Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations

Claudia C. von Bastian, University of Sheffield
Chris Blais, Arizona State University
Gene A. Brewer, Arizona State University
Máté Gyurkovics, University of Illinois at Urbana-Champaign
Craig Hedge, Cardiff University
Patrycja Kałamała, Jagiellonian University in Krakow
Matt E. Meier, Western Carolina University
Klaus Oberauer, University of Zurich
Alodie Rey-Mermet, Swiss Distance University Institute
Jeffrey N. Rouder, University of California Irvine
Alessandra S. Souza, University of Zurich
Lea M. Bartsch, University of Zurich
Andrew R. A. Conway, Claremont Graduate University
Christopher Draheim, Georgia Institute of Technology
Randall W. Engle, Georgia Institute of Technology
Naomi P. Friedman, University of Colorado Boulder
Gidon T. Frischkorn, University of Zurich
Daniel E. Gustavson, Vanderbilt University Medical Center
Iring Koch, RWTH Aachen
Thomas S. Redick, Purdue University
Bridget A. Smeekens, University of North Carolina at Greensboro
Peter S. Whitehead, Duke University
Elizabeth A. Wiemers, Purdue University

Word count: 14,343

Author Contributions

The manuscript is based on presentations and discussions at a one-day workshop ("Attentional control hackathon: Does an attentional control construct exist and, if so, how can we reliably measure it?") held in November 2018 in New Orleans, LA, which was organized by Claudia von Bastian (CVB) and Michael Kane. CVB drafted the structure and coordinated the writing of the manuscript. CVB and Gene Brewer drafted the introduction. Matt Meier (MM), Klaus Oberauer, and Alodie Rey-Mermet drafted the section on theoretical and psychometric considerations. Máté Gyurkovics and Patrycja Kałamała drafted the empirical evidence section, with contributions from Naomi Friedman and Daniel Gustavson. Alessandra Souza (AS) and CVB conducted the survey and drafted the assessment section with contributions from Thomas Redick. Craig Hedge, Chris Blais, and Jeffrey Rouder drafted the section on analytical considerations with contributions from AS and CVB. CVB and MM drafted the summary and conclusion section. All authors provided comments and/or revisions to the manuscript. In addition, we gratefully acknowledge the contributions of Michael Kane who co-organised the workshop, and Eddie Christopher, Julia Haaf, Aakriti Kumar, Vanessa Loaiza, Jessie Martin, Cody Mashburn, Akira Miyake, Matthew Robison, Kevin Rosales, Jean-paul Snijder, Jason Tsukahara, and Matt Welhalf, who attended the workshop and contributed to the discussions covered in this manuscript.

The data and analysis scripts are available on the Open Science Framework (https://osf.io/kw2a9/).

Correspondence concerning this article should be addressed to Claudia C. von Bastian, Department of Psychology, University of Sheffield, Cathedral Court, 1 Vicar Lane, Sheffield S1 2LT. E-mail: c.c.vonbastian@sheffield.ac.uk

Abstract

Attentional control as an ability to regulate information processing during goal-directed behavior is critical to many theories of human cognition and thought to predict a large range of everyday behaviors. However, in recent years, failures to reliably assess individual differences in attentional control have sparked a debate concerning whether attentional control, as currently conceptualized and assessed, can be regarded as a valid psychometric construct. In this consensus paper, we summarize the current debate from theoretical, methodological, and analytical perspectives. First, we propose a consensus-based definition of attentional control and the cognitive mechanisms that potentially contribute to individual differences in attentional control. Next, guided by the findings of an in-depth literature survey, we discuss the psychometric considerations that are critical when assessing attentional control. We then provide suggestions for recent methodological and analytical approaches that can alleviate the most common concerns. We conclude that, to truly advance our understanding of the construct of attentional control, we must develop a theory-driven and empirically supported consensus on how we define, operationalize, and assess attentional control. This consensus paper presents a first step to achieve this goal; a shift toward transparent reporting, sharing of materials and data, and cross-laboratory efforts will further accelerate progress.

Advancing the understanding of individual differences in attentional control: Theoretical, methodological, and analytical considerations

Attentional control – also referred to as cognitive control, executive function, or executive control – reflects the ability to regulate information processing during goal-directed behavior (E. K. Miller & Cohen, 2001; Miyake et al., 2000; Posner & Snyder, 1979). Attentional control measures have been associated with many psychologically meaningful outcomes, including psychopathology (Snyder, Miyake, & Hankin, 2015), intelligence (Kane & Engle, 2002), cognitive development (Blakemore & Choudhury, 2006; Zelazo & Müller, 2002), impulsivity (Sharma, Markon, & Clark, 2014), addiction (Hester & Garavan, 2004), and age-related cognitive declines (Hasher, Stoltzfus, Zacks, & Rypma, 1991). Given the predictive utility of attentional control across these domains, it is critical for researchers to better understand the structure, function, and mechanisms underlying attentional control, and how it is implemented to direct and redirect information processing during goal pursuit. Experimental manifestations of attentional control are manifold and include varieties of the congruency or conflict effect (Posner & Snyder, 1979; Stroop, 1935), response inhibition (Hallett, 1978), error-related slowing (Rabbitt, 1966), sequential congruency effects (Gratton, Coles, & Donchin, 1992), costs of switching between tasks or completing multiple tasks simultaneously (Koch, Poljac, Müller, & Kiesel, 2018), and monitoring and updating of information (Miyake et al., 2000).

Attentional control effects are often measured by contrasting experimental conditions demanding high levels of attentional control to conditions demanding low levels of attentional control, allowing for subtraction of performance measures between these conditions to isolate attentional control processes. A classic example is the Stroop effect, in which participants are prompted to indicate the print color of a word while ignoring the semantic meaning of the word. If the word's print color and meaning are congruent (e.g., "blue" printed in blue ink), little attentional control is required to name the print color. However, if print color and meaning are incongruent (e.g., "blue" printed in green ink), participants need extra time to overcome this conflicting information before responding accurately. The difference in performance between conditions – the congruency effect – is then typically interpreted as a proxy of attentional control processes. Smaller congruency effects are assumed to reflect more efficient overcoming of conflicting information, which is thought of as better attentional control. Experimental attentional control effects are replicable and easily demonstrated across a wide range of tasks, motivating the hypothesis that individual variation in the size of these experimental effects reflects a common, more general attentional control ability.

Researchers interested in individual differences in attentional control ability typically combine experimental and differential approaches to psychological research. Specifically, experimental effect sizes derived from diverse attentional control tasks (see Table 1 for an overview of commonly used paradigms) are assessed for each participant, and the variance shared between these effects is interpreted as a hypothetical psychometric construct representing general attentional control ability. This construct is then correlated with other outcomes such as academic achievement. In this example, the logic is that individuals with greater ability to control interference in the attentional control task will likely also be academically more successful.

Based on the assumption that they tap the same psychometric construct, experimental effects derived from different attentional control tasks such as the Stroop and the Simon task have often been treated interchangeably as markers of attentional control. Critically, however, some research has reported only weak correlations amongst attentional control effects (Friedman et al., 2016; Frischkorn, Schubert, & Hagemann, 2019; Gustavson, Panizzon, Franz, et al., 2018; Keye, Wilhelm, Oberauer, & van Ravenzwaaij, 2009; Paap & Greenberg, 2013; Rey-Mermet, Gade, Souza, von Bastian, & Oberauer, 2019; Shilling, Chetwynd, & Rabbitt, 2002; von Bastian, Souza, & Gade, 2016; Whitehead, Brewer, & Blais, 2019), which has not inspired much confidence in attentional control as a sound psychometric construct (Draheim, Mashburn, Martin, & Engle, 2019; Hedge, Powell, & Sumner, 2018; Paap & Sawi, 2016; Rey-Mermet, Gade, & Oberauer, 2018; Rouder & Haaf, 2019). From a differential psychologist's perspective, weak correlations are problematic for regarding attentional control a psychometrically valid construct, as they suggest that either individual differences cannot be measured reliably, or that there is only a small degree of overlap in variance reflected by conceptually similar empirical indicators of the construct.

Much of the difficulties in establishing a coherent psychometric construct of attentional control arises from the use of experimental effects to measure individual differences. This is different to other areas of individual differences research (e.g., personality or intelligence), where most measures were designed to capture differences between people directly without relying on comparisons between experimental conditions. The combination of experimental and differential approaches to psychology effectively traverses the "desert," as Cronbach (1957) referred to the divide between these two research traditions in his APA presidential address. Considering both traditions side-by-side demonstrates the inherent methodological challenge of this approach to use experimental effects as indicators of an ability (e.g., Tucker-Drob, 2011). Experimental psychologists are primarily interested in how functional relations between independent variables are affected by environmental variations. To maximize these effects, experimental psychologists aim to reduce unsystematic sources of variance by using within-subject designs. Cognitive researchers have a long history of exploiting this approach to isolate cognitive processes. In contrast, differential psychologists characterize this source of unsystematic variance as individual differences to study how people's differences on certain psychological dimensions are related to their differences on other dimensions. The use of within-subjects experimental contrasts as indicators of individual differences of an ability inherits the strength of the experimental approach of isolating specific processes (such as the processes involved in overcoming cognitive conflict), but this comes at the cost of subtracting out the largest part of the variance between individuals, risking that there is too little true variance left for a reliable and valid psychometric measure.

The current state of research on individual differences in attentional control is that weak between-task correlations are observed despite the robust experimental demonstrations of attentional control. This raises important unresolved questions: Are individual differences in attentional control effects generalizable across tasks or are they task-specific in nature? Are the paradigms currently in use a reliable and valid way to capture an individual's ability to control their attention? Given the central role of attentional control in our understanding of individual differences in human cognition, it is imperative to address these questions. In this paper, we therefore consider the construct of attentional control from theoretical, methodological, and analytical perspectives and discuss critical concerns regarding current approaches in researching individual differences in attentional control. First, we argue that, as attentional control researchers, we need to understand better *what* theoretical properties of attentional control we aim to assess. Second, based on a survey of the current body of literature, we critically examine *how* individual differences in attentional control are assessed. Third, we discuss the problems arising from traditional ways to *analyze* attentional control data and present recent advancements with the potential to alleviate at least some of the highlighted concerns. However, we also lay out that past attempts to improve the psychometric properties of the most prominent measures, such as the Stroop task, were not particularly successful. Therefore, we suggest that it is time to move beyond these traditional measures and, as a research field, rethink how to assess attentional control in order to better understand it as a psychometric construct. A recurrent theme is that transparent reporting, sharing of materials and data, and cross-laboratory efforts are necessary to accelerate progress in the operational definition, measurement and, eventually, understanding of individual differences in attentional control.

Attentional Control as an Individual Differences Construct: Theoretical and Psychometric Considerations

The attentional-control construct was first used to explain impaired behavior when patients were confronted with novel situations (Burgess, 1997; Norman & Shallice, 1980). Attentional control was conceptualized as a necessary process occurring in tasks that require participants to: (1) specify or formulate a goal; (2) plan sequences of behaviors and actions to reach this goal; (3) compare the different sequences with respect to their success and efficiency; (4) select the most appropriate sequence; (5) initiate the selected plan; (6) monitor performance when carrying out the sequence of behaviors or actions; and (7) if necessary, to suppress alternative action tendencies (Norman & Shallice, 1986; Rabbitt, 1997). This conceptualization of attentional control is so broad that it makes it difficult to isolate attentional control from other constructs. For example, with this definition, attentional control is difficult to distinguish from intelligence because both constructs represent the ability to reason and solve problems.

To enable a better understanding of the construct and an efficient accumulation of knowledge, it is desirable to create a definition that is specific enough to distinguish attention control from other cognitive-ability constructs but still broad enough to encompass behavior in a variety of situations. Attentional control has been defined in several ways across the years. Here, we propose a consensus that captures the common core of existing definitions: The basic ability in attentional control is maintaining an operative goal, and goal-relevant information, in the face of distraction. Here, we define an operative goal as one that substantially influences current information processing and action. We define distraction as a continuous variable, reflecting the degree to which goal-irrelevant information receives priority in information processing over goal-relevant information. We distinguish among three types of distraction: that caused by the perceived environment, by self-generated information, or by habits (Friedman & Miyake, 2004; Hasher, Lustig, & Zacks, 2007; Nigg, 2000; Stahl et al., 2014).

First, with *distraction caused by the perceived environment*, attention should be controlled to avoid being captured by salient events. For example, when you are driving home, your attention could be captured by a billboard showing an advertisement for a new restaurant. Therefore, to pursue your goal of safely driving home, your attention should be directed to the traffic. In the lab, such situations are often operationalized with tasks such as the flanker task or the attention-capture paradigm (see Table 1).

Second, with *distraction caused by self-generated information*, attention must be controlled to avoid capture by irrelevant thoughts (i.e., mind wandering) or no-longer-relevant memory contents (e.g., proactive interference). In the example of driving home, your attention could be captured by things you intend to do once you arrive home. Methods used to assess this source of distraction are, for example, thought probes presented during extended periods of an attention-demanding task to measure mind wandering (Weinstein, 2018; Wiemers & Redick, 2019), the recall of paired associates in which the same retrieval cue is paired with different recall targets for measuring proactive interference, or overcoming the interference from no-longer relevant memory contents in an updating task (Table 1).

Third, with *distraction caused by habits*, attention must be controlled to avoid or stop the performance of a habit. For example, when approaching a red traffic light, your habit is to stop. However, if you see a police officer signaling you to move forward, you should suppress this habit to follow the police officer’s instructions. In the laboratory, this source of distraction is typically assessed with the Stroop, Simon, antisaccade, global-local, go/no-go, and stop-signal tasks (Table 1). Task switching may represent a special case for this source of distraction. Here, participants are asked to *alternate* between two tasks and, to minimize task-switch costs, overcome the short-term habit of repeating the last-executed task.

Table 1

Description of Paradigms Typically Used to Measure Attentional Control

Paradigm	Description	Common Measures
	Distraction caused by the perceived environment	
Flanker	Participants see a row of characters. They are asked to identify the central stimulus (e.g., an arrow or a letter) while ignoring the flanking stimuli. In an incongruent trial, the central stimulus and the flanking stimuli afford a different response; in a congruent trial, they afford the same response; in a neutral trial, the flanking stimuli are associated with none of the responses or only the target is presented.	Performance difference between incongruent and congruent trials or between incongruent and neutral trials (flanker congruency or conflict effect) or performance in the incongruent trials only
Matching	Participants decide whether two visual objects (e.g., shapes) are the same (match) or not (mismatch). In the interference condition, one of the visual objects is overlaid by another object (a distractor) that should be ignored. No distractor is presented in the control condition.	Performance difference between the interference and the control condition
Search	Participants search for a target item (e.g., red square) among distractors (e.g., green circles). In attentional-capture variants, performance in capture conditions where one distractor has one of the target-defining features (e.g., a red circle) is compared to performance in no-capture conditions where the distractors have no target-defining features. In cued-search variants, participants are asked to report features of stimuli (e.g., whether the letter F is presented mirrored or normal) shown at (typically 2 or 4) cued locations. In the non-cued locations, lure stimuli (e.g., other Fs) are presented that should be ignored.	Average time to complete the search trials or performance difference between capture- and no-capture conditions

	Distraction caused by self-generated information	
Keep track	Participants keep track of a stream of stimuli (e.g., letters) of unpredictable length. In some variants, participants are asked to report the most recent stimuli (e.g., the last four letters). In other variants, the stimuli are sampled from different categories (e.g., animals, colors, and countries) and participants are asked to recall the most recent exemplar from each category. Variants designed for children, older adults, or patients (e.g., self-ordered pointing task, boxes task) often require participants to monitor and remember their responses to the stimuli they have been exposed to before.	Proportion of correctly recalled stimuli
N-back	Stimuli (e.g., letters) are presented sequentially. Participants compare each stimulus to the one *N* positions back (e.g., 2 stimuli back in a 2-back) and to press a key if it was the same (e.g., in a 2-back sequence Q, P, M, P, the P would be a match).	Difference between correct responses to target stimuli (hits) and wrong responses to non-target stimuli (false alarms), *d*', or overall proportion of correct responses
Proactive interference	Participants need to remember new information and overcome interference from previously learned information. For example, in paired-associates recall, participants first learn a list of word pairs. After a memory test, participants are then asked to learn a second list where half the words are paired with the same words as before, and the other half with new words. In some variants, participants are asked to recall the most recent word pairings with the help of cues (one of the words of each word pair). In the recent-negatives paradigm, participants decide whether a probe matches an item from the most recent list (positive probe) or not (negative probe). Some negative probes match an item in the previous list but not the current list; these recent negative probes are harder to reject than new probes.	Performance difference between the two lists, between trials with only one list (no proactive interference) and trials with two lists (proactive interference), or between new negative probes and recent negative probes
Thought probes	At pseudo-random points during a task (e.g., a flanker task, or a reading task), participants are asked whether they are thinking about task-relevant or task-irrelevant topics.	Proportion of probes in which participants reported task-unrelated thoughts
Updating	Participants are asked to transform and substitute memoranda (e.g., numbers) based on sequentially presented new information (e.g.,	Proportion of correctly recalled stimuli

	arithmetic operations) and to remember that new information (e.g., the result of the arithmetic operation). After an unpredictable number of updating steps, participants are asked to recall the most recent information.	
Vigilance	Participants are shown a constant stimulus (e.g., a row of zeros) for an unpredictable, prolonged time. Participants are asked to respond as quickly as possible as soon as the stimulus display changes (e.g., zeros start counting up).	RT means or variability (often in the last blocks of trials in the task), or proportion of correct responses
	Distraction caused by habits	
Antisaccade	Participants see a cue flash on one side of the visual field and must make a quick saccade to the opposite side; in some versions, participants are also asked to identify a briefly presented stimulus at the opposite side. Success in this task requires participants to suppress the reflexive saccade to the cue and to perform a saccade in the opposite direction.	Proportion errors in identifying the stimulus
Global-local	Participants see a large, global pattern made up of small, local elements (e.g., a large letter “Y” made up of small “V”s). In the global version of the task, participants are asked to identify the global element while ignoring the local elements. In the local version of the task, participants are asked to identify the local elements while ignoring the global pattern. In an incongruent trial, the local elements are different from the global pattern (e.g., small “V”s making up a “Y”); in a congruent trial, the local elements match the global pattern (e.g., small “V”s making up a “V”); in a neutral trial, a stimulus associated to none of the responses (e.g., small “Z”s making up a “V”).	Performance difference between incongruent and congruent trials, or between incongruent and neutral trials, or performance in the incongruent trials only
Go/no-go	Participants are asked to press a key when a stimulus (e.g., an “O”) appears, except when a pre-specified stimulus (e.g., an “X”) is presented. In this case, they are instructed to withhold their response.	Proportion of errors on no-go trials and/or RT variability on correct go trials
Random sequence generation	Participants are prompted to generate random sequences (e.g., of numbers or letters) at the pace of 1 number per second and overcome the habit of counting in a stereotyped way.	Degree to which the generated sequences deviate from randomness

Simon	Participants are asked to identify the color of stimulus (e.g., red or blue). The stimulus (e.g., a circle or a letter) is presented at either the left or the right location on the screen. This location is either congruent or incongruent with the location (left or right) of the response buttons.	Performance difference between incongruent and congruent trials (Simon congruency or conflict effect), or performance in the incongruent trials only
Stroop	Participants are asked to indicate the print color of color words while ignoring the meaning of the word. In an incongruent trial, the color word is printed in a different color (e.g., the word "red" printed in green); in a congruent trial, the color word is printed in the same color (e.g., the word "green" printed in green); in a neutral trial, a non-color word or a row of letters is printed in color (e.g., "cat" or "XXX" printed in green).	Performance difference between incongruent and congruent trials, or between incongruent and neutral trials (Stroop congruency or conflict effect), or performance in the incongruent trials only
Stop-signal	Participants are asked to perform an ongoing task (e.g., a word categorization) unless the stop-signal (i.e., a tone or a change in color frame) occurs. In this case, they are instructed to withhold their responses. Typically, the time between the presentation of the stimulus and the stop signal is adapted such that participants can only stop their reaction successfully in 50% of the trials.	Duration of the stopping process (stop-signal response time, SSRT)
Task switching	Participants are asked to alternate between two or more tasks (e.g., deciding whether a digit is larger or smaller than five or whether a digit is odd or even). Participants either switch between tasks across consecutive trials (switch trials) or repeat the decision (repetition trials).	Performance difference between switch and repetition trials[1]

Note. RT = reaction time.

Models of Attentional Control

For a precise conceptualization of the attentional-control construct, we should draw on the best elaborated and empirically supported theories that have emerged from

[1] Switching tasks can also be used to assess global switch costs or mixing costs by comparing mixed-tasks blocks to single-task blocks in which participants perform only one task throughout. These measures do not reflect the switching between tasks but have been shown to be correlated with attentional control (Smith et al., 2019).

experimental cognitive psychology and cognitive neuroscience. Here, we focus on formalized computational models because they make precise (i.e., falsifiable) predictions for empirical findings. A review of these models reveals the following mechanisms that contribute to attentional control: (1) a process of activating or maintaining the representations of task goals (1a) and task rules (1b) for top-down control of cognition (Herd et al., 2014; Munakata et al., 2011; Verguts & Notebaert, 2009; Wiecki, Sofer, & Frank, 2013); (2) a mechanism of attentional filtering to minimize the influence of irrelevant, potentially distracting information (Ulrich, Schröter, Leuthold, & Birngruber, 2015; White, Ratcliff, & Starns, 2011); (3) a global inhibitory mechanism that slows down or stops all action indiscriminately (Munakata et al., 2011; Wiecki et al., 2013); and (4) a process of clearing or removing a task representation once it is no longer needed (Herd et al., 2014; Oberauer, Souza, Druey, & Gade, 2013).

We elaborate on these mechanisms because they arguably mean different things in the context of different tasks. *Task representations* (1 and 4) include representations of the overall goal that describes what is to be achieved (e.g., in the Stroop task, "indicate the print colors"), and of the task set, that is, the relevant stimulus and response categories and the intended mapping between them (e.g., in a 4-color manual Stroop task, mappings from each of the four colors in the stimulus set to the four response keys assigned to them). Goal representations are needed for top-down control on other processes to keep them on task; task sets are needed for efficient translation of stimulus information into response activation. *Attentional filtering* (2) can refer to different aspects of attention, such as filtering out distractors in irrelevant spatial locations (as in the flanker task), or filtering out irrelevant feature dimensions (such as the word meaning in the color Stroop task, the stimulus location in the Simon task, or the currently irrelevant stimulus dimension in a task-switching paradigm). In some models, such filtering emerges through biased competition, which is directly controlled by the strength of goal representations and/or the strength with which goal representations influence ongoing processing (Cohen, Dunbar, & McClelland, 1990; Herd et al., 2014; Munakata et al., 2011; Verguts & Notebaert, 2009). *Global inhibition* (3) plays two roles: In tasks inducing interference, such as the flanker, Simon and Stroop tasks, some models assume that the degree of conflict between alternative responses is continuously monitored (Botvinick, Braver, Barch, Carter, & Cohen, 2001), and a high degree of conflict temporarily raises global inhibition to give top-down influence of task representations time to exert their influence (Erb, Moher, Sobel, & Song, 2016). In tasks requiring task interruption (e.g., the go/no-go or stop-signal tasks), detection of a stimulus demanding an interruption of the ongoing task strongly boosts global inhibition to pause all actions (Wiecki et al., 2013; but see Chatham et al., 2012, for evidence speaking against global inhibition as a source of individual differences in performance). These mechanisms play their roles at different times during cognitive control, some serving more proactive (1a, 1b, 4), others more reactive control (2, 3; Braver, 2012).

These mechanisms, however, do not perfectly match on the types of distraction discussed above. Table 2 shows which of the attentional control mechanisms plausibly contribute to the efficiency of control in some of the most popular experimental paradigms. Critically, the paradigms involve different, partially overlapping combinations of attentional control mechanisms, so that individual differences in each of these mechanisms cannot be expected to drive individual differences in attentional control indices from all of these tasks.

The overall goal representation is the only mechanism involved in all tasks, arguably as a necessary but not sufficient condition for efficient control. A strong goal representation is thus necessary for top-down control on other processes. In some cases, this strong goal representation is associated to strong task sets which are necessary for efficient translation of stimulus information into response activation. For example, strong task-set representations are helpful in paradigms in which the currently relevant task set needs to compete with a prepotent but potentially conflicting task set. One example is the antisaccade task, where the task set that maps lateral stimuli onto saccades in the opposite direction has to overcome the habitual task set that maps stimuli onto saccades in the same direction.

However, strong task-set representations are not necessarily helpful. For example, in the flanker task, incongruent flanker information creates conflict by activating the wrong response via the same task set through which the target activates the correct response; strengthening this task set will facilitate both processes to the same extent, leading to no benefit. Similarly, in the stop-signal task, a stronger representation of both relevant task sets (one for the ongoing task, one for stopping) will make both the production of the go response and the stop process faster, but will not make the stop process win the race more often. In task-switching paradigms, stronger task representations speed up response selection within each task but not when switching to a new task, thereby increasing switch costs (Grange & Cross, 2015; Horoufchin, Philipp, & Koch, 2011). Moreover, if no-longer relevant task representations are not efficiently removed, the interference arising from sticking to that past task representation can increase the cost of switching to another task (Herd et al., 2014).

Table 2

Mechanisms Influencing the Efficiency of Attentional Control Measures in Common Paradigms

	Dependent Measure					
Mechanism	Flanker effect	Anti-saccade accuracy	Simon effect	Stroop effect	Stop-Signal RT	Task switch cost
(1a) Goal representations	↑	↑	↑	↑	↑	↑
(1b) Task representations		↑	↑	↑		
(2) Attentional filtering	↑		↑	↑		↑
(3) Global inhibition					↑	
(4) Removal						↑

Note. An upward arrow (↑) indicates that the mechanism leads to better performance on indicators of attentional control in a task; a downward arrow (↓) indicates that it leads to worse performance. RT = reaction time.

Factor Structure of the Attentional Control Construct

Defining attentional control enables us to formulate testable hypotheses about its psychometric structure. Our current definition of attentional control (i.e., maintaining an operative goal and goal-relevant information in the face of distraction) results in at least four possible structures (see Figure 1). First, attentional control could represent a unitary construct (see Hommel & Wiers, 2017). In this case, the common ability underlying attentional control is the consistent activation and maintenance of goal/task-relevant information, irrespective of the source of distraction. Second, the structure of the attentional control construct may be similar to the structure of intelligence (see the Cattell-Horn-Carroll model; Carroll, 1993), with several first-order constructs that are all positively correlated. This entails that their shared variance could also be represented by a higher-order construct. The higher-order construct would represent the ability to activate, maintain and implement goals and tasks (see Miyake & Friedman, 2012); several taxonomies are possible for the lower-order constructs. A first way to conceptualize the lower-order constructs is to assume that they represent specific abilities to avoid the different types of distraction (i.e., from the perceived environment, self-generated information, and habits). Alternatively, the lower-order constructs could represent the mechanisms identified in the computational models (i.e., task representation, attentional filtering, global inhibition, and removal of task representations).

A third possible structure of attentional control is a collection of multiple approximately orthogonal (i.e., uncorrelated) or separable factors. Similar to the Big Five personality traits proposed in personality research (Goldberg, 1993), this would result in sets of indicators that cluster together (through their mutual correlation) to form factors. Here, the orthogonal constructs could be the same as the first-order constructs proposed above, with the only difference being that there is no shared variance between them to justify a general, higher-order construct. It is also possible that some factors are correlated but not others, resulting in a structure in-between hierarchical and orthogonal models.

Finally, it is also possible that no psychometric construct of attentional control can be established. One reason for this could be that if attentional control were based on learning processes such as proposed in some computational models (Verguts & Notebaert, 2009), the learning process could lead to extremely task- and person-specific control structures, resulting in no correlations between the indicators of the different tasks and thus in no coherent psychometric construct of attentional control. Another reason could be that idiosyncratic strategies or task-specific variation are simply so large that they dwarf smaller variation due to differences in attentional control. Finally, it is also possible that attentional control cannot be established due to either a lack of true individual variation in attentional control or because the measures used are too psychometrically poor to identify a factor.

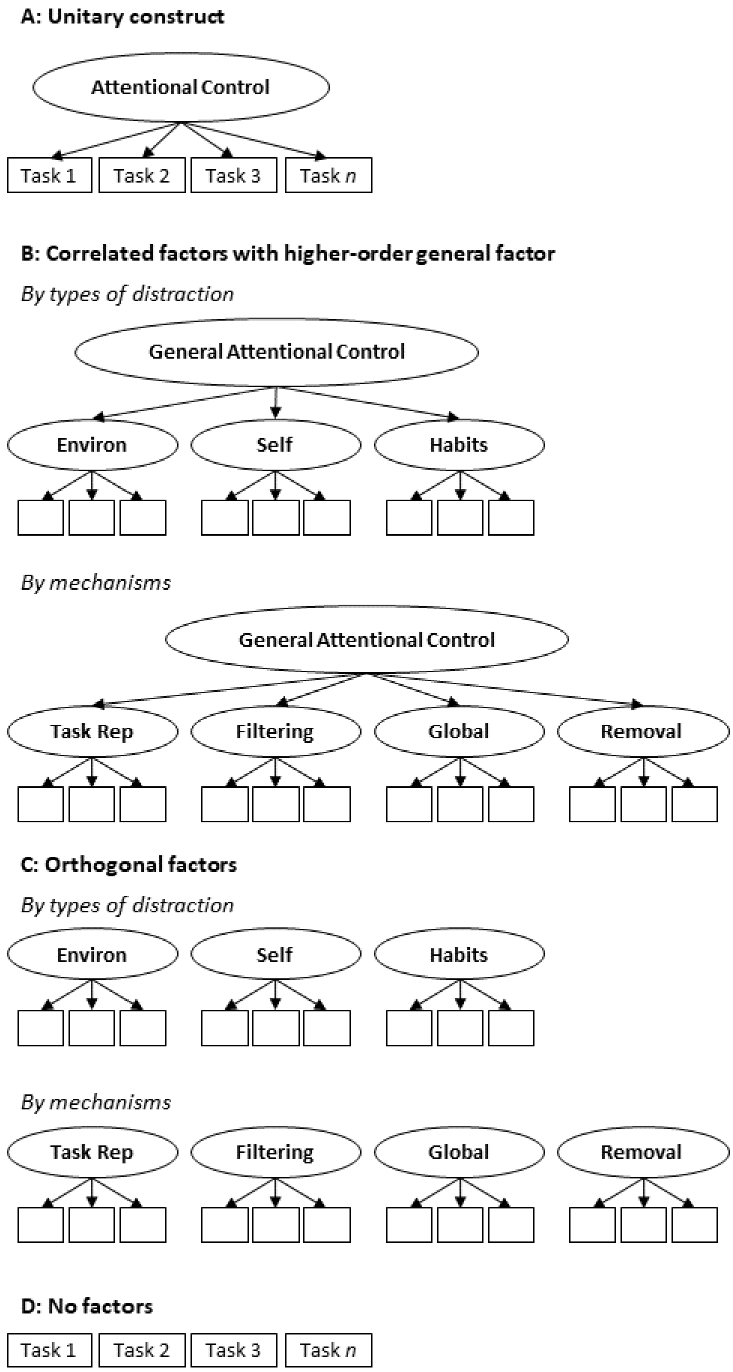

Figure 1. Four possible factor structures of attentional control measures illustrated with latent-variable models. Rectangles denote manifest (observed) variables (e.g., antisaccade accuracy or Stroop effect), ellipses represent latent factors based on the shared variance between the manifest variables. Single-headed arrows denote linear regressions. A) Unitary factor structure in which all measures load on one attentional control latent factor. B) Hierarchical factor structure with a higher-order attentional control construct representing the shared variance of several lower-order specific attentional control factors. The lower-order factors could represent the types of distraction faced in attentional control tasks (from the perceived environment, self-generated information, and habits) or the mechanisms of attentional control (task representations, attentional filtering, global inhibition, and removal of task representations). C) Orthogonal or separable factors where measures cluster by the source of distraction or mechanism, but the types of distraction or mechanisms are (largely) unrelated

to each other. D) Measures do not correlate and, therefore, no latent factor(s) can be established. Environ = distraction from the perceived environment; Self = distraction from self-generated information; Habits = distraction from habits; Task Rep = task representations; Filtering = attentional filtering; Global = global inhibition.

Validity of Attentional Control

A critical aspect of defining attentional control and its psychometric structure is to determine its convergent and divergent validity. The convergent validity of attentional control would be convincing if there were a large set of empirical indicators stemming from heterogeneous sources (e.g., from different experimental tasks, cross-sectional and longitudinal observations, or behavioral-genetic and neuroscientific evidence) that are theory-driven and interchangeable. Importantly, the different indicators should correlate in agreement with theoretical predictions. Additionally, divergent validity from other cognitive-ability constructs, such as associative learning, semantic memory, episodic memory, processing speed, and intelligence should be demonstrated. Attentional control may substantially correlate with one or several of these constructs. However, at the same time, attentional control should retain some variance which cannot be explained by the other constructs.

Once the convergent and divergent validity of attentional control are established, we can test the construct's criterion validity by correlating the unique variance related to attentional control with outcome variables that we theoretically assume to be influenced by attentional control. For example, such real-world outcomes could include academic performance, behavioral problems such as substance abuse or delinquency, or results from questionnaires in which self-control or impulsivity is assessed by an observer (e.g., a teacher in the case of assessing a child's self-control).

A further contribution to construct validity is to confirm that empirical indicators of attentional control reflect the parameters governing attentional control in computational models. Attentional control indicators should correlate with the parameters that are assumed to reflect mechanisms relevant to attentional control. For instance, finding that parameters governing the strength of a goal representation correlate with the congruency effect in a simulated Stroop task (Herd et al., 2014) increases our confidence that the Stroop congruency effect taps an attentional control construct. At the same time, computational modeling can also reveal that processes other than attentional control substantially contribute to the variance observed in classic attentional control paradigms (flanker, Simon, and Stroop task; Hedge, Powell, Bompas, & Sumner, 2020).

Empirical Findings on Structure and Validity of the Attentional Control Construct

Over the past 20 years, the psychometric structure of attentional control has been extensively studied using latent-factor modeling (Friedman & Miyake, 2017; Karr et al., 2018; Miyake et al., 2000). In these studies, multiple experimental tasks assumed to tap into the same construct are administered, and their common variance is extracted as a latent factor (see Figure 1). Latent factors represent the shared variance between several indicators and thereby separate the true common variance from the indicator-specific and error variance. Substantial loadings of the indicators on the latent factor considerably reduce the measurement error and serve as evidence for the coherence (i.e., convergent validity) of the cognitive construct in question. Furthermore, latent factors can also account for the interdependencies (i.e., covariances) between the extracted factors. Thus, this type of modeling is a useful tool in establishing the psychometric structure and validity of theoretical constructs.

The seminal work by Miyake et al. (2000) has gained particular prominence as reflected by its exceptional number of citations (12,758 times according to Google Scholar on July 17, 2020). Based on the covariance among classic measures of attentional control, Miyake et al. (2000) proposed a model with multiple correlated first-order constructs and provided first evidence for convergent validity of the attentional control construct. They identified three distinct yet correlated latent factors of attentional control: the inhibition of a prepotent response, shifting between task sets, and updating of contents in working memory. In more recent incarnations of this influential model (Friedman & Miyake, 2017; Miyake & Friedman, 2012; first presented in Friedman et al., 2008), a common factor that accounts for covariance across all attentional control measures (i.e., inhibition, shifting, and updating tasks) has emerged, along with two orthogonal factors that capture the covariance specific to shifting and updating. This nested-factor (bifactor) structure suggests that there are both domain-general and domain-specific attentional control mechanisms. Notably, in this model, the covariance across inhibition tasks is fully accounted for by the common factor. This common factor has been proposed to reflect active goal maintenance and the top-down bias of attention (Friedman & Miyake, 2017), a notion that is consistent with the basic definition of attentional control we proposed above. Below, we briefly review the evidence from cross-sectional, longitudinal, genetic-behavioral, and neuroscientific studies concerning the factor structure and validity of the attentional control construct.

Cross-Sectional Evidence

Following Miyake et al.'s (2000) seminal paper, inhibition, shifting, and updating have become the most frequently assessed attentional control constructs in the literature (Karr et al., 2018), but the evidence for their construct validity is mixed. Attempts to establish the three-factor and nested-factor structure have often been successful (Fleming, Heintzelman, & Bartholow, 2016; Friedman et al., 2008; Gustavson, Panizzon, Franz, et al., 2018; Ito et al., 2015; Schnitzspahn, Stahl, Zeintl, Kaller, & Kliegel, 2013; Vaughan & Giovanello, 2010), and a recent systematic review and re-analysis of existing data sets by Karr et al. (2018) revealed overall modest support for the nested-factor model across adult samples. with the evidence being particularly strong for studies with large sample sizes and more rigorous procedures. However, the review also highlighted that only few studies even considered alternative models (i.e., models with different lower-order constructs and/or structures, e.g., Adrover-Roig, Sesé, Barceló, & Palmer, 2012; Chuderski, Taraday, Nęcka, & Smoleń, 2012; Fournier-Vicente, Larigauderie, & Gaonac'h, 2008). Moreover, Karr et al. reported low rates of model acceptance (i.e., the rate at which a model showed satisfactory fit to the data) and model selection (i.e., the rate at which a selected model fitted the data better than alternative models). Karr et al. attributed these issues to relatively small sample sizes, high model complexity, and poor reliability of the experimental measures used in the sampled studies.

Moreover, recent studies have questioned the convergent validity of the attentional control construct, in particular the inhibition construct (e.g., Rey-Mermet et al., 2018, 2019). Whereas factor loadings and/or between-task correlations are typically acceptable for measures of updating (Ecker, Lewandowsky, Oberauer, & Chee, 2010; Singh, Gignac, Brydges, & Ecker, 2018) and shifting (e.g., von Bastian & Druey, 2017), they are often weak for inhibition measures (De Simoni & von Bastian, 2018; Friedman & Miyake, 2004; Guye & von Bastian, 2017; Hedge, Powell, Bompas, Vivian-Griffiths, & Sumner, 2018; Hull, Martin, Beier, Lane, & Hamilton, 2008; Keye et al., 2009; Oschwald, Schättin, von Bastian, & Souza, 2018; Paap & Greenberg, 2013; Rey-Mermet et al., 2018, 2019; von Bastian & Oberauer, 2013; von Bastian et al., 2016; Whitehead et al., 2019; Wilhelm, Hildebrandt, & Oberauer, 2013). In some of these studies, inhibition measures did not share enough variance to establish a latent factor. In other cases, latent factors were dominated by a single task (typically the antisaccade task; for a detailed discussion, see Rey-Mermet et al., 2019; see also Magnusdottir et al., 2019), thus reflecting only a small fraction of the systematic variance across the measures. Taken together, evidence available from cross-sectional studies overall supports a factor structure with multiple, correlated first-order constructs, with the critical limitation that other structures – except nested solutions with fewer factors – have rarely been tested. However, cross-sectional evidence is mixed for the construct validity of attentional control, and in particular inhibition, as reflected by weak between-task correlations.

Longitudinal Evidence

Evidence from existing longitudinal studies generally support the three-factor and the nested-factor model. In these studies, the factors reflecting inhibition, shifting and updating, as well as the common factor were highly correlated across 6-7 year spans in adolescence and middle age (range in rs = .87 to .99; Friedman et al., 2016; Gustavson, Panizzon, Elman, et al., 2018), and across short test-retest intervals in childhood or late middle age (range in *r*s = .98 to 1.0; Ettenhofer, Hambrick, & Abeles, 2006; Willoughby, Kuhn, Blair, Samek, & List, 2017). Correlations among measures attributed to inhibition at any one point in time were low (range *r*s = .15 to .32). However, across 6-7 year intervals, inhibition measures at Wave 1 consistently correlated with different inhibition tasks at Wave 2 to roughly the same extent as they correlated with each other at one point in time (range *r*s = .13 to .27), but not as highly as each measure correlated with itself across 5-6 -year intervals (range *r*s = .21 to .58). Therefore, although the covariance among inhibition measures is small overall, it is reliable and stable within individuals. Similar patterns, but with stronger correlations, are observed for shifting and updating measures.

Behavioral-Genetic Evidence

Results from twin studies suggest that the variance shared between these measures is primarily genetic in origin. Heritability estimates for inhibition, shifting, updating, and the common factor are high from childhood through middle age (Engelhardt, Briley, Mann, Harden, & Tucker-Drob, 2015; Friedman et al., 2016; Gustavson, Panizzon, Elman, et al., 2018). Heritability estimates are smaller but still considerable for individual measures from each domain (T. Lee et al., 2012; Schachar, Forget-Dubois, Dionne, Boivin, & Robaey, 2010). In other words, the small amount of variance that is shared across attentional control measures within an individual also seems to be shared across different individuals to the extent that those individuals share genetic variation.

Neuroscientific Evidence

Most studies investigating neural correlates of attentional control focus on within-subject changes by contrasting brain activity in low- and high-demand conditions; studies examining individual differences, and in particular such using a latent-factor approach, are rare. However, mapping brain regions to task performance can still inform about commonalities and differences in terms of the neural circuitries involved when performing various attentional control tasks. In their meta-analysis, Niendam et al. (2012) found evidence for a domain-general, superordinate control network involving the dorsolateral prefrontal cortex, the anterior cingulate cortex, and the parietal cortex that is activated across most attentional control tasks (e.g., Stroop, switching, and N-back tasks). This superordinate control network can be speculated to form the neural substrate of a general attentional control factor (e.g., the common factor proposed in the nested-factor model by Miyake & Friedman, 2012; see also Figure 1B).

In addition to this domain-general neural network, domain-specific networks may exist as well (e.g., Nigbur, Ivanova, & Stürmer, 2011). In a recent review of the dynamics of attentional control, Gratton, Cooper, Fabiani, Carter, and Karayanidis (2018) attempted to link Miyake et al.’s (2000) three-factor model to distinct brain networks, with mixed results. Whereas there is some evidence to suggest that shifting performance is associated with activity in the dorsal frontoparietal network, inhibition and updating do not seem to correspond consistently to separable networks. However, it should be emphasized that there may not be a one-to-one mapping between psychological constructs and brain networks as any one cognitive process might emerge from a complex interplay between different networks (Gratton et al., 2018). Future large-scale neuroscientific studies with task batteries assessing multiple aspects of attentional control are needed to investigate how the interactions of various brain networks support the theoretical constructs related to attentional control.

Interim Summary

Taken together, the present evidence suggests that classic measures of attentional control reflect a good deal of task-specific variance, but do share some common variance that is heritable, temporally stable, and related to unique neural markers. Overall, however, the covariances between measures loading on the shifting and updating constructs seem to be stronger than covariance among inhibition measures, resulting in a fickle inhibition factor that does not emerge consistently across studies. On the one hand, unsuccessful attempts to establish a coherent latent factor using the measures related to inhibition lend credence to the view that there is no such construct as domain-general inhibition (Egner, 2008; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). On the other hand, studies that do report a latent factor for inhibition suggest that the small variance shared among inhibition measures is important, as it can be related to real-world behaviors, such as attention problems and a liability to externalizing psychopathology behaviors (Friedman et al., 2007; Young et al., 2009; but see Eisenberg et al., 2019). Nevertheless, the most cautious conclusion we can draw from the evidence thus far is that the convergent validity of the inhibition construct is yet to be established. Notably, studies rarely tested the convergent validity considering the underlying mechanisms of attentional control (for an exception, see Friedman & Miyake, 2004).

How Do We Assess Attentional Control? A Survey of Individual Differences Studies

The inconsistent conclusions from previous studies could result from the large variation in methodologies they employed. The most obvious difference across studies is the diversity of paradigms used to measure attentional control (see Table 1). This heterogeneity is further compounded with variations in the specific stimuli used, the timing parameters, the amount of conflict created (e.g., how often conflicting information is presented), the number of trials administered, and how attentional control scores are derived.

Given the state of the literature, two questions seem pertinent: To what extent do individual differences studies converge in the paradigms used to assess attentional control? And do these tasks do a good job in terms of reliable and valid assessment of individual differences? To gain more insight into these questions, we surveyed studies assessing individual differences in attentional control to identify the most popular attentional control tasks and gauge the task scoring methods, reliability, correlations with other tasks, and loadings on a latent factor. We selected studies that (1) involved a sample of cognitively healthy participants, (2) reported zero-order correlations and/or factor loadings in a latent-variable measurement model of at least two independent, performance-based measures of attentional control that involve distraction caused by the perceived environment, by self-generated information, or habits, (3) were published in a peer-reviewed journal or academic book, and (4) were written in English language. We screened the 40 studies included in Karr et al. (2018) and 50 additional studies that our team of authors were aware of. Upon closer inspection, we excluded 6 studies that reported neither zero-order correlations nor factor loadings of independent measures of attentional control (Bettcher et al., 2016; Frazier et al., 2015; Huizinga, Dolan, & van der Molen, 2006; Keye, Wilhelm, Oberauer, & Stürmer, 2013; Keye et al., 2009; Schnitzspahn et al., 2013), 5 studies that reported data from a sample that overlapped with the sample in another study already included in the review (Friedman et al., 2016, 2006; Friedman, Miyake, Robinson, & Hewitt, 2011; Gustavson et al., 2019; Smith, Banich, & Friedman, 2019), 2 studies that reported data from only one eligible measure (C. Hughes, Ensor, Wilson, & Graham, 2009; Usai, Viterbori, Traverso, & De Franchis, 2014), and 1 study that reported a reanalysis of data reported in studies already included in the review (Unsworth, 2015). The final set of 76 included studies yielded 85 samples of participants: 29 samples of children or adolescents, 5 of older adults, 3 of younger and older adults, and the remaining reported data were samples of younger adults (see Appendix for a list of included studies). In total, the included samples provided a pool of 577 measures. Reliabilities were reported for 406 of the measures assessed in 59 samples (51 studies). We could extract 2114 zero-order task correlations from 77 samples (70 studies), and 470 factor loadings from 71 samples (63 studies). Measures were categorized by the type of distraction involved (environment, self-generated, or habits) and the attentional control latent factor they are typically thought to reflect (inhibition, shifting, updating, or mind wandering). We also extracted the sample size and number of trials, and we coded the type of paradigm and performance score. Coded data and analysis scripts are available on the Open Science Framework (https://osf.io/kw2a9/).

Figure 2 shows the frequency of usage of different paradigms and the type of performance scores extracted. On average, 6.8 of the extracted measures were administered per sample. Although studies varied in terms of the exact stimuli used, most administered tasks were variations of the paradigms summarized in Table 1. The more popular tasks in Figure 2 match those Miyake and colleagues (2000) used in their seminal paper to assess inhibition (Stroop, antisaccade, and stop-signal), shifting (task switching), and updating (keep track), but studies surveyed varied in the combination of tasks used. Distraction by the environment was most commonly assessed with flanker tasks, with only few studies using matching or search paradigms. Distraction by self-generated information was most frequently measured with keep-track, n-back and other updating paradigms. Thought probes were used somewhat less often, probably because the assessment of mind wandering has gained momentum only in the last ten years or so (for reviews, see Mooneyham & Schooler, 2013; Seli, Risko, Smilek, & Schacter, 2016; Weinstein, 2018). Distraction by habits was assessed most frequently overall and predominantly with paradigms reflecting the inhibition factor, including variations of Stroop, go/no-go, antisaccade, stop-signal, and Simon paradigms. Task switching paradigms made up about a third of distraction by habits measures.

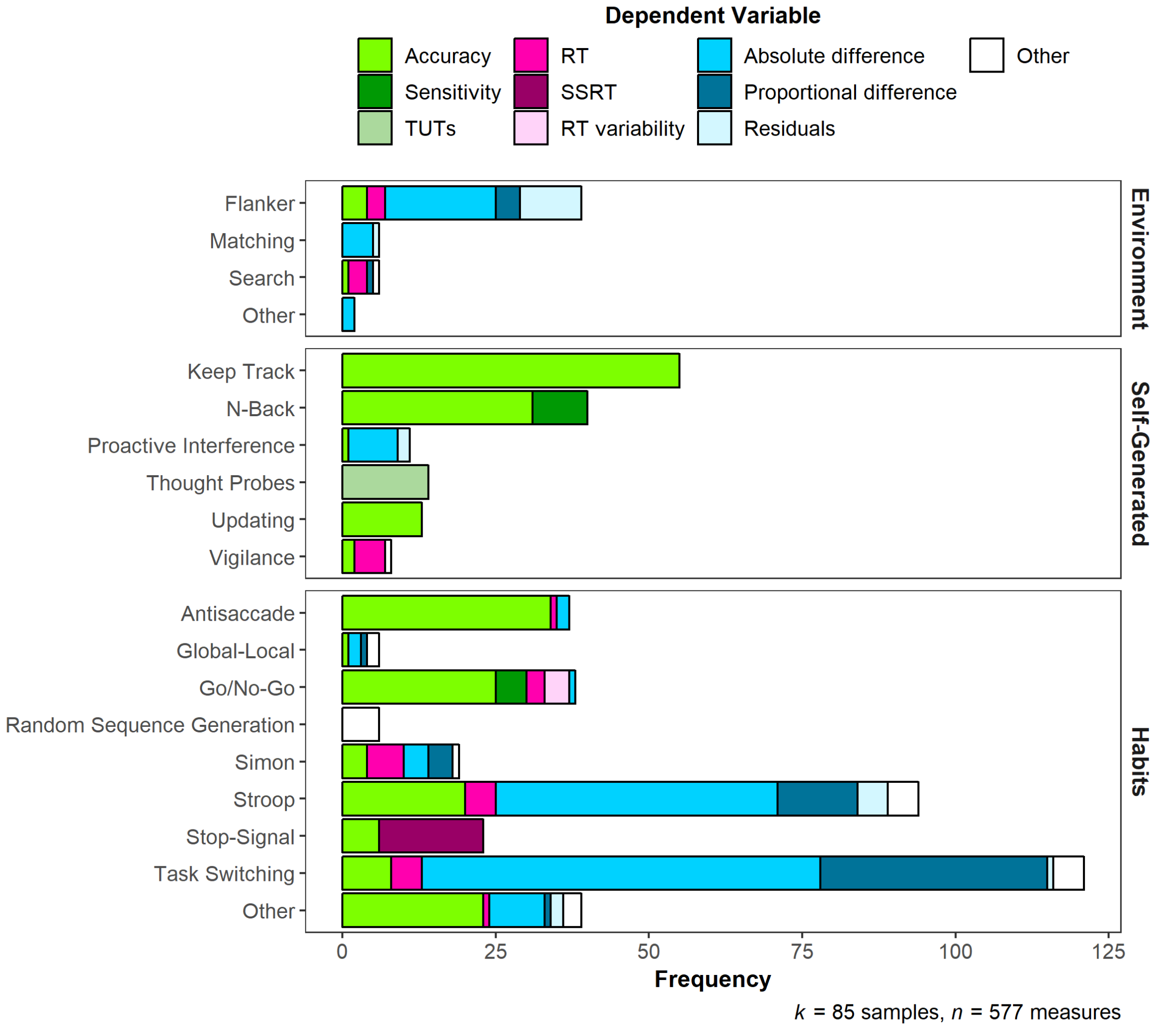

Figure 2. Frequency of usage of each type of paradigm in the survey and of the different types of performance scores extracted from them. TUTs = task-unrelated thoughts; RT = reaction time; SSRT = stop-signal reaction time.

Some researchers also employed tasks that are not closely related to the paradigms presented in Table 1 to measure attentional control (categorized as “Other” in Figure 2), including neuropsychological tests such as the Trail-Making Test or the Wisconsin Card-Sorting Test. This is potentially problematic, as neuropsychological tests may capture much more than attentional control because these tasks tend to be more complex. Moreover, they often involve only a single trial designed to diagnose clinical impairments, not to measure individual differences in non-clinical populations.

As illustrated in Figure 2, the heterogeneity in assessing attentional control is also reflected in the type of scores researchers extract from the tasks. For most paradigms, researchers can select from multiple performance indicators, resulting in often varying indicators for the same paradigm. First, researchers can choose between scores based on accuracy or reaction times (RTs). Second, researchers can compute these scores based on different sets of trials. One option is to compute scores using only the experimental condition demanding high levels of attentional control, or to average performance across all conditions implemented in the experiment; here, we will refer to these types of measures as single scores. Our literature survey showed that researchers relied on single scores in 58% of the cases (primarily on accuracy, 73%). Updating and mind wandering were assessed exclusively by single scores (primarily accuracy and proportion of task-unrelated thoughts, respectively). However, when assessing inhibition and shifting, researchers often want to isolate the contribution of attentional control by contrasting conditions demanding high and low levels of inhibition or shifting. Here, we will refer to these types of measures as condition-difference scores. In the literature surveyed, researchers opted for computing condition-difference scores for 72% of the shifting measures and 46% of the inhibition measures (primarily based on RTs, 84%). In most cases, researchers computed the absolute difference in performance (66%), followed by proportional differences (25%). Residuals of explaining variance on the high demanding condition based on the low demanding condition were reported for only 9% of the measures. The exact reasoning for selecting one method over another is often not clear from the publication, or sometimes researchers report that the selection was based on the size of the correlation with other measures. Either way, these choices led to inconsistent methods in assessing attentional control.

Are Attentional Control Paradigms Reliable?

Although reporting of reliability should be standard practice when studying individual differences (Parsons, Kruijt, & Fox, 2019), reliability coefficients could be extracted for only 70% of the measures. Most of the 406 reported reliability coefficients were for internal consistency (e.g., Cronbach's α, split-half reliability); test-retest reliability was reported in only 3% of the cases. The median reliability across all measures was .81, with a range, however, from .01 to .98. Figure 3 presents the reliabilities of the measures reported in the individual studies as a function of the dependent variable used. Median reliability was highest for mind wandering measures (.90; M = .89, SD = .08), followed by updating (.83; M = .78, SD = .15) and shifting measures (.81; M = .77, SD = .16). Although median reliability was overall only slightly lower for inhibition measures (.79; M = .72, SD = .22), the spread of reliability coefficients was larger for indicators of this factor. Median reliability was particularly low for inhibition condition-difference scores (.66; M = .63, SD = .24), which was markedly lower than for shifting condition-difference scores (.82; M = .78, SD = .14). Low reliability is problematic, because a task score that is not reliable does not consistently rank-order individuals and, thus, cannot be a good predictor of other abilities.

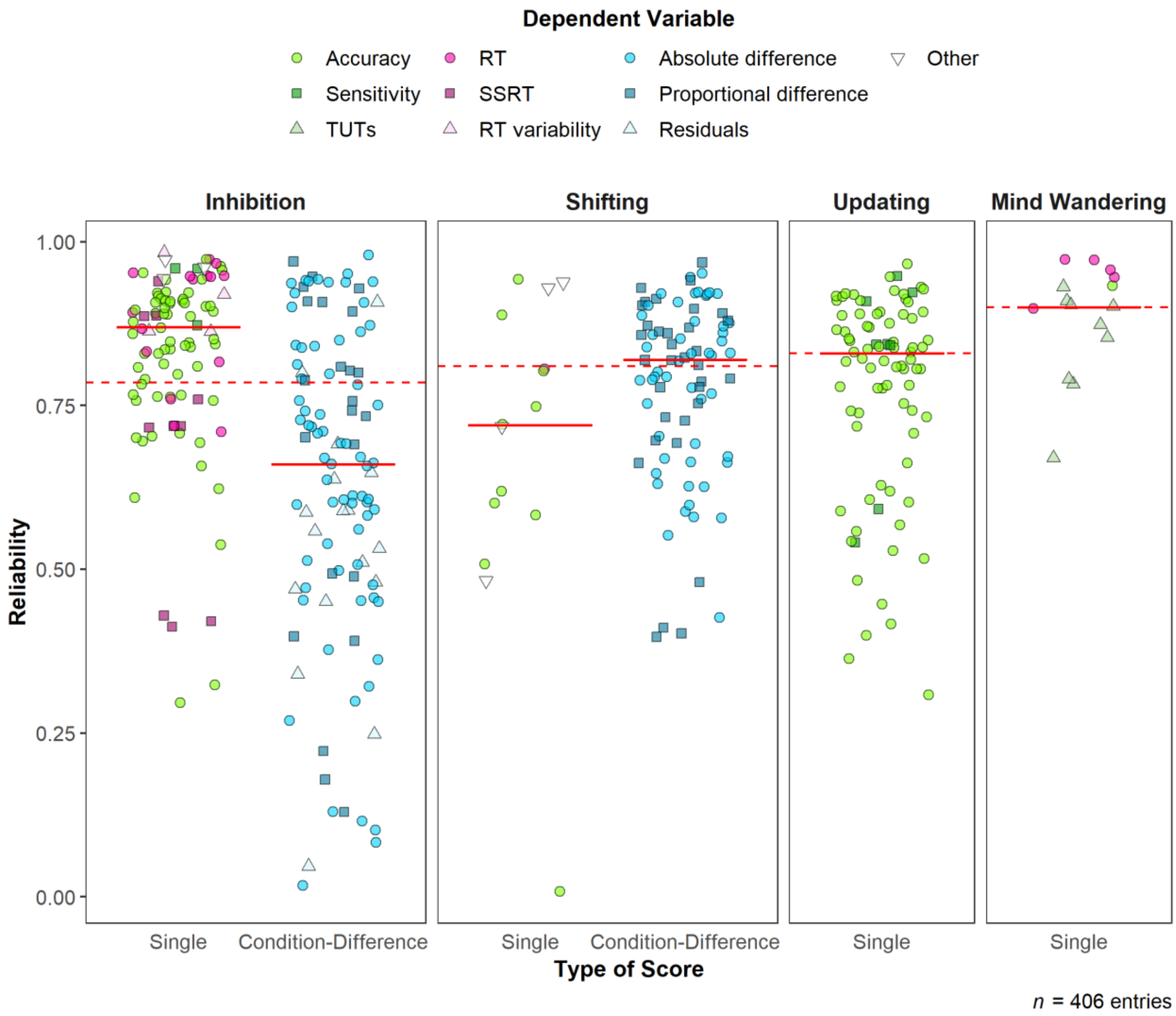

Figure 3. Reliabilities of measures as a function of scoring method and attentional control factor assessed. Horizontal lines represent median reliability, dashed lines represent the median across types of scores.

Low reliability of inhibition condition-difference scores has been attributed to these indicators relying on measuring an experimental contrast (i.e., the condition demanding high levels attentional control in contrast to the condition demanding low levels of it) and thereby subtracting out a large part of the performance variance between individuals (Enkavi et al., 2019; Hedge, Powell, & Sumner, 2018). If the true variance in the experimental effect is small, it is easily swamped by error variance, leading to low reliability. Hence, it is possible that reliabilities of condition-difference scores are higher for shifting than for inhibition paradigms because the latter may yield smaller effects. We explored this possibility by computing the ratio between means and standard deviations of the condition-difference RT effects. Indeed, on average, shifting effect sizes were larger (absolute differences: 1.81, proportional differences: 1.53) than inhibition effect sizes (absolute differences: 1.61, proportional differences: 1.21), and effect sizes were positively correlated with reliability, r = .55 (n = 162; inhibition: r = .51, n = 80; shifting: r = .52, n = 82).[2]

Experimental effects and, thus, reliability can be increased by increasing the true variance of the experimental effect, or by reducing the measurement noise – the latter can be achieved by increasing the number of trials. Figure 4 presents reliabilities for the dependent measures of the three most frequently used inhibition paradigms as a function of the number of trials. In the studies surveyed, reliabilities were positively correlated with the number of trials for the single scores used in antisaccade paradigms, r = .21 (n = 29), but appeared to be unrelated to the number of trials in Stroop paradigms, r = .01 (n = 54). Upon closer inspection, we found that the correlation hovered around zero for condition-difference scores, r = .04 (n = 46), and that reliabilities tended to be lower for larger numbers of trials for single scores, r = -.11 (n = 8). This negative relationship was more pronounced for the single scores used in, stop-signal paradigms, r = -.74 (n = 11).

[2] Note that these and all following correlation coefficients do not account for clustering in the data arising from including multiple measures assessed in the same sample. However, a hierarchical regression model that did include a random effect of sample on the intercept (and, thereby, did account for clustering of effect sizes within samples) also showed that effect sizes significantly predicted reliability, p < .001, with decisive evidence favoring the inclusion of effect size as predictor (Bayes factor = 1,488,194,608).

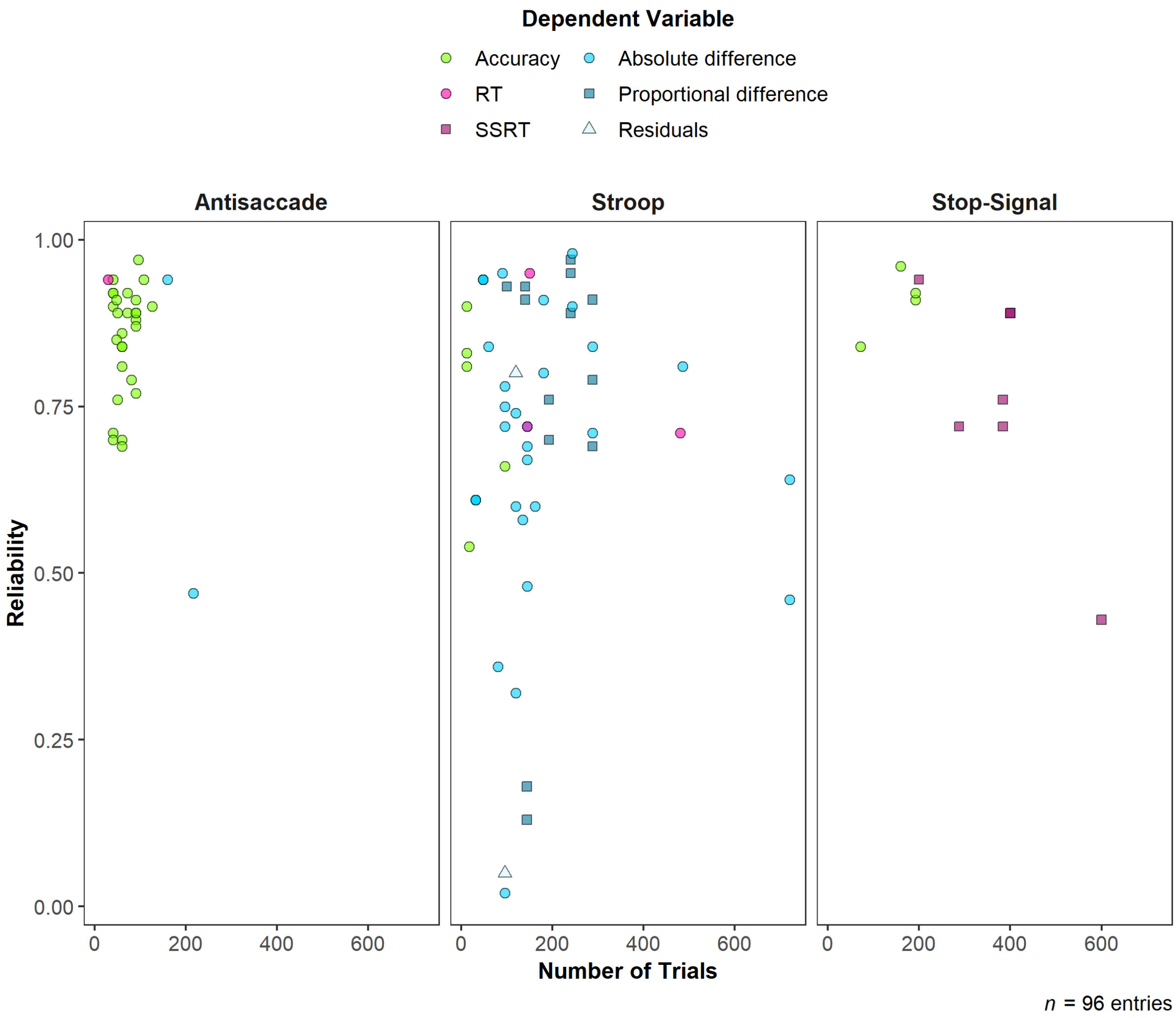

Figure 4. Reliabilities of measures from the three most frequently used inhibition paradigms as a function of number of trials and type of dependent variable.

At first glance, this appears puzzling because we would expect reliabilities to increase, and not decrease, with more trials (Enkavi et al., 2019; Hedge, Powell, & Sumner, 2018). One explanation goes back to a more general issue of difference scores that affects condition-difference scores as well as some single scores that are based on differences such as SSRTs or *d*': the higher the correlation between the components of a difference (such as between congruent and incongruent trials in a Stroop task, or between participants' latencies and the individual stop-signal delay in a stop-signal task), the lower the reliability of the difference score (Lord, 1963). Therefore, although running more trials will improve the reliability of the components and, thereby, also increase the reliability of the difference score, a larger number of trials still cannot overcome the limited reliability of difference scores if the components are highly correlated. Another explanation could be that with increasing length, a paradigm may no longer measure attentional control due to habituation of the source of distraction (e.g., when the originally predominant response is overwritten by the response required in the task context). Hence, for each task context, there will be an ideal number of trials for maximizing reliability of the difference score while avoiding habituation of the distractors. Moreover, considerations regarding the optimal number of trials for each task context will need to be balanced with practical concerns such as length of the overall testing session.

Taken together, moving forward, the literature would benefit from seeking to develop new or to refine existing tasks that focus on increasing between-subject variability while reducing measurement noise with an optimized number of trials. Developing theoretically driven tasks that target the differentiation between individuals and that have low measurement noise, from which consistent task scores are extracted, is likely to increase the robustness of the assessment of individual differences in attentional control.

Do Attentional Control Tasks Measure A Coherent Construct?

Figure 5 presents a scatterplot relating inter-task correlations to the sample size in each study. Studies reported consistently low inter-task correlations (median r = .16) and most studies do not report correlations larger than r = .30. The median sample size was N = 159 (ranging from N = 36 to 1225), which is far below the number that simulations have shown is needed for such low correlations to stabilize (N = 250; Schönbrodt & Perugini, 2013).

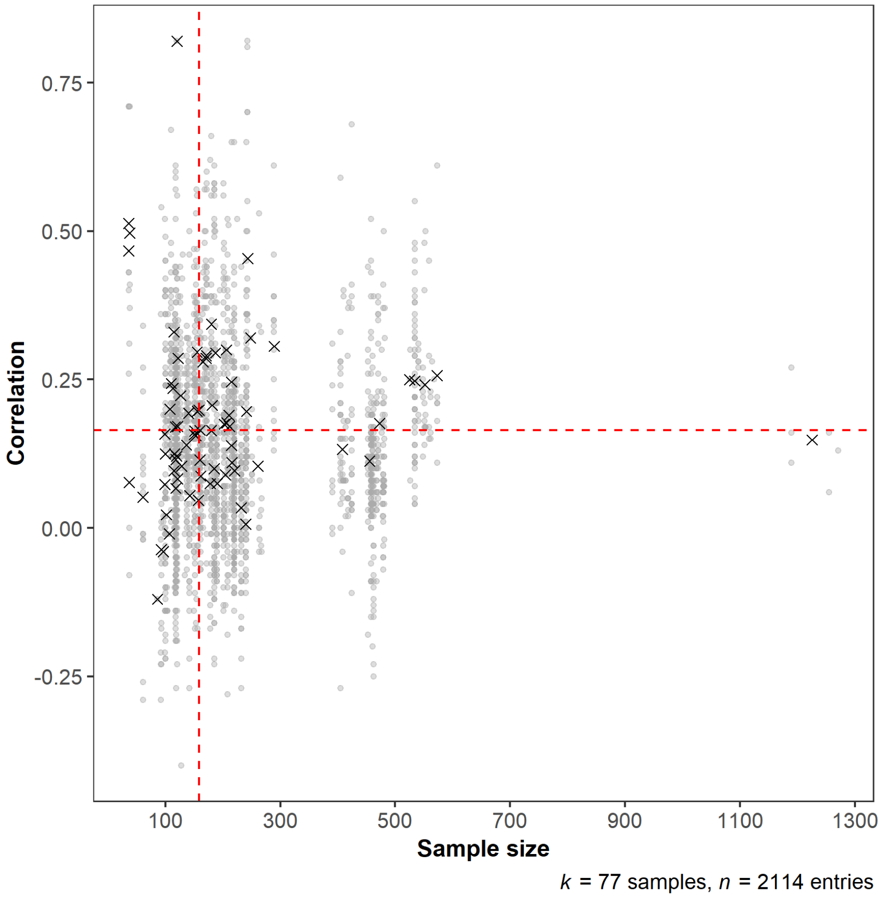

Figure 5. Correlations between all attentional control measures used in each sample (grey dots) and their mean correlation in that sample (× symbols) in relation to sample size. Dashed lines represent median values across samples.

Figure 6 presents correlations split by paradigm and whether that task was correlated with other tasks involving the same source of distraction (e.g., habit-habit) or a different source of distraction (e.g., habit-environment). Paradigms involving distraction from self-generated information showed the highest same-source correlations, but they correlated overall only weakly with paradigms involving other sources of distraction. Considering only same-source correlations, measures of antisaccade and go/no-go paradigms also tended to show somewhat higher correlations. Notably, for many paradigms, correlations with same-source paradigms were within a similar range as different-source correlations.

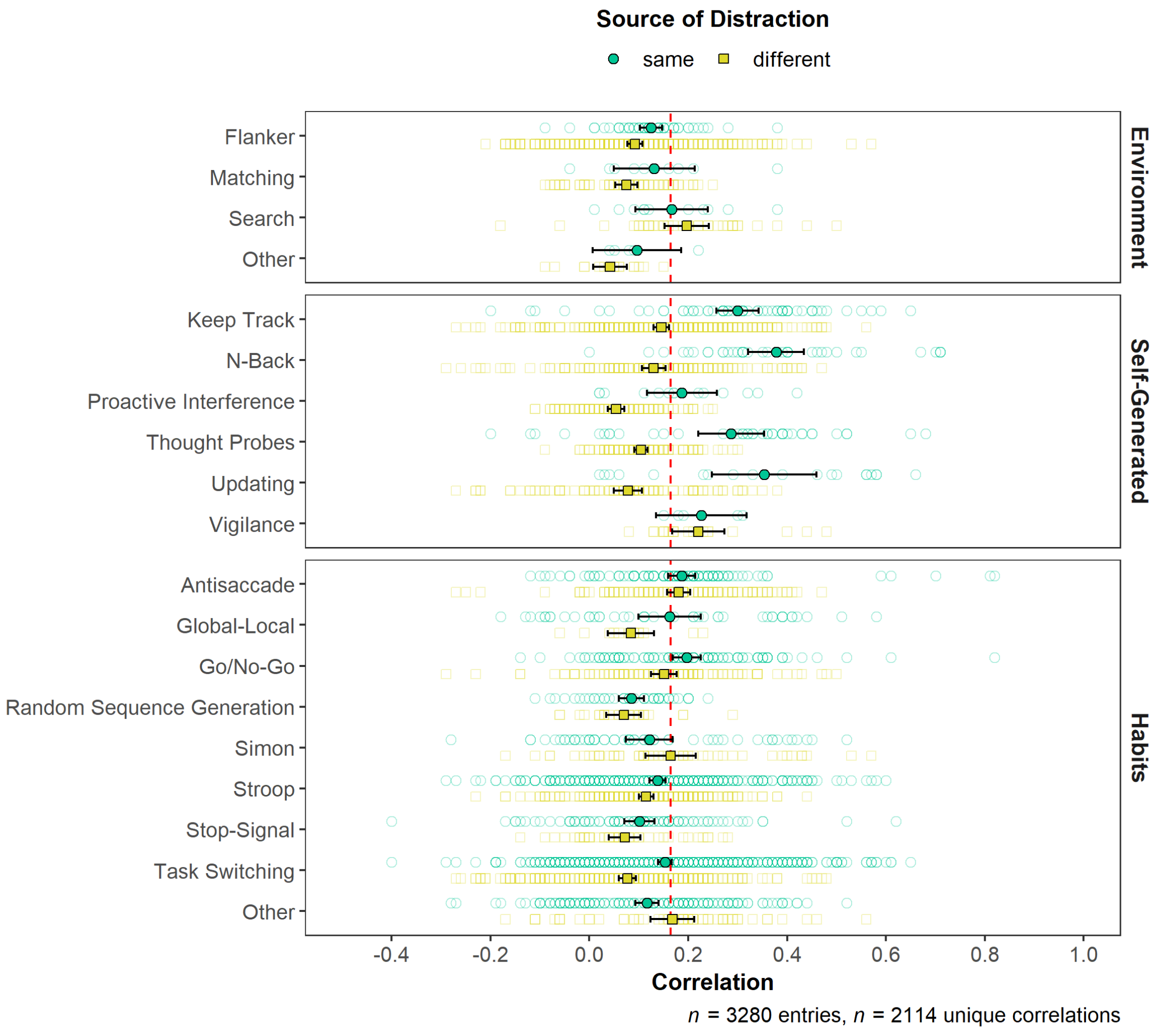

Figure 6. Mean correlations between measures obtained for each paradigm in each sample (unfilled symbols) as a function of whether the second measure involved the same source of distraction (e.g., habits-habits) or a different source of distraction (e.g., habits-self-generated). Filled symbols represent the median correlation across samples, and the error-bars represent the 95% confidence interval. The dashed line represents median values across samples.

Correlations between any two measures determine their loadings on a common latent attentional control factor. Thus, it is not surprising that many of the factor loadings derived from the surveyed attentional control measures are low, as illustrated in Figure 7. Matching the patterns observed for correlations (Figure 6), paradigms involving distraction from self-generated information show generally higher factor loadings, whereas paradigms involving distraction from the environment, and especially flanker paradigms, yield relatively lower factor loadings. From the paradigms involving distraction from habits, antisaccade measures stand out as loading most strongly on the latent factor; indeed, they are often found to dominate the attentional control construct derived. This has led some authors to criticize these models as lacking validity, because the latent variable does not capture as much measure variance from the other measures as it does for the antisaccade (for a discussion, see Rey-Mermet et al., 2019). Go/no-go, Simon, and task switching paradigms tended to yield factor loadings around the median and, thus, might be more suitable for constructing balanced latent factors.

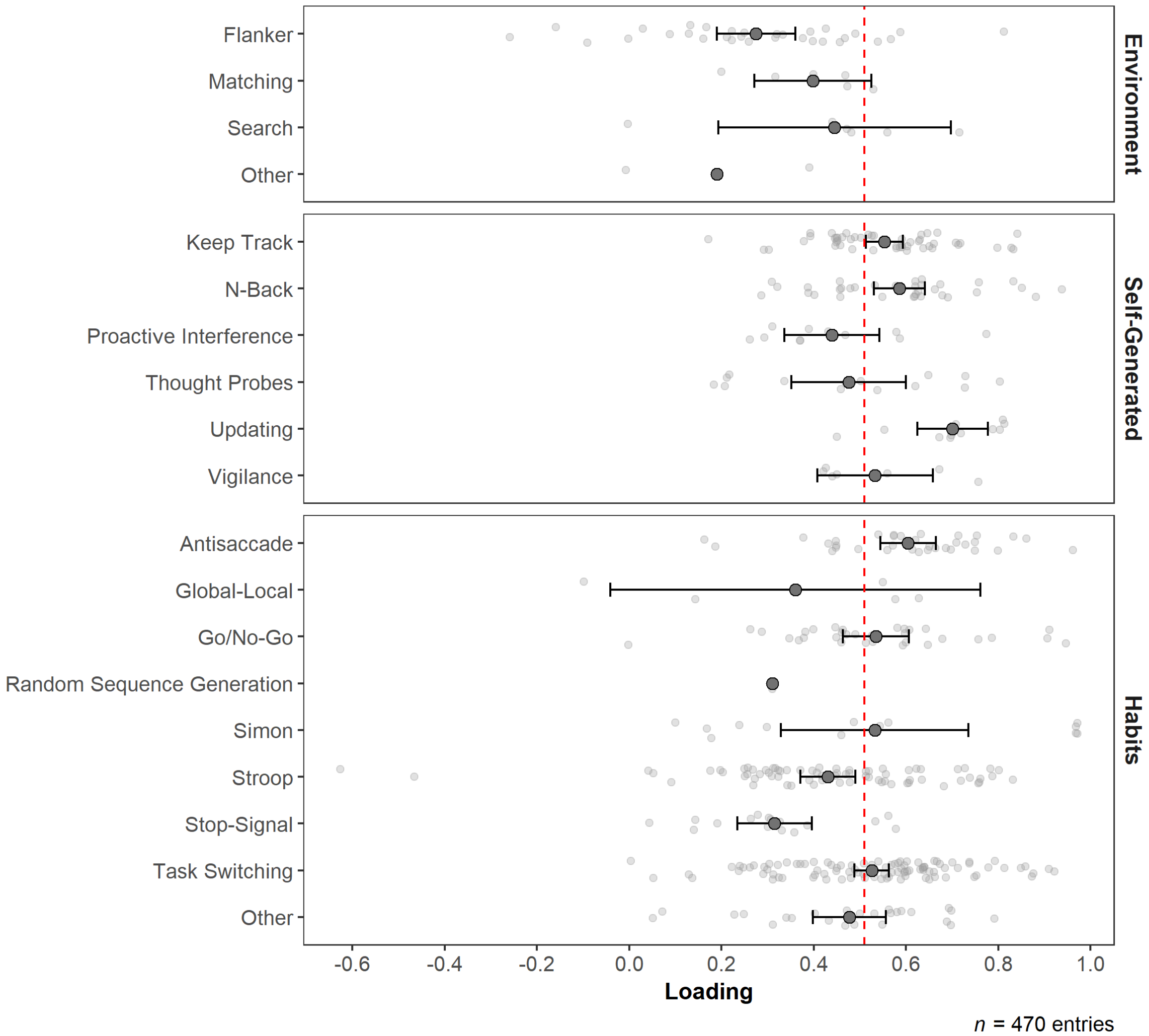

Figure 7. Loadings for each paradigm on a latent attentional control factor. The filled circles represent the mean factor loadings across studies and the error-bars the 95% confidence interval (displayed for paradigms with at least 3 entries). The dashed line represents the median factor loading across all paradigms.

Which Attentional Control Paradigms Work “Best”?

The findings of our survey demonstrate that the currently most widely used attentional control paradigms yielded (1) variable reliabilities, (2) low between-measures correlations, and, in some cases, (3) low loadings on a latent attentional control factor. Low reliabilities pose a serious problem to the measurement of individual differences in attentional control. One difficulty in respect to (2) and (3) is to define desirable levels of between-measures correlations or factor loadings. Here, we refer to them as low in a purely descriptive manner. Whether they are considered “low” in an evaluative manner depends on our theoretical assumptions of how much our measures should tap into a common factor. As pointed out by Bandalos and Gerstner (2016), the criterion for whether a factor loading is considered meaningful should be determined by the breadth of the theoretical concept under evaluation. As a field, we should reflect on what to expect given our theoretical assumptions and the indicators that we select to measure attentional control. For example, when administering multiple versions of the same paradigm (e.g., an arrow flanker and a letter flanker), we would reasonably expect a high(er) correlation; similarly, we could expect higher correlations between measures involving the same source of distraction and/or underlying attentional control mechanism. Alternatively, if we use paradigms that vary strongly in terms of lower-level processes, we might be satisfied with lower factor loadings from each measure as it would suggest attentional control was separated from these processes that differ across paradigms. Agreeing on what to expect will provide benchmarks to compare our data to.

Based on our survey, are there any paradigms that do a better job in measuring attentional control? It depends on how we define “better”. If we define “better” as a task with high reliability and (somewhat) larger correlations and factor loadings, the measures involving distraction from self-generated information – especially those typically used to assess the updating latent factor – stand out. However, as a large portion of performance variance in these measures is due to maintenance in memory (Frischkorn, von Bastian, Souza, & Oberauer, 2020), using updating measures (alone) may not be ideal to assess attentional control more broadly. From the paradigms that assess attentional control more directly, the antisaccade and the go/no-go paradigm stand out. What characteristics of these paradigms make them more successful? One main characteristic of antisaccade and go/no-go tasks is that researchers typically extract single scores from these paradigms rather than condition-difference scores. The practice of using condition-difference scores has a long history in psychology (Donders, 1969; for a review, see Draheim et al., 2019). It is rooted in the assumption that it allows for a purer measurement of attentional control by eliminating contamination from general performance factors such as speed of processing, task strategy, or memory (although there are reasons to question this assumption, see Hedge et al., 2020; Hedge, Powell, Bompas, et al., 2018; J. Miller & Ulrich, 2013; Wickelgren, 1977), and one cannot simply substitute a condition-difference score with performance in just the condition with high attentional control demands without considering the contribution of general processing factors (Goodhew & Edwards, 2019). Indeed, one could argue that the single scores used in antisaccade and go/no-go tasks often conflate attentional control with other cognitive abilities. For example, it is conceivable that the accuracy of responses in antisaccade trials depends not solely on the efficiency of inhibiting saccades towards a sudden onset stimulus but also on the speed of saccades towards the target and the ability to discriminate its identity. Hence, in antisaccade trials, low or high levels of accuracy may be partially related to motor speed and discrimination abilities. Thus, in implementations of this task that do not control for these baseline cognitive abilities (e.g., by contrasting an antisaccades condition to a prosaccades condition)[3], the derived accuracy score may conflate attentional control with

[3] Note that this approach achieves its purpose only if prosaccades performance is not at ceiling, in which case the subtraction would be moot or even detrimental, as random variation around ceiling would increase noise in the measure. An alternative approach could be to calibrate the presentation duration of

other abilities (see Rey-Mermet et al., 2019). This increases the score's reliability, its correlations with other tasks, and its power to predict other cognitive abilities. However, the same argument could be made for single-score measures of other abilities that could, at least partially, also be confounded with processing speed (e.g., working memory, see Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). Moreover, one could argue that even basic processing speed paradigms – such as vigilance tasks – require some goal maintenance and distraction avoidance and, therefore, assess attentional control (Cepeda, Blackwell, & Munakata, 2013).

Taken together, avoiding condition-difference scores might come at the price of providing less clarity regarding the share of variance attentional control explains: Are these tasks truly measuring variance due to attentional control, or could they be measuring variance due to other cognitive abilities as well? And where do we draw the theoretical line between what is attentional control and what are other cognitive abilities? This does not mean that tasks that use condition-difference scores are automatically better; they may still contain (sometimes substantial) variance from other processes (Hedge et al., 2020) and lack construct validity to the degree that they do not capture construct variance.

Importantly, reliability and shared variance alone cannot determine the value of a paradigm for every context. Another criterion for determining what makes a "better" paradigm is to consider its sensitivity for predicting future behaviors (Yarkoni & Westfall, 2017) and real-world outcomes such as psychopathology (e.g., Hutton & Ettinger, 2006). It is possible that paradigms that do not discriminate well between healthy young individuals may still be sensitive to aggregate differences between groups, for example when comparing patient groups with healthy controls. A possible reason for this is that these groups are often sampled from the extreme ends of the spectrum. For example, the Mini-Mental State Exam

---

antisaccade trials based on participants' speed in prosaccade blocks, thereby controlling for differences in motor processes (Rey-Mermet et al., 2019).

(MMSE; Folstein, Folstein, & McHugh, 1975) is a good screening tool for detecting dementia, but not a good measure of individual differences of cognitive ability in healthy young adults. Another possibility is that attentional control effects may be strongly situational, with fluctuations in performance on the individual level but robust effects on the group aggregate level. For example, such patterns have been observed for the implicit associations test, a measure for implicit biases: Whereas individual differences in implicit biases are only weakly correlated with behavioral outcomes, implicit biases assessed on the aggregate level of countries or states have been predictive of the levels of disparities or discrimination (De Schryver, Hughes, Rosseel, & De Houwer, 2016; Payne, Vuletich, & Lundberg, 2017).

Taken together, the antisaccade and the go/no-go paradigms seem to fare “better” regarding reliability, inter-task correlations, and factor loadings, possibly because these tasks avoid condition-difference scores (see Draheim et al., 2019). However, some researchers question their construct validity (e.g., Rey-Mermet et al., 2019). In a nutshell, it might be premature at this point to single out “better” tasks. More groundwork is necessary to establish the robustness and construct validity of the many paradigms used to assess attentional control and the performance scores they yield. Moreover, especially in more applied settings, it will be important to consider the sensitivity and predictive power of attentional control paradigms.

The Role of Sample Diversity

In our survey of the literature, we discovered that only a few studies investigated individual differences in attentional control with identical task batteries but in different samples. However, differences in the characteristics of the participants’ samples across studies is an important source of variability: homogeneous samples of introductory psychology students at selective universities may produce different (mean and variability) levels of performance than samples that recruit community volunteers across a wider range of demographic variables, even if identical attentional control tasks are administered. A change in the distribution of scores can also affect reliabilities, intercorrelations among, and the factor structure of latent variables derived from, attentional control tasks. Specifically, homogeneous student samples could entail a restriction of true variance in the attentional-control scores, resulting in low reliability and inter-task correlations. To get a better idea of how sampling decisions can affect the conclusions of latent-variable studies, we can look to the working memory individual differences literature. For example, Shah and Miyake (1996) and Kane et al. (2004) arrived at different conclusions about the strength of domain-general versus domain-specific working memory constructs (see also Redick & Lindsey, 2013; Unsworth & Engle, 2007, for discussion of and analyses pertaining to heterogeneous versus homogeneous samples in working memory studies). Participants' age across studies is also important to consider when comparing results. For example, St. Clair-Thompson (2010) showed that the same battery of memory span tasks yielded a different factor structure in children versus young adults: backwards digit span loaded onto a working memory latent variable in children but loaded onto a short-term memory latent variable in adults. These examples demonstrate that researchers should also consider the role of sample composition in conclusions about the factor structure of the attentional control construct.

Addressing Poor Indicators of Attentional Control: Analytical Considerations

The findings from our survey of existing studies demonstrate the problems of how we currently assess attentional control. One problem of most of the existing paradigms is that they were originally designed to estimate attentional control effects at the group level. This means that differentiating between individuals was not the focus of their design, and individual differences can in fact be disadvantageous to obtain robust group effects. With this methodological limitation, it is hard to tell whether the difficulty to establish an attentional control latent factor is the result of poor statistical properties of our measures, the impurity of our measures, or whether it indeed means that there is no such thing as a (domain-general) attentional control ability.

Here, we discuss different analytical approaches that have been proposed to address these issues; Table 3 summarizes their strengths and weaknesses. These approaches are not mutually exclusive and not specific to the study of attentional control, and they overlap with recent recommendations for the translation of experimental paradigms to the study of individual differences (Goodhew & Edwards, 2019). We stress that there is unlikely to be a single approach that works best for every dataset and research question, and we are not suggesting that every study of individual differences in attentional control should apply all these techniques. Instead, we aim to provide a guide to the tools that are available to inform best practice.

Table 3

Strengths and Weaknesses of Approaches to Measurement and Analysis of Attentional Control

Approach	Characteristics
Single scores (e.g., mean RTs or accuracy in incongruent trials)	+ Easy to calculate + Good reliability - Sensitive to speed-accuracy trade-offs and other strategies and biases - Interpretation complicated by confounds from variance unrelated to attentional control
Condition-difference scores (e.g., attentional control effects in mean RT or accuracy)	+ Easy to calculate + Widely used - Sensitive to general processing speed (J. Miller & Ulrich, 2013), speed-accuracy trade-offs and other strategies and biases - Risk of poor reliability - Requires large numbers of trials for reliable measurement
Integrative scores (e.g., inverse-efficiency, linear integrated speed-accuracy score, balanced-integration score)	+ Easy to calculate + Incorporates multiple elements of performance + Improved control for speed-accuracy trade-offs and other strategies and biases (but see Liesefeld & Janczyk, 2019) - Reliability of difference scores still potentially problematic - Relative weighting of RT and accuracy can lack theoretical justification - No single agreed-upon method
Adaptive/threshold procedures	+ Improved control for speed-accuracy trade-offs + Avoids reliability issues of difference scores - Unclear psychometric properties
Distributional measures (e.g., delta plots slopes, conditional accuracy functions)	+ Links to theoretical mechanisms - Lack a single representative score - Poor reliability - Sensitive to speed-accuracy trade-offs and other strategies and biases
Hierarchical models	+ Estimate correlations without attenuation by trial noise - Requires large numbers of trials for reliable measurement
Cognitive measurement models	+ Incorporates multiple elements of performance + Improved control for speed-accuracy trade-offs and other strategies and biases + Makes theoretical assumptions explicit and allows for direct theoretical interpretation - Some models are difficult to estimate - Relies on strong assumptions about underlying mechanisms

Computing the Dependent Measure: Reaction Times, Accuracy, or Both?

Existing attentional control paradigms differ in whether performance variability is (expected to be) reflected in RTs, accuracy, or both. In principle, researchers can take three approaches. The first approach is to design tasks that promote variation in RT across conditions while accuracy is maintained above a threshold level. This approach is implemented by designing tasks in which it is easy to achieve perfect accuracy, so that variability in performance is pushed into RTs which then serve as dependent variable. For example, naming the colors in a Stroop task correctly is very simple despite the additional time it costs to overcome the conflict from the incongruent print color information. The second approach is the reverse: design tasks that push the variability in performance into accuracy, which is then analyzed as dependent variable. A typical example for this approach is the antisaccade task where executing the antisaccade and correctly identifying the stimulus before it is masked is challenging enough to generate inter-individual variation. The third approach is to let accuracy and RT covary together so that the dependent measures are the joint consideration of RT and accuracy. One option is to combine them into a single (integrative) score, for example by computing inverse-efficiency (Townsend & Ashby, 1983), rate-correct (Woltz & Was, 2006), linear-integrated speed-accuracy (Vandierendonck, 2017), or balanced-integration scores (Liesefeld & Janczyk, 2019), or by using binning procedures (Draheim, Hicks, & Engle, 2016; M. M. Hughes, Linck, Bowles, Koeth, & Bunting, 2014). Another option is to analyze RTs and accuracies with a process model such as the diffusion model (Ratcliff, 1979) or a race model (Brown & Heathcote, 2008).

Most studies in our survey extracted either accuracies or RTs rather than considering them jointly. However, the choice of whether to use RT or accuracy is not trivial. A recent meta-analysis, including but not limited to attentional control tasks, showed that the correlation between attentional control effects measured in RT and accuracy taken from the same tasks is low (mean $r = .17$) and heterogeneous (Hedge, Powell, Bompas, et al., 2018).[4] Thus, even if we use condition-difference scores to isolate mechanisms of attentional control from other factors, we cannot take for granted that either RT or accuracy-based measures exclusively map on to attentional control ability.

Interpreting findings from solely RT-based paradigms can be challenging for at least two reasons: Studies vary in how raw RT data are processed prior to analysis, and the derived RT measures are sensitive to speed-accuracy trade-offs (Draheim et al., 2019; Rey-Mermet et al., 2019). Frequently applied data preprocessing procedures include data transformations and trimming/outlier removal. Researchers typically use data transformations to meet the assumptions of the statistical analyses they wish to use (see also Ratcliff, 1993). For example, RT distributions are often positively skewed, and researchers can select from a range of non-linear transformations (e.g., logarithmic, reciprocal, or arcsine) to approximate normality and homoscedasticity as required by many analyses (Box & Cox, 1964). In addition, many researchers remove outlier RTs, because individual RTs can strongly influence measures of central tendency and, thereby, the dependent measures derived from attentional control tasks. Methods to detect and remove outliers include distribution-based approaches (e.g., excluding RTs smaller or larger than $M \pm 2$ *SD*s, or median ± 3 median absolute deviations), absolute cutoff values (e.g., excluding RTs smaller than 250 ms), or a mix of both (e.g., excluding RTs smaller than 250 ms and larger than $M + 2$ *SD*s), and they can be applied globally for the full sample of participants, separately per participant, or separately per participant and condition. Furthermore, as RT data are prone to sequential effects such as post-error slowing (e.g., Laming, 1979a, 1979b; Rabbitt, 1966), some researchers remove RTs from trials following an error. This large number of (combinations of) options complicates comparing findings across the literature and may contribute to the inconsistencies in conclusions (for some examples

[4] Note that if performance variance is pushed into either RTs or accuracy by design, one would not expect a correlation between RTs and accuracy given the lack of variance in the other variable.

where applying different procedures led to diverging conclusions, see Balota, Aschenbrenner, & Yap, 2013; Zhou & Krott, 2016). The lack of consensus regarding preprocessing methods increases the “researchers degrees of freedom”, that is, it adds potential flexibility in picking a method that produces the desired results (Simmons, Nelson, & Simonsohn, 2011). Questionable research practices aside, also researchers with the best intentions will find it challenging to decide which procedure to use. For example, should you apply the procedure that you perceive as standard in your field? Or should you instead choose the procedure that yields the highest reliability? A way to address this problem is through applying a “multiverse” of scoring methods (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016), that is, running the analysis for a broad range of possible scoring methods to test whether the substantive results are sensitive to variations in scoring methods (for an example, see Rey-Mermet et al., 2018).

A second challenge in interpreting solely RT-based paradigms is their sensitivity to speed-accuracy trade-offs.[5] These arise because, in any given situation, people may decide to sacrifice accuracy to respond more quickly (thereby reducing the condition-related variability in RTs), or they may decide to respond slowly but very accurately (unduly inflating RT condition difference scores). Importantly, people can differ in the extent to which they favor speed over accuracy (or vice versa), which may further vary depending on the particular attentional control task. To the extent that individual differences in speed-accuracy trade-offs are unrelated to individual differences in attentional control, they add noise to the measurement of attentional control. Therefore, Draheim et al. (2019) proposed to avoid using paradigms in which participants are instructed to focus on both speed and accuracy by designing tasks in which only speed or accuracy are important. In these tasks, performance does not depend on the speed of a response (e.g., in variations of the antisaccade paradigm

[5] Note that speed-accuracy trade-offs can also occur for paradigms solely based on accuracies (Hedge et al., 2020; Hedge, Powell, Bompas, et al., 2018).

where the direction of the saccade is tracked and used as dependent variable; Unsworth, Schrock, & Engle, 2004), making RT-variance irrelevant for explaining individual differences and allowing researchers to draw inferences solely based on accuracy.

Another alternative proposed by Draheim et al. (2019) is to use adaptive or threshold paradigms that adjust the task difficulty based on an individual's performance. A typical example is the stop-signal task, where the time between stimulus presentation of the stop signal is adapted so that participants can successfully stop their reaction in 50% of the trials. In adaptive paradigms, the measure of interest is the value of the task parameter that was adapted to reach the desired performance criterion. For example, in the stop-signal task, the dependent measure (i.e., the SSRT) is the interval between stimulus presentation and stop signal at which participants achieve the 50% response stopping-success threshold. Whether the adaptive approach can overcome the problems associated with traditional paradigms is unclear at this point. In our survey, the stop-signal task did not show higher reliability, factor loadings, or correlations with other tasks than non-adaptive paradigms. Furthermore, little is known about the reliability and validity of the adaptive procedures to extract the measures, for example whether repeated cycles of adjustment of task difficulty correlate with each other and retain the rank-ordering of individuals, or whether adjustment criteria (which may differ between tasks) properly discriminate between individuals in terms of attentional control ability.

One criticism of designing tasks that are based solely on accuracy or another single score across all trials is that they do not isolate the variance contributed by attentional control to solving the task (see the above discussion of the possible problems with the antisaccade task). To avoid this problem while addressing the issue of speed-accuracy trade-offs, Rey-Mermet et al. (2019) implemented response deadline versions of traditional paradigms (including Stroop, flanker, Simon, go/no-go, stop-signal, and antisaccade tasks) while maintaining the contrast between conditions with low and high attentional control demands. Using a threshold method, stimuli display durations were calibrated individually in a baseline phase with low attentional-control demands. The resulting display duration was then used as the response deadline in the conditions with high attentional control demands, thereby accounting for individual differences in processing speed. Moreover, the response deadline ensured that, independent of individual differences in speed-accuracy trade-offs, lower attentional control yielded lower accuracies: If participants slowed down, they missed the deadline more often; if they speeded up and sacrificed accuracy, they made more errors. The resulting task scores had mostly acceptable reliability and produced traditional attentional control effects (e.g., congruency effects). Nevertheless, inter-task correlations remained low, and the tasks did not load on a coherent factor of attentional control. These findings suggest that accounting for speed-accuracy trade-offs may improve the reliability of attentional control measures. Yet, these findings also question whether low correlations and low loadings on a latent factor of attentional control are fully explained by a failure to take speed-accuracy trade-offs into account.

Another method to account for construct-irrelevant variance is to assess potential sources of contamination (e.g., processing speed) and use latent-variable analysis to explicitly model them, thereby isolating the construct-relevant variance. For example, Draheim, Tsukahara, Martin, Mashburn, and Engle (2020) assessed attentional control with accuracy-based measures – an antisaccade task, a vigilance task, and a visual working memory task (visual arrays) with a selection demand[6] – and examined to what extent they predicted

[6] The visual arrays task used – a change detection paradigm – and the dependent measure derived (Cowan's *k*) are typically used to assess working memory capacity. Moreover, an exploratory factor analysis showed that although this measure loaded most strongly on the attentional control factor (.37), it loaded almost equally strong on the working memory factor (.30) and the fluid intelligence factor (.27). Therefore, there is disagreement among the present authors whether this paradigm reflects attentional control adequately. See Martin et al. (2020) for an argument as to why individual differences in visual arrays *may* primarily reflect attention.

working memory capacity and fluid intelligence. To account for potential contamination of attentional control from individual differences in general speed, Draheim and colleagues also included a processing speed factor in their model. They found that the attentional control factor and the processing speed factor indeed shared a substantial amount of variance – about 41%. Critically, however, the attentional control factor accounted for substantial incremental variance in working memory capacity (69%) and fluid intelligence (38%) above and beyond processing speed, whereas processing speed contributed no statistically significant variance to these constructs after accounting for attention control. Thus, this approach allowed Draheim and colleagues to assess and account for the contamination of attentional control from individual differences in processing speed.

Addressing Trial Noise: Hierarchical Models

An additional, often neglected element that contributes to poor measurement properties is trial-to-trial variability of RT and accuracy (henceforth called trial noise). Trial noise is measurement noise that attenuates between-task correlations (Enkavi et al., 2019; Hedge, Powell, & Sumner, 2018; Rouder, Kumar, & Haaf, 2019; Whitehead et al., 2019). Rouder et al. (2019) surveyed 15 attentional-control studies that reported RT-based tasks and found that trial noise (expressed as *SD*) was between 100 ms and 250 ms, with a median of about 175 ms. They also found that the *SD* of true individual effects was about 25 ms, which constitutes only 1/7 of the trial noise.

One of the challenges in studying individual differences in attentional control is in overcoming this trial noise. Experimenters can remedy trial noise by running more trials to get a more precise estimate of each individual’s true score (Whitehead et al., 2019). Furthermore, hierarchical models are useful to estimate the extent to which trial noise attenuates the correlations between attentional control tasks. In hierarchical models, trial noise and true variability of experimental effects are estimated separately. As a consequence, the estimates of the true variability of the attentional control effect in one task can then be correlated with that of another task without attenuation (Matzke et al., 2017; Rouder & Haaf, 2019). However, applying hierarchical models alone cannot make up for all problems coming with current implementations of attentional control paradigms. For example, Rouder et al. (2019) estimated the true variability of attentional control effects in data from four attentional control tasks (color Stroop, number Stroop, arrow flanker, and letter flanker task) reported by Rey-Mermet et al. (2018). Correlations remained low for 5 out of the 6 possible between-task combinations even when accounting for trial noise (*r*s ≤ .12). One reason for the low correlations could be that the number of trials per condition (*n* = 72) was relatively small for hierarchical modeling. Another reason could be that the true variability is not large in these tasks to begin with.

Taken together, if researchers wish to understand the structure of the attentional control construct, they should account for trial noise, which is often the largest source of nuisance variation in attentional control paradigms. Hierarchical models are useful for estimating and accounting for trial noise. However, in situations where there is particularly little variability in the true effects, even the deattenuated correlations may remain low.

Formalizing Conceptual Frameworks: Cognitive Models

Formal modeling approaches for data from attentional control tasks include (but are not limited to) evidence accumulation models, connectionist models, and dynamic neural field models (Bompas & Sumner, 2011; for reviews and examples, see Botvinick et al., 2001; Chuderski & Smolen, 2016; Herd et al., 2014; Munakata et al., 2011; Oberauer et al., 2013; Schall, Palmeri, & Logan, 2017; Ulrich et al., 2015; Verguts & Notebaert, 2009; Wiecki, et al. 2013; White et al., 2013; White, Servant, & Logan, 2018). An advantage of formal modeling approaches is that they make explicit that there are multiple processes that underlie performance in attentional control tasks. In some cases, they also provide a method to jointly account for RTs and accuracy. Here, we highlight ways in which models can be applied to aid our understanding of individual differences in attentional control and the considerations that arise from doing so.

How do attentional control effects manifest in behavior? A first step in formulating a formal model of the mechanisms underlying attentional control is to identify its signatures in behavior that distinguish it from other factors (e.g., basic processing speed or speed-accuracy trade-offs) that may confound its measurement. Attentional control paradigms that rely on congruency manipulations all produce congruency effects – on average, incongruent trials produce more errors and slower RTs than congruent trials – but they show different patterns when we examine how these effects emerge in fast compared to slow responses (Pratte, Rouder, Morey, & Feng, 2010). These patterns can be observed using distributional analysis methods such as conditional accuracy functions (CAFs) and delta functions. Figure 8 shows the diversity of CAF (top panel) and delta function (bottom panel) patterns from four commonly used attentional control tasks.

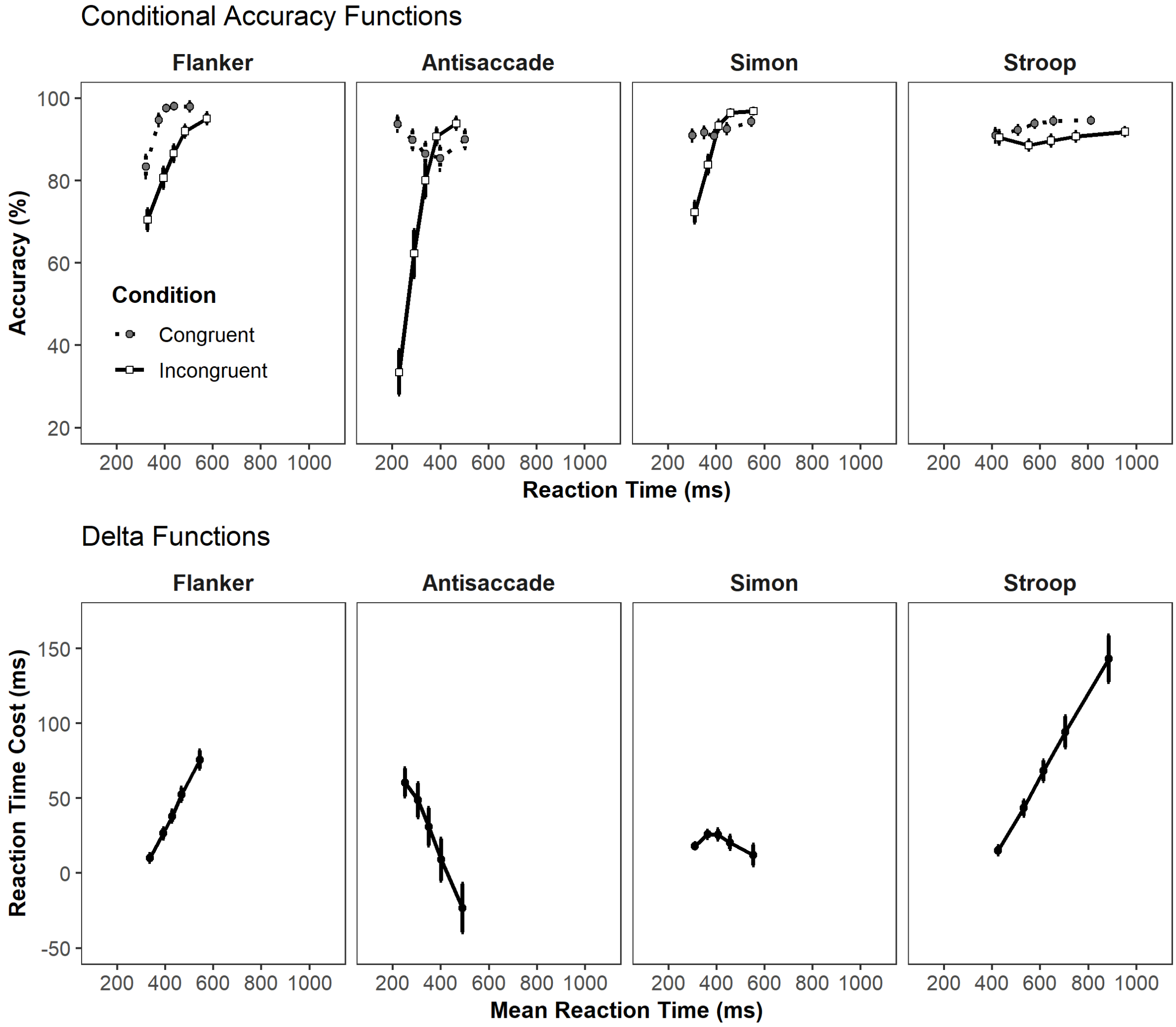

Figure 8. Conditional accuracy functions (CAFs; top row) and delta functions (bottom row) for four attentional control tasks. Flanker and Stroop data are from Hedge, Powell, and Sumner (2018). Antisaccade data are healthy control participants from Wiecki et al. (2016 from the Track-HD dataset; Tabrizi et al., 2009). Simon data are from Hedge, Powell, Bompas, et al. (2018). All tasks included congruent and incongruent (e.g., prosaccade and antisaccade) trials that were intermixed within blocks.

CAFs are constructed by first binning RTs from each condition into quantiles and then plotting the accuracy for each quantile. For many attentional control tasks, it is assumed that conflicting stimulus features are processed quickly and automatically, resulting in fast errors to incongruent stimuli, so that congruency effects are larger in the earlier parts of the CAF. However, this is not true for all tasks. As illustrated in the top panel of Figure 8, the Stroop task is an example where errors are not limited to fast responses.

Delta functions are constructed by computing the difference between RT quantiles in each condition and plotting them against the mean of those quantiles (De Jong, Liang, & Lauber, 1994). For example, in the bottom panel of Figure 8, the leftmost data points are the difference between the shortest 20% of congruent RTs and the shortest 20% of incongruent RTs. Typically, means and standard deviations are larger in more difficult conditions relative to easier conditions. This pattern almost always manifests as an increasing delta-plot function, indicating that the congruency effect is largely present in the slowest responses. The increasing delta plot pattern is common across many tasks, for example in studies where the strength of to-be-detected stimuli is manipulated (e.g., the intensity or duration of a light, see Pratte et al., 2010). However, there are a few tasks that consistently show the opposite pattern. For example, Simon interference is often characterized by decreasing delta plot functions, indicating that the congruency effect is largest for the shortest RTs. It becomes smaller, and sometimes even reverses, for longer RTs. Negative going delta slopes have been interpreted to reflect a strong influence of the conflicting response early on, which is subsequently suppressed (Ridderinkhof, 2002). An alternative account assumes that irrelevant information is subject to spontaneous activation decay (Hommel, 1994).

Although distributional analyses have previously been used to examine individual differences (e.g., Forstmann et al., 2008; Ridderinkhof, Scheres, Oosterlaan, & Sergeant, 2005), they are subject to many of the measurement concerns that we highlighted above. Most notably, they involve calculations based on subsets of trials and, thus, need larger trial numbers than other methods to be estimated reliably. Nevertheless, they illustrate patterns of behavior that are not well represented by condition means. If general mechanisms of attentional control exist, the diversity in temporal dynamics of behavior indicates that they manifest differently in different tasks (Ulrich et al., 2015).

Testing theories and selecting an appropriate model. A researcher who is interested in applying a formal modelling approach to understanding attentional control will inevitably face the question of which model(s) they should use. This involves both theoretical and pragmatic considerations. Theoretically, the question of whether a model is appropriate for a task is inherently tied into the question of what the underlying mechanisms are, and this question is not specific to the study of individual differences. One approach is to begin with a single architecture, and implement a theoretical mechanism thought necessary to account for key features of empirical data (e.g., conflict detection; Botvinick et al., 2001). Another approach is to qualitatively and/or quantitatively compare how well models with different assumptions/architectures can account for the data (Chuderski & Smolen, 2016; White et al., 2011).

Multiple models can provide adequate accounts, or some models can provide better accounts than others in some contexts but not others. So which model is the best model for understanding individual differences? Lewandowsky and Farrell (2011) suggest the choice of model can be guided by principles of sufficiency and verisimilitude. A model can be sufficient if it can account for key empirical data patterns and expected experimental manipulations (see Oberauer et al., 2018, as an example of benchmarks in the field of working memory). Verisimilitude, or partial truth value (Popper, 1963), appeals to the explanatory successes and novel predictions that a model demonstrates over alternative approaches. For example, if evidence accumulation models provide a way in which we can understand why attentional control effects in RTs and accuracy do not correlate across tasks (Hedge, Powell, Bompas, et al., 2018), then there can be value in applying such a model even if we know that they are a simplification. It is worth noting though that, just like experimental tasks, some models may work well in within-subject contexts but are not ideally suited for studying individual differences. Complex models may implement mechanisms with a strong basis in evidence, and they may provide the best fits to the average performance in empirical data (e.g., Ratcliff & Tuerlinckx, 2002; Usher & McClelland, 2001). However, simpler models are sometimes better for capturing individual differences in the parameters that researchers are primarily interested in (Dutilh et al., 2019; Lerche & Voss, 2016; Miletić, Turner, Forstmann, & van Maanen, 2017; van Ravenzwaaij, Donkin, & Vandekerckhove, 2017; van Ravenzwaaij & Oberauer, 2009).

Quantifying latent processes. If we accept that there are multiple psychological processes underlying a given task, and we have an appropriate model that captures those processes, then one avenue is to fit the model to the data of each individual and examine correlations between parameters. In principle, this enables us to examine individual differences in latent psychological processes of interest, while at the same time controlling for processes that may confound interpretations of RTs and accuracy in isolation. For example, recent applications of the diffusion model for attentional control tasks (Ulrich et al., 2015) to data from flanker, Simon and Stroop tasks revealed between-task correlations in parameters representing information processing, response caution, and the duration of perceptual and motor processes (Hedge et al., 2020; Hedge, Vivian-Griffiths, Powell, Bompas, & Sumner, 2019; see also Eisenberg et al., 2019). However, they observed no correlations in parameters representing conflict processing. Notably, simulations from the diffusion model for attentional control tasks show that underlying correlations in both conflict and non-conflict parameters can drive correlations in behavioral measures of attentional control (e.g., attentional control effects on RTs; Hedge et al., 2020). These findings suggest that there may be commonality between tasks in specific processes that is often not captured by measures of average performance.

As noted above, it is important to evaluate the appropriateness of a model for a given purpose and dataset. There are also methodological and pragmatic considerations. Formal models often require at least several hundred trials per condition to obtain adequate fits, though this is also true for adequately measuring attentional control without cognitive models (Hedge, Powell, & Sumner, 2018; Rouder et al., 2019; Whitehead et al., 2019). Other obstacles may be the skills required to implement models, the availability of existing implementations, and the time required to fit large datasets. For example, readily available implementations of a simplified diffusion model provide instantaneous parameter estimates (EZ-diffusion model; Wagenmakers, van der Maas, & Grasman, 2007). However, for other models and implementations, fitting data of a single participant can take hours, days, or so long as to be impractical for formal fitting approaches (Bompas, Hedge, & Sumner, 2017; White et al., 2018).

Taken together, applying formal models to attentional control data has the unique advantage of enabling precise and falsifiable predictions. However, existing models do not fit all purposes and can be challenging to implement. Inevitably, researchers will have to balance these considerations with the value a model adds to a given research question.

Summary and Future Directions

To better understand attentional control as a psychometric construct, we need to reconsider what we want to measure and how. The traditionally broad definition of attentional control resulted in a multitude of paradigms ostensibly assessing it. However, a growing number of unsuccessful attempts to establish robust correlations between measures from those paradigms suggest that new conceptual, methodological, and analytical approaches are necessary to advance our understanding of individual differences and mechanisms of attention control.

Clarifying the Definition of Attentional Control

Although the assumption of a general ability to control attention is plausible and appealing by its simplicity, a closer examination of proposed theoretical mechanisms of attentional control and existing methods to assess attentional control suggests that different experimental paradigms engage different, only partially overlapping mechanisms. Thus, from a theoretical perspective, it is no surprise that some researchers find it difficult to establish a coherent psychometric construct of attentional control. To advance our understanding of individual differences in attentional control, it will be critical to select our measures based on the involved mechanisms.

As a working definition, we propose that attentional control is maintaining an operative goal, and goal-relevant information, in the face of distraction caused by the perceived environment, by self-generated information, or by habits. Future studies employing measures that can tap into these types of distraction to varying degrees will allow for more informative, theory-driven model comparisons and for gauging the convergent and divergent validity of the attentional control construct. However, as our survey of existing individual differences studies revealed, even where tasks are assumed to involve execution of the same mechanisms – such as the Stroop and the Simon task, or even different variations of the Stroop task – correlations between task congruency effects can be small, suggesting that our conception of the mechanisms involved may not be accurate. Thus, the conceptual frameworks that researchers are operating from may need further refinement (e.g., see Simpson & Carroll, 2019) or even rethinking (Doebel, 2020). Future research would further benefit from formalizing theoretical assumptions as parameters in computational models to provide more precise and falsifiable predictions.

Improving Methodology and Analysis Across Laboratories

Our survey further highlighted that existing studies vary greatly in the measures used. Even where the same or similar paradigms are employed, studies differ in how these paradigms are administered (e.g., stimuli, timing, or number of trials per condition) and how performance is scored. Our attempt to identify the most optimal measures for investigating individual differences in attentional control painted a bleak picture: Considering reliabilities, between-task correlations, and loadings on a common attentional control factor, most currently used measures appear unfit for this purpose. Critically, where reliabilities, correlations, and loadings were relatively higher (i.e., some implementations of antisaccade and go/no-go paradigms), no consensus exists regarding their construct validity.

The notoriously low reliabilities, especially of condition-difference scores, likely contribute to low correlations among attentional control tasks; notably though, low correlations have also been observed for the more reliable measures (e.g., the antisaccade task). Therefore, there is a need to develop better suited tasks with both strong measurement properties and high construct validity. Although recent attempts to avoid condition-difference scores by pushing between-subjects variance into accuracy yielded mixed results (Draheim et al., 2020; Rey-Mermet et al., 2019), the more successful tasks should be further validated and built upon across diverse samples of participants.

Task variations, even those that seem subtle, might add undesired variance between studies and complicate if not prevent replication of findings (see also Brenninkmeijer, Derksen, & Rietzschel, 2019). Cross-study comparisons are further complicated by differences in how raw data are processed and which dependent measures are analyzed. Some methodological choices will yield better measures than others, and some analytical choices will lead to more robust conclusions. To efficiently move forward, we encourage researchers to adopt transparent and open practices. Where there is no consensus, it is critical that researchers make those choices explicit and easily discoverable. By adopting transparent practices, researchers will be able to accurately assess the choices and work of others and, thus, be better able to incorporate practices that seem to result in progress (or drop those that do not) in their own work. Previously, the norms in the field (broadly defined) were such that not all aspects of the research process were reported (e.g., instructions, tasks, data, analysis scripts). This may have resulted in less than complete understanding of the methodological and analytical procedures and potentially led to redundant efforts. Moreover, by transparently reporting cross-laboratory differences in methodological and analytical procedures, we may be able to identify practices and scenarios that moderate attentional control effects.

Freely accessible, easily portable, and adaptable computer-based tasks can help to improve methodological consistency. Open study materials can reduce unintended task variations, differences in administration, analysis, and interpretation of results, thereby enabling a more cumulative investigation of attentional control. Examples for open attentional control materials can be found for a range of free, open-source platforms such as Tatool Web (see Task Library on www.tatool-web.com, von Bastian, Locher, & Ruflin, 2013). To evaluate the robustness of their analytical conclusions, researchers could adopt a “multiverse analysis” (Steegen et al., 2016), which involves conducting the analyses based on all reasonable alternative data preprocessing scenarios. Automatized analytical workflows (R Core Team, 2019) can enhance the feasibility of such an approach.

It is clear from the broad range of conceptual definitions, tasks, and analytical choices that we have highlighted here, that there are many ways of capturing data from attentional control investigations, each of which may provide a different window to a potential attentional control construct. For many studies it will be neither feasible nor desirable to apply every technique that we have considered. Nevertheless, we regularly collect data that can be used to inform questions other than our own. We can maximize the value of these datasets through the adoption of open science practices and cross-laboratory collaboration. For example, making data available on open science repositories allow them to be used in meta-analyses, or as benchmark data for the evaluation of new or existing methods (Hedge, Powell, Bompas, et al., 2018; Rouder & Haaf, 2019; Rouder et al., 2019). Moreover, as is the case for experimental methods, sharing analytical procedures will be critical for identifying differences in analytical procedures that may produce divergent results across laboratories.

Conclusion

Researchers investigating attentional control face a theoretical challenge where promising lines of research have stalled. That we, as a research community, identified the complexities of individual differences in attentional control and are now making concerted efforts to investigate these complexities is a tremendous progress. Moreover, recently, new encouraging methodological approaches have been explored to meet this challenge. To further advance our understanding of attentional control, we think a more transparent, collaborative, and ultimately cumulative approach will provide the best avenue for rapid theoretical development. In this collaborative approach, the field would be best served by transparently exploring (and reporting) the methodological and statistical practices outlined above. Moreover, if the data produced by this cumulative effort do not support the construct of attentional control as currently defined, a new conceptualization must be adopted and thoroughly tested. Given the work underway and the spirit of collaboration exhibited in crafting this article, we believe that we can eventually overcome the theoretical and methodological challenges to better understand attentional control.

References

*References marked with an asterisk indicate studies included in the literature survey.

*Adrian, J., Moessinger, M., Charles, A., & Postal, V. (2019). Exploring the contribution of executive functions to on-road driving performance during aging: A latent variable analysis. *Accident Analysis & Prevention*, *127*, 96–109. https://doi.org/10.1016/j.aap.2019.02.010

*Adrover-Roig, D., Sesé, A., Barceló, F., & Palmer, A. (2012). A latent variable approach to executive control in healthy ageing. *Brain and Cognition*, *78*(3), 284–299. https://doi.org/10.1016/j.bandc.2012.01.005

*Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology*, *105*(4), 286–305. https://doi.org/10.1016/j.jecp.2009.09.006

*Allom, V., & Mullan, B. (2014). Individual differences in executive function predict distinct eating behaviours. *Appetite*, *80*, 123–130. https://doi.org/10.1016/j.appet.2014.05.007

*Arán-Filippetti, V. (2013). Structure and invariance of executive functioning tasks across socioeconomic status: Evidence from Spanish-speaking children. *Spanish Journal of Psychology*, *16*, 1–15. https://doi.org/10.1017/sjp.2013.102

Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: The influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1563–1571. https://doi.org/10.1037/a0032186

Bandalos, D. L., & Gerstner, J. J. (2016). Using factor analysis in test construction. In K. Schweizer & C. DiStefano (Eds.), *Psychological assessment—science and practice: Vol. 3. Principles and methods of test construction: Standards and recent advances* (pp. 26–51). Boston, MA: Hogrefe Publishing. Retrieved from https://psycnet.apa.org/record/2016-44627-003

*Bartholow, B. D., Fleming, K. A., Wood, P. K., Cowan, N., Saults, J. S., Altamirano, L., Miyake, A., Martins, J., & Sher, K. J. (2018). Alcohol effects on response inhibition: Variability across tasks and individuals. *Experimental and Clinical Psychopharmacology*, *26*(3), 251–267. https://doi.org/10.1037/pha0000190

*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

Bettcher, B. M., Mungas, D., Patel, N., Elofson, J., Dutt, S., Wynn, M., Watson, C. L., Stephens, M., Walsh, C. M., & Kramer, J. H. (2016). Neuroanatomical substrates of executive functions: Beyond prefrontal structures. *Neuropsychologia*, *85*, 100–109. https://doi.org/10.1016/j.neuropsychologia.2016.03.001

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*(3–4), 296–312. https://doi.org/10.1111/j.1469-7610.2006.01611.x

Bompas, A., Hedge, C., & Sumner, P. (2017). Speeded saccadic and manual visuo-motor decisions: Distinct processes but same principles. *Cognitive Psychology*, *94*, 26–52. https://doi.org/10.1016/j.cogpsych.2017.02.002

Bompas, A., & Sumner, P. (2011). Saccadic Inhibition Reveals the Timing of Automatic and Voluntary Signals in the Human Brain. *Journal of Neuroscience*, *31*(35), 12501–12512. https://doi.org/10.1523/JNEUROSCI.2234-11.2011

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

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Brenninkmeijer, J., Derksen, M., & Rietzschel, E. (2019). Informal laboratory practices in psychology. *Collabra: Psychology*, *5*(1), 45. https://doi.org/10.1525/collabra.221

*Brocki, K. C., & Tillman, C. (2014). Mental set shifting in childhood: The role of working memory and inhibitory control. *Infant and Child Development*, *23*(6), 588–604. https://doi.org/10.1002/icd.1871

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

*Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence*, *40*(5), 458–469. https://doi.org/10.1016/j.intell.2012.05.006

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. M. Rabbitt (Ed.), *Theory and Methodology of Frontal and Executive Function* (pp. 81–116). Hove, UK: Psychology Press.

*Carlson, S. M., White, R. E., & Davis-Unger, A. C. (2014). Evidence for a relation between executive function and pretense representation in preschool children. *Cognitive Development*, *29*(1), 1–16. https://doi.org/10.1016/j.cogdev.2013.09.001

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cepeda, N. J., Blackwell, K. A., & Munakata, Y. (2013). Speed isn't everything: Complex processing speed measures mask individual differences and developmental changes in executive control. *Developmental Science*, *16*(2), 269–286. https://doi.org/10.1111/desc.12024

Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M. T., & Munakata, Y. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0031546

*Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition*, *42*(3), 448–463. https://doi.org/10.3758/s13421-013-0366-x

*Chuderski, A. (2015). High intelligence prevents the negative impact of anxiety on working memory. *Cognition and Emotion*, *29*(7), 1197–1209. https://doi.org/10.1080/02699931.2014.969683

Chuderski, A., & Smolen, T. (2016). An integrated utility-based model of conflict evaluation and resolution in the Stroop task. *Psychological Review*, *123*(3), 255–290. https://doi.org/10.1037/a0039979

*Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*(3), 278–295. https://doi.org/10.1016/j.intell.2012.02.010

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671–684. https://doi.org/10.1037/h0043943

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 731–750. https://doi.org/10.1037/0096-1523.20.4.731

De Schryver, M., Hughes, S., Rosseel, Y., & De Houwer, J. (2016). Unreliable yet still replicable: A comment on LeBel and Paunonen (2011). *Frontiers in Psychology, 6*, 2039. https://doi.org/10.3389/fpsyg.2015.02039

*De Simoni, C., & von Bastian, C. C. (2018). Working memory updating and binding training: Bayesian evidence supporting the absence of transfer. *Journal of Experimental Psychology: General*, *147*(6), 829–858. https://doi.org/10.1037/xge0000453

*Del Missier, F., Mäntylä, T., & Bruine de Bruin, W. (2010). Executive functions in decision making: An individual differences approach. *Thinking & Reasoning*, *16*(2), 69–97. https://doi.org/10.1080/13546781003630117

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science*, *15*(4), 942–956. https://doi.org/10.1177/1745691620904771

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

Draheim, C., Hicks, K. L., & Engle, R. W. (2016). Combining Reaction Time and Accuracy: The Relationship Between Working Memory Capacity and Task Switching as a Case Example. *Perspectives on Psychological Science*, *11*(1), 133–155. https://doi.org/10.1177/1745691615596990

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. (2020). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, Advance online publication. https://doi.org/10.1037/xge0000783

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., … Donkin, C. (2019). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. *Psychonomic Bulletin & Review*, *26*(4), 1051–1069. https://doi.org/10.3758/s13423-017-1417-2

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 170–189. https://doi.org/10.1037/a0017891

Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*(10), 374–380. https://doi.org/10.1016/j.tics.2008.07.001

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 2319. https://doi.org/10.1038/s41467-019-10301-1

*Engelhardt, L. E., Briley, D. A., Mann, F. D., Harden, K. P., & Tucker-Drob, E. M. (2015). Genes unite executive functions in childhood. *Psychological Science*, *26*(8), 1151–1163. https://doi.org/10.1177/0956797615577209

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477. https://doi.org/10.1073/pnas.1818430116

Erb, C. D., Moher, J., Sobel, D. M., & Song, J.-H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, *152*, 114–126. https://doi.org/10.1016/j.cognition.2016.03.015

Ettenhofer, M. L., Hambrick, D. Z., & Abeles, N. (2006). Reliability and stability of executive functioning in older adults. *Neuropsychology*, *20*(5), 607–613. https://doi.org/10.1037/0894-4105.20.5.607

*Fleming, K. A., Heintzelman, S. J., & Bartholow, B. D. (2016). Specifying associations between conscientiousness and executive functioning: Mental set shifting, not prepotent response inhibition or working memory updating. *Journal of Personality*, *84*(3), 348–360. https://doi.org/10.1111/jopy.12163

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Forstmann, B. U., Jahfari, S., Scholte, H. S., Wolfensteller, U., van den Wildenberg, W. P. M., & Ridderinkhof, K. R. (2008). Function and structure of the right inferior frontal cortex predict individual differences in response inhibition: A model-based approach. *Journal of Neuroscience*, *28*(39), 9790–9796. https://doi.org/10.1523/JNEUROSCI.1465-08.2008

*Fournier-Vicente, S., Larigauderie, P., & Gaonac’h, D. (2008). More dissociations and interactions within central executive functioning: A comprehensive latent-variable analysis. *Acta Psychologica*, *129*, 32–48. https://doi.org/10.1016/j.actpsy.2008.04.004

Frazier, D. T., Bettcher, B. M., Dutt, S., Patel, N., Mungas, D., Miller, J., … Kramer, J. H. (2015). Relationship between insulin-resistance processing speed and specific executive function profiles in neurologically intact older adults. *Journal of the International Neuropsychological Society*, *21*(8), 622–628. https://doi.org/10.1017/S1355617715000624

Friedman, N. P., Haberstick, B. C., Willcutt, E. G., Miyake, A., Young, S. E., Corley, R. P., & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science*, *18*(10), 893–900. https://doi.org/10.1111/j.1467-9280.2007.01997.x

*Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

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., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340. https://doi.org/10.1037/dev0000075

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.

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. T., 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. T., von Bastian, C. C., Souza, A. S., & Oberauer, K. (2020). *Individual differences in updating are not related to reasoning ability and working memory capacity*. Manuscript submitted for publication. https://doi.org/10.31234/osf.io/92uhb
Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26
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
Grange, J. A., & Cross, E. (2015). Can time-based decay explain temporal distinctiveness effects in task switching? *Quarterly Journal of Experimental Psychology*, *68*(1), 19–45. https://doi.org/10.1080/17470218.2014.934696
Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506. https://doi.org/10.1037/0096-3445.121.4.480
Gratton, G., Cooper, P., Fabiani, M., Carter, C. S., & Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. *Psychophysiology*, *55*(3), e13016. https://doi.org/10.1111/psyp.13016
Gustavson, D. E., Franz, C. E., Panizzon, M. S., Reynolds, C. A., Xian, H., Jacobson, K. C., Toomey, R., Lyons, M. J., & Kremen, W. S. (2019). Genetic and environmental associations among executive functions, trait anxiety, and depression symptoms in middle age. *Clinical Psychological Science*, *7*(1), 127–142. https://doi.org/10.1177/2167702618805075
Gustavson, D. E., Panizzon, M. S., Elman, J. A., Franz, C. E., Reynolds, C. A., Jacobson, K. C., Friedman, N. P., Xian, H., Toomey, R., Lyons, M. J., & Kremen, W. S. (2018). Stability of genetic and environmental influences on executive functions in midlife. *Psychology and Aging*, *33*(2), 219–231. https://doi.org/10.1037/pag0000230
*Gustavson, D. E., Panizzon, M. S., Franz, C. E., Friedman, N. P., Reynolds, C. A., Jacobson, K. C., Xian, H., Lyons, M. J., & Kremen, W. S. (2018). Genetic and environmental architecture of executive functions in midlife. *Neuropsychology*, *32*(1), 18–30. https://doi.org/10.1037/neu0000389
*Guye, S., & von Bastian, C. C. (2017). Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging*, *32*(8), 732–746. https://doi.org/10.1037/pag0000206
Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research*, *18*(10), 1279–1296. https://doi.org/10.1016/0042-6989(78)90218-3
Hasher, L., Lustig, C., & Zacks, R. (2007). Inhibitory mechanisms and the control of attention. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in Working Memory* (pp. 227–249). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0009
Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 163–169. https://doi.org/10.1037/0278-7393.17.1.163
*Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology*, *20*(5), 511–528. https://doi.org/10.1037/0894-4105.20.5.511
Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2020). *Strategy and processing speed*

eclipse individual differences in control ability in conflict tasks. Manuscript submitted for publication. https://doi.org/10.31234/osf.io/vgpxq

Hedge, C., Powell, G., Bompas, A., Vivian-Griffiths, S., & Sumner, P. (2018). Low and variable correlation between reaction time costs and accuracy costs explained by accumulation models: Meta-analysis and simulations. *Psychological Bulletin*, *144*(11), 1200–1227. https://doi.org/10.1037/bul0000164

*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., Vivian-Griffiths, S., Powell, G., Bompas, A., & Sumner, P. (2019). Slow and steady? Strategic adjustments in response caution are moderately reliable and correlate across tasks. *Consciousness and Cognition*, *75*, 102797. https://doi.org/10.1016/j.concog.2019.102797

Herd, S. A., O′Reilly, R. C., Hazy, T. E., Chatham, C. H., Brant, A. M., & Friedman, N. P. (2014). A neural network model of individual differences in task switching abilities. *Neuropsychologia*, *62*, 375–389. https://doi.org/10.1016/j.neuropsychologia.2014.04.014

Hester, R., & Garavan, H. (2004). Executive dysfunction in cocaine addiction: Evidence for discordant frontal, cingulate, and cerebellar activity. *Journal of Neuroscience*, *24*(49), 11017–11022. https://doi.org/10.1523/JNEUROSCI.3321-04.2004

*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

*Holochwost, S. J., Propper, C. B., Wolf, D. P., Willoughby, M. T., Fisher, K. R., Kolacz, J., Volpe, V. V., & Jaffee, S. R. (2017). Music education, academic achievement, and executive functions. *Psychology of Aesthetics, Creativity, and the Arts*, *11*(2), 147–166. https://doi.org/10.1037/aca0000112

Hommel, B. (1994). Spontaneous decay of response-code activation. *Psychological Research*, *56*(4), 261–268. https://doi.org/10.1007/BF00419656

Hommel, B., & Wiers, R. W. (2017). Towards a unitary approach to human action control. *Trends in Cognitive Sciences*, *21*(12), 940–949. https://doi.org/10.1016/j.tics.2017.09.009

Horoufchin, H., Philipp, A. M., & Koch, I. (2011). The dissipating task-repetition benefit in cued task switching: Task-set decay or temporal distinctiveness? *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 455–472. https://doi.org/10.1037/a0020557

Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the rransition to school: A latent variable approach. *Developmental Neuropsychology*, *35*(1), 20–36. https://doi.org/10.1080/87565640903325691

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior Research Methods*, *46*(3), 702–721. https://doi.org/10.3758/s13428-013-0411-5

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017–2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

*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

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology*, *43*(3), 302–313. https://doi.org/10.1111/j.1469-8986.2006.00403.x

*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

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., 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

Keye, D., Wilhelm, O., Oberauer, K., & Stürmer, B. (2013). Individual differences in response conflict adaptations. *Frontiers in Psychology*, *4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00947

Keye, D., Wilhelm, O., Oberauer, K., & van Ravenzwaaij, D. (2009). Individual differences in conflict-monitoring: Testing means and covariance hypothesis about the Simon and the Eriksen flanker task. *Psychological Research*, *73*(6), 762–776. https://doi.org/10.1007/s00426-008-0188-9

*Khng, K. H., & Lee, K. (2009). Inhibiting interference from prior knowledge: Arithmetic intrusions in algebra word problem solving. *Learning and Individual Differences*, *19*(2), 262–268. https://doi.org/10.1016/j.lindif.2009.01.004

*Khng, K. H., & Lee, K. (2014). The relationship between Stroop and stop-signal measures of inhibition in adolescents: Influences from variations in context and measure estimation. *PLoS ONE*, *9*(7), e101356. https://doi.org/10.1371/journal.pone.0101356

*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

Koch, I., Poljac, E., Müller, H., & Kiesel, A. (2018). Cognitive structure, flexibility, and plasticity in human multitasking—An integrative review of dual-task and task-switching research. *Psychological Bulletin*, *144*(6), 557–583. https://doi.org/10.1037/bul0000144

*Lambek, R., & Shevlin, M. (2011). Working memory and response inhibition in children and adolescents: Age and organization issues. *Scandinavian Journal of Psychology*, *52*(5), 427–432. https://doi.org/10.1111/j.1467-9450.2011.00899.x

Laming, D. (1979a). Autocorrelation of choice-reaction times. *Acta Psychologica*, *43*(5), 381–412. https://doi.org/10.1016/0001-6918(79)90032-5

Laming, D. (1979b). Choice reaction performance following an error. *Acta Psychologica*, *43*(3), 199–224. https://doi.org/10.1016/0001-6918(79)90026-X

*Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N. A. M., & Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: Executive functioning, patterns, numeracy, and arithmetic. *British Journal of Educational Psychology*, *82*, 82–99. https://doi.org/10.1111/j.2044-8279.2010.02016.x

Lee, T., Mosing, M. A., Henry, J. D., Trollor, J. N., Ames, D., Martin, N. G., Wright, M. J., & Sachdev, P. S. (2012). Genetic influences on four measures of executive functions and their covariation with general cognitive ability: The older Australian twins study. *Behavior Genetics*, *42*(4), 528–538. https://doi.org/10.1007/s10519-012-9526-1

*Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59–80. https://doi.org/10.1348/026151003321164627

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

*Lerner, M. D., & Lonigan, C. J. (2014). Executive function among preschool children: Unitary versus distinct abilities. *Journal of Psychopathology and Behavioral Assessment*, *36*(4), 626–639. https://doi.org/10.1007/s10862-014-9424-3

Lewandowsky, S., & Farrell, S. (2011). *Computational Modeling in Cognition: Principles and Practice*. Thousand Oaks, CA: Sage Publications. https://doi.org/10.4135/9781483349428

Liesefeld, H. R., & Janczyk, M. (2019). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods*, *51*(1), 40–60. https://doi.org/10.3758/s13428-018-1076-x

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). Madison, WI: University of Wisconsin Press.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. *Psychology of Learning and Motivation*, *43*, 163–214. https://doi.org/10.1016/S0079-7421(03)01014-4

Magnusdottir, B. B., Faiola, E., Harms, C., Sigurdsson, E., Ettinger, U., & Haraldsson, H. M. (2019). Cognitive measures and performance on the antisaccade eye movement task. *Journal of Cognition*, *2*(1). https://doi.org/10.5334/joc.52

Martin, J. D., Tsukahara, J. S., Draheim, C., Mashburn, C. A., Shipstead, Z., Vogel, E. K., & Engle, R. W. (2020). *The visual arrays task: Visual storage capacity or attention control?* Manuscript submitted for publication. https://doi.org/10.31234/osf.io/u92cm

*Martins, J. S., Bartholow, B. D., Cooper, M. L., Von Gunten, C. D., & Wood, P. K. (2018). Associations between executive functioning, affect-regulation drinking motives, and alcohol use and problems. *Psychology of Addictive Behaviors*, *32*(1), 16–28. https://doi.org/10.1037/adb0000324

*Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K., Long, J. D., & Zelazo, P. D. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher*, *41*(9), 375–384. https://doi.org/10.3102/0013189X12459883

Matzke, D., Ly, A., Selker, R., Weeda, W. D., Scheibehenne, B., Lee, M. D., & Wagenmakers, E.-J. (2017). Bayesian inference for correlations in the presence of measurement error and estimation uncertainty. *Collabra: Psychology*, *3*(1), 25. https://doi.org/10.1525/collabra.78

*McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of Experimental Child Psychology*, *108*(3), 453–468. https://doi.org/10.1016/j.jecp.2010.08.009

*McCabe, D. P., Roediger III, H. L., McDaniel, M. A., & Balota, D. A. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*(2), 222–243. https://doi.org/10.1037/a0017619

*McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, *141*(2), 302–320. https://doi.org/10.1037/a0025250

Miletić, S., Turner, B. M., Forstmann, B. U., & van Maanen, L. (2017). Parameter recovery for the Leaky Competing Accumulator model. *Journal of Mathematical Psychology*, *76*, 25–50. https://doi.org/10.1016/j.jmp.2016.12.001

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*(1), 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167

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

*Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, *13*(3), 395–423. https://doi.org/10.1080/15248372.2011.585478

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*, 49–100. https://doi.org/10.1006/cogp.1999.0734

*Monette, S., Bigras, M., & Lafrenière, M.-A. (2015). Structure of executive functions in typically developing kindergarteners. *Journal of Experimental Child Psychology*, *140*, 120–139. https://doi.org/10.1016/j.jecp.2015.07.005

Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(1), 11–18. https://doi.org/10.1037/a0031569

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O’Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences*, *15*(10), 453–459. https://doi.org/10.1016/j.tics.2011.07.011

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(2), 241–268. https://doi.org/10.3758/s13415-011-0083-5

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology*, *122*(11), 2185–2194. https://doi.org/10.1016/j.clinph.2011.03.030

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220–246. https://doi.org/10.1037/0033-2909.126.2.220

Norman, D. A., & Shallice, T. (1980). Attention to Action: Willed and Automatic Control of Behavior (Technical Report 8006). University of California, Center for Human Information Processing.

Norman, D. A., & Shallice, T. (1986). Attention to Action. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation* (pp. 1–18). Boston, MA: Springer. https://doi.org/10.1007/978-1-4757-0629-1_1

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A. A., Conway, A., Cowan, N., … Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. E. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: Experiments and a computational model. *Cognitive Psychology*, *66*(2), 157–211. https://doi.org/10.1016/j.cogpsych.2012.11.001

*Okada, B. M., & Slevc, L. R. (2018). Individual differences in musical training and executive functions: A latent variable approach. *Memory & Cognition*, *46*(7), 1076–1092. https://doi.org/10.3758/s13421-018-0822-8

*Oschwald, J., Schättin, A., von Bastian, C. C., & Souza, A. S. (2018). Bidialectalism and bilingualism: Exploring the role of language similarity as a link between linguistic ability and executive control. *Frontiers in Psychology*, *9*, 1997. https://doi.org/10.3389/fpsyg.2018.01997

*Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., & Sawi, O. (2016). The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods*, *274*, 81–93. https://doi.org/10.1016/j.jneumeth.2016.10.002

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in Methods and Practices in Psychological Science*, *2*(4), 378–395. https://doi.org/10.1177/2515245919879695

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248. https://doi.org/10.1080/1047840X.2017.1335568

*Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29*(2), 187–204. https://doi.org/10.1037/a0036085

Popper, K. R. (1963). *Conjectures and Refutations*. London, UK: Routledge.

Posner, M. I., & Snyder, C. R. R. (1979). Attention and Cognitive Development. In G. A. Hale & M. Lewis (Eds.), *Information processing and cognition: The Loyola symposium*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4613-2985-5

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception & Psychophysics*, *72*(7), 2013–2025. https://doi.org/10.3758/APP.72.7.2013

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Rabbitt, P. M. (1997). *Methodology of Frontal and Executive Function*. East Sussex, UK: Psychology Press.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, *86*(3), 446–461. https://doi.org/10.1037/0033-2909.86.3.446

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

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, *9*(3), 438–481. https://doi.org/10.3758/BF03196302

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *20*(6), 1102–1113. https://doi.org/10.3758/s13423-013-0453-9

*Redick, T. S., Shipstead, Z., Meier, M. E., Montroy, J. J., Hicks, K. L., Unsworth, N., Kane, M. J., Hambrick, D. Z., & Engle, R. W. (2016). Cognitive predictors of a common multitasking ability: Contributions from working memory, attention control, and fluid intelligence. *Journal of Experimental Psychology: General*, *145*(11), 1473–1492. https://doi.org/10.1037/xge0000219

*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

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research*, *66*(4), 312–323. https://doi.org/10.1007/s00426-002-0104-7

Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2005). Delta plots in the study of individual differences: New tools reveal response inhibition deficits in AD/HD that are eliminated by methylphenidate treatment. *Journal of Abnormal Psychology*, *114*(2), 197–215. https://doi.org/10.1037/0021-843X.114.2.197

*Ropovik, I. (2014). Do executive functions predict the ability to learn problem-solving principles? *Intelligence*, *44*, 64–74. https://doi.org/10.1016/j.intell.2014.03.002

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. (2019). *Why most studies of individual differences with inhibition tasks are bound to fail*. Manuscript in preparation. https://doi.org/10.31234/osf.io/3cjr5

Schachar, R. J., Forget-Dubois, N., Dionne, G., Boivin, M., & Robaey, P. (2010). Heritability of response inhibition in children. *Journal of the International Neuropsychological Society*, *17*(2), 238–247. https://doi.org/10.1017/S1355617710001463

Schall, J. D., Palmeri, T. J., & Logan, G. D. (2017). Models of inhibitory control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1718), 20160193. https://doi.org/10.1098/rstb.2016.0193

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
*Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence*, *32*(4), 329–347. https://doi.org/10.1016/j.intell.2004.06.006
Seli, P., Risko, E. F., Smilek, D., & Schacter, D. L. (2016). Mind-wandering with and without intention. *Trends in Cognitive Sciences*, *20*(8), 605–617. https://doi.org/10.1016/j.tics.2016.05.010
Shah, P., & Miyake, A. (1996). The separability of working memory ressources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology : General*, *125*, 4–27.
Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*(2), 374–408. https://doi.org/10.1037/a0034418
Shilling, V. ., Chetwynd, A., & Rabbitt, P. M. . (2002). Individual inconsistency across measures of inhibition: an investigation of the construct validity of inhibition in older adults. *Neuropsychologia*, *40*(6), 605–619. https://doi.org/10.1016/S0028-3932(01)00157-9
*Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, and Psychophysics*, *77*(6), 1863–1880. https://doi.org/10.3758/s13414-015-0899-0
*Shipstead, Z., Lindsey, D. R. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. https://doi.org/10.1016/j.jml.2014.01.004
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632
Simpson, A., & Carroll, D. J. (2019). Understanding early inhibitory development: Distinguishing two ways that children use inhibitory control. *Child Development*, *90*(5), 1459–1473. https://doi.org/10.1111/cdev.13283
Singh, K. A., Gignac, G. E., Brydges, C. R., & Ecker, U. K. H. (2018). Working memory capacity mediates the relationship between removal and fluid intelligence. *Journal of Memory and Language*, *101*, 18–36. https://doi.org/10.1016/j.jml.2018.03.002
*Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, *152*, 199–211. https://doi.org/10.1016/j.cognition.2016.03.017
Smith, L. L., Banich, M. T., & Friedman, N. P. (2019). Individual differences in mixing costs relate to general executive functioning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(4), 606–613. https://doi.org/10.1037/xlm0000613
Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*, 328. https://doi.org/10.3389/fpsyg.2015.00328
St Clair-Thompson, H. L. (2010). Backwards digit recall: A measure of short-term memory or working memory? *European Journal of Cognitive Psychology*, *22*(2), 286–296. https://doi.org/10.1080/09541440902771299
*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
Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency

through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

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

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic Modeling of Elementary Psychological Processes*. Cambridge, UK: Cambridge University Press.

Tucker-Drob, E. M. (2011). Individual differences methods for randomized experiments. *Psychological Methods*, *16*(3), 298–318. https://doi.org/10.1037/a0023349

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. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110–128. https://doi.org/10.1016/j.intell.2015.01.005

*Unsworth, N., Brewer, G. A., & Spillers, G. J. (2012). Variation in cognitive failures: An individual differences investigation of everyday attention and memory failures. *Journal of Memory and Language*, *67*(1), 1–16. https://doi.org/10.1016/j.jml.2011.12.005

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

*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., & McMillan, B. D. (2014). Similarities and differences between mind-wandering and external distraction: A latent variable analysis of lapses of attention and their relation to cognitive abilities. *Acta Psychologica*, *150*, 14–25. https://doi.org/10.1016/j.actpsy.2014.04.001

*Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences*, *30*(4), 194–200. https://doi.org/10.1027/1614-0001.30.4.194

*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., Schrock, J. C., & Engle, R. W. (2004). Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

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

*Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science Quarterly*, *51*(4), 388–402.

Usai, M. C., Viterbori, P., Traverso, L., & De Franchis, V. (2014). Latent structure of executive function in five- and six-year-old children: A longitudinal study. *European Journal of Developmental Psychology*, *11*(4), 447–462.

https://doi.org/10.1080/17405629.2013.840578

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

*van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*(5), 427–449. https://doi.org/10.1016/j.intell.2006.09.001

*van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology, 31*(1), 70–87. https://doi.org/10.1111/j.2044-835X.2012.02079.x

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

van Ravenzwaaij, D., & Oberauer, K. (2009). How to use the diffusion model: Parameter recovery of three methods: EZ, fast-dm, and DMAT. *Journal of Mathematical Psychology, 53*(6), 463–473. https://doi.org/10.1016/j.jmp.2009.09.004

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods, 49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

*Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences on executive processes on instrumental activities of daily living. *Psychology and Aging, 25*(2), 343–355. https://doi.org/10.1037/a0017729

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences*, *13*(6), 252–257. https://doi.org/10.1016/j.tics.2009.02.007

*von Bastian, C. C., & Druey, M. D. (2017). Shifting between mental sets: An individual differences approach to commonalities and differences of task switching components. *Journal of Experimental Psychology: General*, *146*(9), 1266–1285. https://doi.org/10.1037/xge0000333

von Bastian, C. C., Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods, 45*(1), 108–115. https://doi.org/10.3758/s13428-012-0224-y

*von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, *69*(1), 36–58. https://doi.org/10.1016/j.jml.2013.02.002

*von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, *145*(2), 246–258. https://doi.org/10.1037/xge0000120

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*(1), 3–22. https://doi.org/10.3758/BF03194023

*Was, C. (2007). Further evidence that not all executive functions are equal. *Advances in Cognitive Psychology*, *3*(3), 399–407. https://doi.org/10.2478/v10053-008-0004-5

Weinstein, Y. (2018). Mind-wandering, how do I measure thee with probes? Let me count the ways. *Behavior Research Methods*, *50*(2), 642–661. https://doi.org/10.3758/s13428-017-0891-9

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

White, C. N., Servant, M., & Logan, G. D. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review*, *25*(1), 286–301. https://doi.org/10.3758/s13423-017-1271-2

Whitehead, P. S., Brewer, G. A., & Blais, C. (2019). Are cognitive control processes reliable? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(5), 765–778. https://doi.org/10.1037/xlm0000632

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9

*Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology*, *44*(2), 575–587. https://doi.org/10.1037/0012-1649.44.2.575

*Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A. C., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of Experimental Child Psychology*, *108*(3), 436–452. https://doi.org/10.1016/j.jecp.2010.08.008

Wiecki, T. V., Antoniades, C. A., Stevenson, A., Kennard, C., Borowsky, B., Owen, G., Leavitt, B., Roos, R., Durr, A., Tabrizi, S. J., & Frank, M. J. (2016). A computational cognitive biomarker for early-stage Huntington's disease. *PLOS ONE*, *11*(2), e0148409. https://doi.org/10.1371/journal.pone.0148409

Wiecki, T. V, Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*(14), 1–10. https://doi.org/10.3389/fninf.2013.00014

Wiemers, E. A., & Redick, T. S. (2019). The influence of thought probes on performance: Does the mind wander more if you ask it? *Psychonomic Bulletin and Review*, *26*(1), 367–373. https://doi.org/10.3758/s13423-018-1529-3

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*(433), 1–22. https://doi.org/10.3389/fpsyg.2013.00433

*Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017). The test-retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology*, *23*, 822–837. https://doi.org/10.1080/09297049.2016.1205009

Woltz, D. J., & Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Memory and Cognition*. https://doi.org/10.3758/BF03193587

*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

*Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PloS One*, *8*(10), e77770. https://doi.org/10.1371/journal.pone.0077770

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Young, S. E., Friedman, N. P., Miyake, A., Willcutt, E. G., Corley, R. P., Haberstick, B. C., & Hewitt, J. K. (2009). Behavioral disinhibition: Liability for externalizing spectrum disorders and its genetic and environmental relation to response inhibition across adolescence. *Journal of Abnormal Psychology*, *118*(1), 117–130. https://doi.org/10.1037/a0014657

Zelazo, P. D., & Müller, U. (2002). Executive Function in Typical and Atypical Development. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 445–460). Madden, MA: Blackwell Publishers. https://doi.org/10.1002/9780470996652.ch20

*Zhang, T., Mou, D., Wang, C., Tan, F., Jiang, Y., Lijun, Z., & Li, H. (2015). Dopamine and executive function: Increased spontaneous eye blink rates correlate with better set-shifting and inhibition, but poorer updating. *International Journal of Psychophysiology*, *96*(3), 155–161. https://doi.org/10.1016/j.ijpsycho.2015.04.010

Zhou, B., & Krott, A. (2016). Data trimming procedure can eliminate bilingual cognitive advantage. *Psychonomic Bulletin and Review*, *23*(4), 1221–1230. https://doi.org/10.3758/s13423-015-0981-6

Appendix

Table A1

Studies Included in the Survey

Study	Sample Age	Sample *N*
Adrian et al. (2019)	Younger and Older Adults	126
Adrover-Roig et al. (2012)	Older Adults	122
Agostino et al. (2010)	Children	155
Allom & Mullan (2014)	Younger Adults	115
Arán-Filippetti (2013)	Children	248
Bartholow et al. (2018)	Younger Adults	216
Benedek et al. (2014)	Younger Adults	240
Brocki & Tillman (2014)	Children	117
Brydges et al. (2012)	Children	215
Carlson et al. (2014)	Children	101
Chuderski (2014)	Younger Adults	243
Chuderski (2015)	Younger Adults	525
Chuderski et al. (2012), sample 1	Younger Adults	160
Chuderski et al. (2012), sample 2	Younger Adults	178
De Simoni & von Bastian (2018)	Younger Adults	185
Del Missier et al (2010)	Younger Adults	116
Engelhardt et al. (2015)	Children	486
Fleming et al. (2016)	Younger Adults	413
Fournier-Vicente et al. (2008)	Younger Adults	180
Friedman & Miyake (2004)	Younger Adults	220
Friedman et al. (2008)	Younger Adults	556
Gustavson, Panizzon, Franz, et al. (2018)	Younger Adults	1248
Guye & von Bastian (2017)	Older Adults	142
Hedden & Yoon (2006)	Older Adults	121
Hedge, Powell, & Sumner (2018)	Younger Adults	103
Himi et al. (2019)	Younger Adults	202
Holochwost et al. (2017)	Children	207
Hull et al. (2008)	Older Adults	100
Ito et al. (2015)	Younger Adults	477
Kane et al. (2016)	Younger Adults	471
Khng & Lee (2009)	Children	157
Khng & Lee (2014)	Children	178
Klauer et al. (2010), Sample 1	Younger Adults	128
Klauer et al. (2010), Sample 2	Younger Adults	118
Lambek & Shevlin (2011), Sample 1	Children	164
Lambek & Shevlin (2011), Sample 2	Children	75
Lee, Ng, et al. (2012)	Children	155
Lehto et al. (2003)	Children	108
Lerner & Lonigan (2014)	Children	289
Martins et al. (2018)	Younger Adults	764

Study	Sample Age	Sample *N*
Masten et al. (2012)	Children	138
McAuley & White (2011), Sample 1	Children	36
McAuley & White (2011), Sample 2	Children	36
McAuley & White (2011), Sample 3	Children	38
McAuley & White (2011), Sample 4	Younger Adults	37
McCabe et al. (2010)	Younger Adults	206
McVay & Kane (2012)	Younger Adults	226
Miller et al. (2012)	Children	116
Miyake et al. (2000)	Younger Adults	137
Monette et al. (2015)	Children	265
Okada & Slevc (2018)	Younger Adults	150
Oschwald et al. (2018)	Younger Adults	99
Paap & Greenberg (2013), Sample 1	Younger Adults	86
Paap & Greenberg (2013), Sample 2	Younger Adults	107
Pettigrew & Martin (2014)	Younger and Older Adults	162
Redick et al. (2016)	Younger Adults	534
Rey-Mermet et al. (2018)	Younger and Older Adults	232
Ropovik (2014)	Children	96
Schweizer & Moosbrogger (2004)	Younger Adults	120
Shipstead et al. (2014)	Younger Adults	215
Shipstead et al. (2015), Sample 2	Younger Adults	573
Slevc et al. (2016)	Younger Adults	93
Stahl et al. (2014)	Younger Adults	190
Unsworth & McMillan (2014)	Younger Adults	241
Unsworth & Spillers (2010)	Younger Adults	181
Unsworth, Spillers, & Brewer (2009)	Younger Adults	155
Unsworth et al. (2010)	Younger Adults	151
Unsworth et al. (2012)	Younger Adults	165
Unsworth et al. (2014)	Younger Adults	171
Unsworth, Miller, et al. (2009)	Younger Adults	138
van der Sluis et al. (2007)	Children	171
van der Ven et al. (2013)	Children	208
Vaughan & Giovanello (2010)	Older Adults	95
von Bastian & Druey (2017)	Younger Adults	119
von Bastian & Oberauer (2013)	Younger Adults	121
von Bastian et al. (2016)	Younger Adults	118
Was (2007)	Younger Adults	180
Wiebe et al. (2008)	Children	187
Wiebe et al. (2011)	Children	219
Willoughby et al. (2017)	Children	115
Wongupparaj et al. (2015)	Younger Adults	110
Xu et al. (2013), Sample 1	Children	140
Xu et al. (2013), Sample 2	Children	165
Xu et al. (2013), Sample 3	Children	152
Zhang et al. (2015)	Younger Adults	61

Note. Mean sample age < 18 years was categorized as children (and adolescents), mean sample age < 60 as younger (and middle-aged) adults, and mean sample age > 60 as older adults. The sample *N* is the average *N* across the measures included in the survey.