern, suggesting that the residuals might not be specific to the experimentally induced attentional control demands. Nevertheless, we found that some factors specific to shifting and inhibition tasks that loaded onto ERP amplitudes measured in conditions with lower as well as greater processing demands could not be explained by general, EF-unrelated factors. These factors are unlikely to represent the shifting and inhibition demands typically manipulated in EF experiments, as they were not exclusive to conditions with higher processing demands. Nonetheless, it is possible that these factors might be indicative of elevated attentional control demands. This is because tasks involving shifting and inhibition often require more conflict monitoring compared to simpler decision tasks that lack response or task-set conflicts. Therefore, these specific factors may reflect a different dimension of cognitive processing, potentially linked to general heightened top-down control in conflict tasks, rather than experimentally induced shifting and inhibition demands per se.

This interpretation aligns with the conflict monitoring theory put forth by Botvinick et al. (2001). According to this theory, the cognitive system is engaged in a continuous process of monitoring the level of conflict encountered during tasks. Depending on the degree of conflict, the system adjusts the extent of top-down control accordingly. This suggests that individual differences in the ability to monitor and respond to conflict, as well as in the modulation of top-down control, could manifest across both low and high processing demand conditions, particularly in tasks that inherently involve some form of conflict. Therefore, the factors we identified could be indicative of this broader conflict-monitoring and control mechanism, rather than being specific to the unique demands of shifting and inhibition.

To further assess the nature of these residual factors, we correlated them with intelligence and WMC, based on the premise that differences in attentional control should be related to higher-order cognitive abilities (e.g., Kane et al., 2008; Kovacs & Conway, 2016). However, no credible correlations were found, suggesting that these factors do not validly capture individual differences in EF-related processes.

In conclusion, our findings indicate that the shared processes among EF tasks are general and not specific to EF abilities. This mirrors our recent behavioral research findings (Löffler et al., 2024), where general, EF-unrelated processing abilities were found to account for the shared variance in EF tasks. This challenges the conventional use of EF tasks in individual differences research, suggesting that their shared variance is primarily indicative of general processing abilities, both at the electrophysiological and behavioral levels.

General, EF-unrelated P3 amplitudes correlated with WMC

We observed a positive correlation between the general P3 factor and WMC. Specifically, individuals with higher WMC tended to show larger P3 amplitudes in both EF as well as simple decision tasks. This is in line with previous research, which also reported a positive relationship between mean P3 amplitudes and WMC (e.g., Gevins, 2000). These findings lend support to the theoretical perspective that the P3 component generally reflects working memory updating processes across various cognitive tasks (Donchin, 1981; Donchin & Coles, 1988; Luck, 1998, 2014; Polich, 2007; Polich & Kok, 1995; Vogel et al., 1998; Vogel & Luck, 2002).

Limitations

In this study, we used two electrophysiological process parameters that are theoretically and experimentally associated with EFs. We measured individual differences in EF abilities by employing the mean amplitudes of the N2 and P3 components. Beyond the N2 and P3, there exist other promising process parameters derived from the EEG signal that could be valuable measurements in capturing individual differences in EFs.

In the time-domain, beside the N2 and P3 components, several other ERP components have been associated with response conflicts and EFs. For instance, differences in the N450 component (Larson et al., 2009; Rey-Mermet, Gade, & Steinhauser, 2019) and the conflict slow pattern (Larson et al., 2009) have been observed between congruent and incongruent conditions in the Stroop task. Additionally, the lateralized readiness potential (LRP) has been shown to reflect congruency effects in inhibition tasks (e.g., Kopp et al., 1996; Schubert et al., 2022; Willemssen et al., 2004).

Moving beyond ERPs, analysis of the frequency or time-frequency domains of the EEG potentially offers other process parameters that could be instrumental in exploring individual differences in EFs. Notably, the power of frequency bands, such as theta power, has shown differences in EF tasks under conditions of low and high processing demands (see for an overview: Cavanagh & Frank, 2014). Additionally, Klimesch (2012) suggested that the individual peak alpha frequency reflects differences in inhibitory control. However, a recently published study by Busch et al. (2024) demonstrated that individual peak alpha frequency and theta power, both measured in a resting state paradigm, did not correlate with individual differences in the performance in inhibition tasks. While the findings by Busch et al. (2024) make it evident that more research is needed to identify valid electrophysiological correlates of EFs, future research could benefit from considering electrophysiological parameters beyond ERPs to better understand the factor structure and individual differences in EFs.

Although our sample size is relatively large for an EEG study, it is important to note that even larger sample sizes could further enhance the robustness of our results, as they provide more accurate estimates of correlation coefficients (Kretzschmar & Gignac, 2019; Schönbrodt & Perugini, 2013).Thus, future research may benefit from replicating our findings in independent and even larger samples.

To further validate the generalizability of our findings, replicating this study with a diverse set of EF tasks would be insightful. Instead of solely relying on the three inhibition tasks used here, alternative tasks like the Simon task, the Stop-Signal task, or the Antisaccade task could be employed. Similarly, different updating and shifting tasks might be explored. However, it is important to acknowledge that the tasks used in our study are well-established and frequently employed in EF research, which lends credibility to our findings.

Conclusion

In the present study, we investigated the factor structure of EFs using electrophysiological measures. We used participants' mean amplitudes of the fronto-central N2 and the parieto-central P3 in executive function tasks and employed structural equation models to distinguish between variances related to general, EF-unrelated processing abilities and variances specific to executive functions.

Careful variance decomposition using structural equation models yielded three key findings: First, individual differences in N2 and P3 amplitudes in EF tasks were, to a large degree, accounted for by N2 and P3 amplitudes measured in simple decision tasks. Second, any variance specific to shifting and inhibition *not* accounted for by these general N2 and P3 factors was highly correlated across conditions with lower and greater control demands. This suggests that these ERP measures are unlikely to reflect the shifting and inhibition demands typically manipulated in EF experiments, but they may still measure overall elevated attentional control demands in conflict trials. Third, we found a positive correlation between P3 amplitudes measured across various tasks and WMC, which corroborates the theoretical perspective of the P3 component as a reflection of working memory updating processes.

Overall, our study brings to light a pivotal conclusion: executive function tasks do not capture interindividual differences in EF abilities in the way traditionally assumed, neither on the behavioral nor on the neurocognitive level. This raises significant questions about the use of such tasks in individual differences or developmental research and underscores the need for a reevaluation of how EF abilities are measured and understood.

Acknowledgments

We want to thank our research assistants – Johanna Hein, Florian Kaulhausen, Jan Göttmann, and Larissa Kunoff – who supported us with the recruitment of participants, data collection, project organization, and data management. This research was supported by the German Research Foundation (DFG) [grant numbers SCHU 3266/1-1 and SCHU 3266/2-1]. We declare no conflicts of interest. The preprocessed data and scripts supporting the findings of this study are available in the Open Science Framework repository at https://osf.io/a6zr9/ (Löffler et al., 2024) and the raw data and the materials are available in the Open Science Framework repository at https://osf.io/4pvz3/ (Löffler & Schubert, 2024).

References

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, *42*(1), 33–42. https://doi.org/10.1111/j.1469-8986.2005.00258.x

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/BF01797193

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295x.108.3.624

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–570. https://doi.org/10.3758/BRM.42.2.563

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Burgoyne, A. P., Tsukahara, J. S., Mashburn, C. A., Pak, R., & Engle, R. W. (2023). Nature and measurement of attention control. *Journal of Experimental Psychology: General*, *152*(8), 2369–2402. https://doi.org/10.1037/xge0001408

Busch, N., Geyer, T., & Zinchenko, A. (2024). Individual peak alpha frequency does not index individual differences in inhibitory cognitive control. *Psychophysiology*, *n/a*(n/a), e14586. https://doi.org/10.1111/psyp.14586

Cameron S. Carter, Matthew M. Botvinick, & Jonathan D. Cohen. (1999). The Contribution of the Anterior Cingulate Cortex to Executive Processes in Cognition. *Reviews in the Neurosciences*, *10*(1), 49–58. https://doi.org/10.1515/REVNEURO.1999.10.1.49

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*, *49*(5), 659–664. https://doi.org/10.1111/j.1469-8986.2011.01349.x

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421. https://doi.org/10.1016/j.tics.2014.04.012

Chen, D. D., Chan, J. S. Y., Pei, L., & Yan, J. H. (2012). Transfer Effects of Manipulating Temporal Constraints on Learning a Two-Choice Reaction Time Task with Low Stimulus-Response Compatibility. *Perceptual and Motor Skills*, *115*(2), 415–426. https://doi.org/10.2466/22.23.25.PMS.115.5.415-426

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

Donders, F. C. (1869). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1

Dong, S., Reder, L. M., Yao, Y., Liu, Y., & Chen, F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. *Brain Research*, *1616*, 146–156. https://doi.org/10.1016/j.brainres.2015.05.003

Draheim, C., Tshukara, J. S., & Engle, R. W. (2023). Replication and extension of the toolbox approach to measuring attention control. *Behavior Research Methods*. https://doi.org/10.3758/s13428-023-02140-2

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242–275. https://doi.org/10.1037/xge0000783

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, *47*(5), 1410–1430. https://doi.org/10.1037/a0023750

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual Differences in Executive Functions Are Almost Entirely Genetic in

Origin. *Journal of experimental psychology. General*, *137*(2), 201–225. https://doi.org/10.1037/0096-3445.137.2.201

Frischkorn, G., Schubert, A.-L., & Hagemann, D. (2019). Processing speed, working memory, and executive functions: Independent or inter-related predictors of general intelligence. *Intelligence*, *75*, 95–110. https://doi.org/10.1016/j.intell.2019.05.003

Gaillard, A. W. K. (1988). Problems and paradigms in ERP research. *Biological Psychology*, *26*(1), 91–109. https://doi.org/10.1016/0301-0511(88)90015-4

Gajewski, P. D., & Falkenstein, M. (2011). Diversity of the P3 in the task-switching paradigm. *Brain Research*, *1411*, 87–97. https://doi.org/10.1016/j.brainres.2011.07.010

Gajewski, P. D., Kleinsorge, T., & Falkenstein, M. (2010). Electrophysiological correlates of residual switch costs. *Cortex*, *46*(9), 1138–1148. https://doi.org/10.1016/j.cortex.2009.07.014

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

Gelman, A., & Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, *7*(4), 457–472.

Gevins, A. (2000). Neurophysiological Measures of Working Memory and Individual Differences in Cognitive Ability and Cognitive Style. *Cerebral Cortex*, *10*(9), 829–839. https://doi.org/10.1093/cercor/10.9.829

Gevins, A., Smith, M. E., Le, J., Leong, H., Bennett, J., Martin, N., McEvoy, L., Du, R., & Whitfield, S. (1996). High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroencephalography and Clinical Neurophysiology*, *98*(4), 327–348. https://doi.org/10.1016/0013-4694(96)00288-X

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations.

Consciousness and Cognition, *69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Groom, M. J., & Cragg, L. (2015). Differential modulation of the N2 and P3 event-related potentials by response conflict and inhibition. *Brain and Cognition*, *97*, 1–9. https://doi.org/10.1016/j.bandc.2015.04.004

Harrell, Jr., F. E. (2019). Package 'hmisc'. *CRAN2018, 2019*, 235–236.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Heil, M., Osman, A., Wiegelmann, J., Rolke, B., & Hennighausen, E. (2000). N200 in the Eriksen-Task: Inhibitory Executive Processes? *Journal of Psychophysiology*, *14*(4), 218–225. https://doi.org/10.1027//0269-8803.14.4.218

Hoofs, H., van de Schoot, R., Jansen, N. W. H., & Kant, Ij. (2018). Evaluating Model Fit in Bayesian Confirmatory Factor Analysis With Large Samples: Simulation Study Introducing the BRMSEA. *Educational and Psychological Measurement*, *78*(4), 537–568. https://doi.org/10.1177/0013164417709314

Hull, R., Martin, R. C., Beier, M. E., Lane, D., & Hamilton, A. C. (2008). Executive function in older adults: A structural equation modeling approach. *Neuropsychology*, *22*(4), 508–522. https://doi.org/10.1037/0894-4105.22.4.508

Ishihara, S. (2000). *The series of plates designed as a test for colour deficiency: 38 plates edition.* Kanehara.

Jäger, A. O., Süß, H., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: BIS-Test*. Hogrefe. https://madoc.bib.uni-mannheim.de/14578/

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Hrsg.), *Variation in Working Memory* (S.

21–48). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0002

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185. https://doi.org/10.1037/bul0000160

Klauer, K. C., Schmitz, F., Teige-Mocigemba, S., & Voss, A. (2010). Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology*, *63*(3), 595–619. https://doi.org/10.1080/17470210903076826

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16.

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*(3), 282–294. https://doi.org/10.1111/j.1469-8986.1996.tb00425.x

Kovacs, K., & Conway, A. R. A. (2016). Process Overlap Theory: A Unified Account of the General Factor of Intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kretzschmar, A., & Gignac, G. E. (2019). At what sample size do latent variable correlations stabilize? *Journal of Research in Personality*, *80*, 17–22. https://doi.org/10.1016/j.jrp.2019.03.007

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence*, *37*(4), 347–364. https://doi.org/10.1016/j.intell.2009.02.003

Larson, M. J., Kaufman, D. A. S., & Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia*, *47*(3), 663–670. https://doi.org/10.1016/j.neuropsychologia.2008.11.013

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods*, *42*(2), 571–585. https://doi.org/10.3758/BRM.42.2.571

Löffler, C., Frischkorn, G. T., Hagemann, D., Sadus, K., & Schubert, A.-L. (2024). The common factor of executive functions measures nothing but speed of information uptake. *Psychological Research*. https://doi.org/10.1007/s00426-023-01924-7

Löffler, C., Frischkorn, G. T., Sadus, K., Hagemann, D., & Schubert, A.-L. (2024, May 1). *The factor structure of executive functions measured with electrophysiological correlates: An event-related potential analysis*. https://doi.org/10.17605/OSF.IO/A6ZR9

Löffler, C., & Schubert, A.-L. (2024, May 1). *Neurocognitive process parameters of cognitive performance*. https://doi.org/10.17605/OSF.IO/4PVZ3

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00213

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

Luck, S. J. (1998). Neurophysiology of selective attention. In *Attention* (S. 257–295). Psychology Press/Erlbaum (UK) Taylor & Francis.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press.

Mashburn, C. A., Barnett, M. K., & Engle, R. W. (2023). Processing speed and executive attention as causes of intelligence. *Psychological Review*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/rev0000439

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773. https://doi.org/10.1080/10705511.2016.1186549

Merkle, E. C., & Rosseel, Y. (2018). blavaan: Bayesian Structural Equation Models via Parameter Expansion. *Journal of Statistical Software*, *85*, 1–30. https://doi.org/10.18637/jss.v085.i04

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383. https://doi.org/10.1016/0010-0285(77)90012-3

Nebe, S., Reutter, M., Baker, D. H., Bölte, J., Domes, G., Gamer, M., Gärtner, A., Gießing, C., Gurr, C., Hilger, K., Jawinski, P., Kulke, L., Lischke, A., Markett, S., Meier, M., Merz, C. J., Popov, T., Puhlmann, L. M., Quintana, D. S., … Feld, G. B. (2023). Enhancing precision in human neuroscience. *eLife*, *12*, e85980. https://doi.org/10.7554/eLife.85980

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task:

Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience*, *3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, *3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, *41*(2), 103–146. https://doi.org/10.1016/0301-0511(95)05130-9

Posner, M. I., & Mitchell, R. F. (1967). Chronometric analysis of classification. *Psychological Review*, *74*(5), 392–409. https://doi.org/10.1037/h0024913

Pratt, N., Willoughby, A., & Swick, D. (2011). Effects of Working Memory Load on Visual Selective Attention: Behavioral and Electrophysiological Evidence. *Frontiers in Human Neuroscience*, *5*. https://www.frontiersin.org/articles/10.3389/fnhum.2011.00057

R. Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing.* URL https://www.R-project.org/

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

Revelle, W. (2020). psych: Procedures for Psychological, Psychometric, and Personality Research. In *Northwestern University, Evanston, Illinois. R package version 2.0.12,*. https://CRAN.R-project.org/package=psych.

Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 501–526. https://doi.org/10.1037/xlm0000450

Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *148*(8), 1335–1372. https://doi.org/10.1037/xge0000593

Rey-Mermet, A., Gade, M., & Steinhauser, M. (2019). Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage*, *188*, 411–418. https://doi.org/10.1016/j.neuroimage.2018.12.031

Rogers, R. D., & Monsell, S. (1995). Costs of a predictible switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*(2), 207–231. https://doi.org/10.1037/0096-3445.124.2.207

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467. https://doi.org/10.3758/s13423-018-1558-y

Rouder, J. N., Kumar, A., & Haaf, J. M. (2023). Why many studies of individual differences with inhibition tasks may not localize correlations. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02293-3

Salmi, J., Vilà-Balló, A., Soveri, A., Rostan, C., Rodríguez-Fornells, A., Lehtonen, M., & Laine, M. (2019). Working memory updating training modulates a cascade of event-related potentials depending on task load. *Neurobiology of Learning and Memory*, *166*, 107085. https://doi.org/10.1016/j.nlm.2019.107085

Scharinger, C., Soutschek, A., Schubert, T., & Gerjets, P. (2015). When flanker meets the n-back: What EEG and pupil dilation data reveal about the interplay between the two central-executive working memory functions inhibition and updating. *Psychophysiology*, *52*(10), 1293–1304. https://doi.org/10.1111/psyp.12500

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schubert, A.-L., Löffler, C., & Hagemann, D. (2022). A neurocognitive psychometrics account of individual differences in attentional control. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001184

Schubert, A.-L., Löffler, C., Hagemann, D., & Sadus, K. (2023). How robust is the relationship between neural processing speed and cognitive abilities? *Psychophysiology*, *60*(2), e14165. https://doi.org/10.1111/psyp.14165

Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*(2), 850–886. https://doi.org/10.1037/a0033981

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421–457.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Sudevan, P., & Taylor, D. A. (1987). The cuing and priming of cognitive operations. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 89–103. https://doi.org/10.1037/0096-1523.13.1.89

Tipper, S. P., & Cranston, M. (1985). Selective Attention and Priming: Inhibitory and Facilitatory Effects of Ignored Primes. *The Quarterly Journal of Experimental Psychology Section A*, *37*(4), 591–611. https://doi.org/10.1080/14640748508400921

Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743. https://doi.org/10.3758/BF03196329

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1656–1674. https://doi.org/10.1037/0096-1523.24.6.1656

von Bastian, C. C., Blais, C., Brewer, G., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., … Wiemers, E. (2020). *Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations*. PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology*, *38*(6), 998–1003. https://doi.org/10.1111/1469-8986.3860998

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

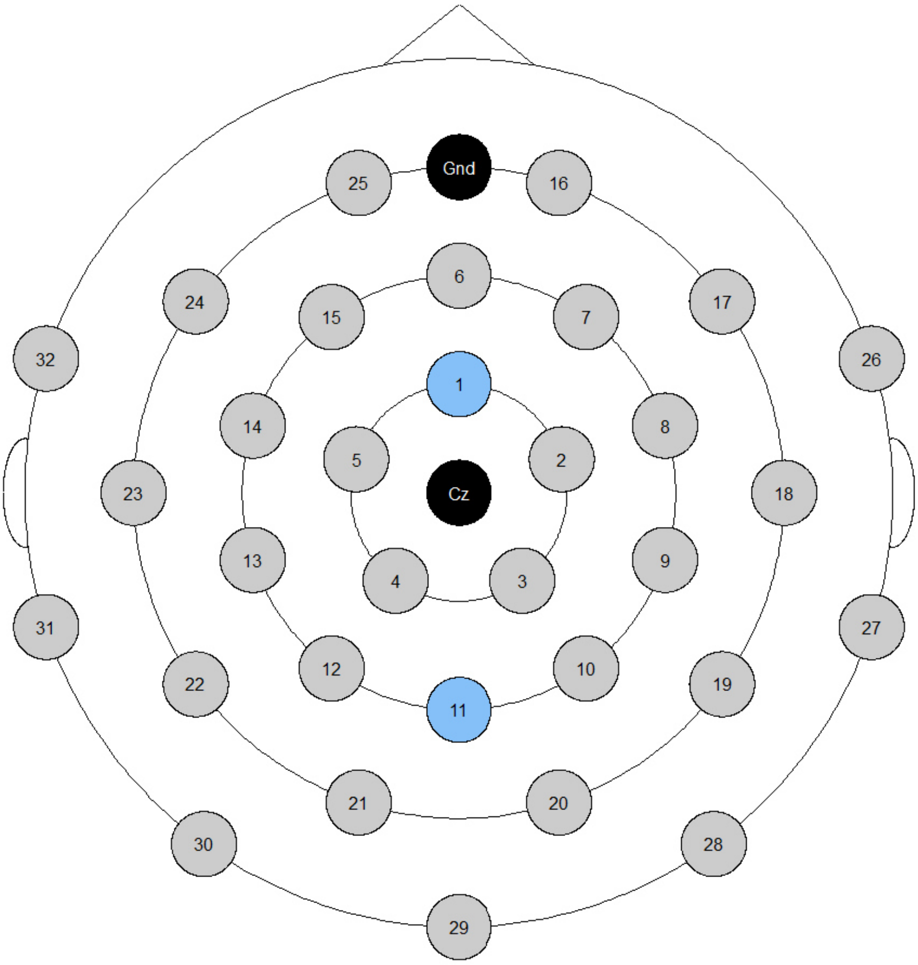
Wilhelm, O., Hildebrandt, A. H., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *0*. https://doi.org/10.3389/fpsyg.2013.00433

Willemssen, R., Hoormann, J., Hohnsbein, J., & Falkenstein, M. (2004). Central and parietal event-related lateralizations in a flanker task. *Psychophysiology*, *41*(5), 762–771. https://doi.org/10.1111/j.1469-8986.2004.00201.x

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931

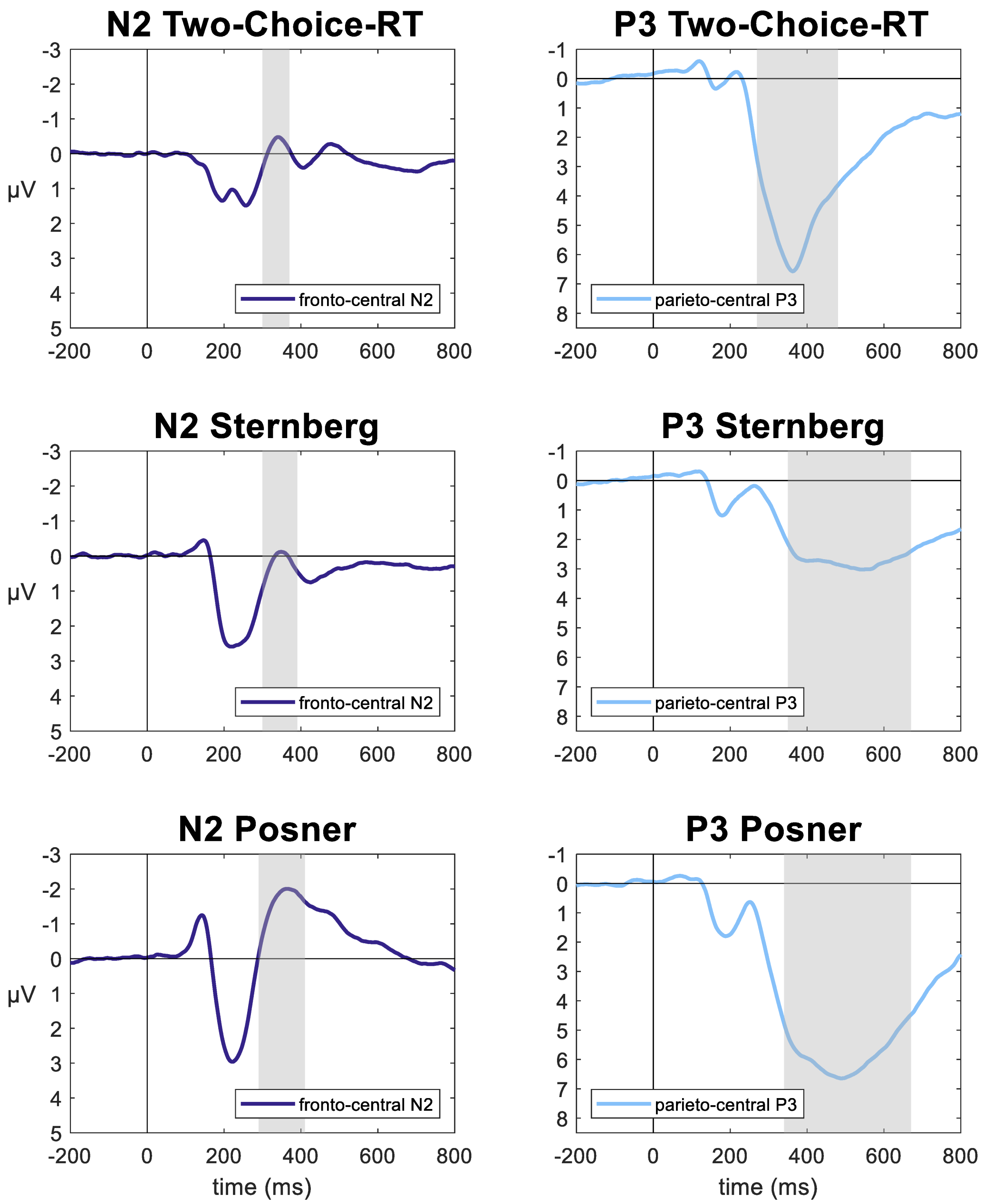
Supplementary Materials

Figure S1 *Electrode montage of 32 equidistant electrodes*



Note. We measured the N2 on the fronto-central electrode 1, and the P3 on parieto-central electrode 11.

Figure S2 *The ERP wave forms of the fronto-central N2 and parieto-central P3 in the elementary cognitive tasks (ECTs)*



Note. The ERPs are locked to the probe stimuli in the corresponding tasks. The gray shaded areas indicate the time intervals in which the mean amplitudes were measured.

Table S1 *The RGB color codes for the colors used in the experminental tasks*

Color	R	G	B
black	0	0	0
gray	140	140	140
green	0	178	30
red	255	25	32
orange	255	150	25
yellow	255	255	25
turquoise	0	255	255
blue	25	50	255
pink	255	25	255

Table S2 *The experminental effects of the EF tasks on the behavioral level*

Task	Dependent variable	*t*-value	*df*	*p*	Cohen's *d*	95% CI
OE-LM	Mean RT	9.69	136	<.001	0.83	0.63 - 1.02
GL	Mean RT	15.02	141	<.001	1.26	1.04 - 1.48
NL	Mean RT	9.94	129	<.001	0.87	0.67 - 1.07
FL	Mean RT	17.34	141	<.001	1.46	1.22 - 1.69
NP	Mean RT	9.361	127	<.001	0.83	0.63 - 1.03
Stroop	Mean RT	22.07	141	<.001	1.85	1.58 - 2.12
NB	A.T. proportion	12.98	132	<.001	1.13	0.91 - 1-34
KT	A.T. proportion	10.87	121	<.001	0.98	0.77 - 1.20
RS	A.T. proportion	8.25	131	<.001	0.72	0.53 - 0.91

Note OE-LM = odd/even-less/more task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task; Mean RT = Mean reaction time; A.T. proportion = Arcsine transformed proportion correct; 95% CI represents the lower and upper boundary of the 95 % confidence interval for Cohen's *d*.

Table S3 *Relibilities of the ERP difference scores*

Task	Reliability N2	Reliability P3
OE-LM	-.11	.30
GL	-.04	.32
NL	.27	.04
FL	.32	.12
NP	-.03	-.08
Stroop	.03	-.16
NB	.42	.57
KT	-.01	.49
RS	.67	.74

Note OE-LM = odd/even-less/more task; GL = Global/Local task; NL = Number/Letter task; FL = Arrow Flanker task; NP = Negative Priming task; Stroop = Stroop task; NB = N-Back task; KT = Keep-Track task; RS = Running-Span task. Reliability scores were estimated with Spearman-Brown corrected correlations.

Declaration in accordance to § 8 (1) c) and d) of the doctoral degree regulation of the Faculty

FAKULTÄT FÜR VERHALTENS- UND EMPIRISCHE KULTURWISSENSCHAFTEN

UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

Promotionsausschuss der Fakultät für Verhaltens- und Empirische Kulturwissenschaften der Ruprecht-Karls-Universität Heidelberg / Doctoral Committee of the Faculty of Behavioural and Cultural Studies of Heidelberg University

Erklärung gemäß § 8 (1) c) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) c) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation selbstständig angefertigt, nur die angegebenen Hilfsmittel benutzt und die Zitate gekennzeichnet habe. / I declare that I have made the submitted dissertation independently, using only the specified tools and have correctly marked all quotations.

Erklärung gemäß § 8 (1) d) der Promotionsordnung der Universität Heidelberg für die Fakultät für Verhaltens- und Empirische Kulturwissenschaften / Declaration in accordance to § 8 (1) d) of the doctoral degree regulation of Heidelberg University, Faculty of Behavioural and Cultural Studies

Ich erkläre, dass ich die vorgelegte Dissertation in dieser oder einer anderen Form nicht anderweitig als Prüfungsarbeit verwendet oder einer anderen Fakultät als Dissertation vorgelegt habe. / I declare that I did not use the submitted dissertation in this or any other form as an examination paper until now and that I did not submit it in another faculty.

Vorname Nachname / First name Family name	Christoph Löffler
Datum / Date	20.01.2025
Unterschrift / Signature	

Dem Dekanat der Fakultät für Verhaltens- und Empirische Kulturwissenschaften liegt eine unterschriebene Version dieser Erklärung vom 20.01.2025 vor.

The Dean's Office of the Faculty of Behavioural and Cultural Studies has a signed version of this declaration dated January 20, 2025.