Journal of Experimental Psychology:
Learning, Memory, and Cognition
2018, Vol. 44, No. 4, 501–526

0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000450

Should We Stop Thinking About Inhibition? Searching for Individual and Age Differences in Inhibition Ability

Alodie Rey-Mermet and Miriam Gade
University of Zurich and
Catholic University of Eichstätt-Ingolstadt

Klaus Oberauer
University of Zurich

Inhibition is often conceptualized as a unitary construct reflecting the ability to ignore and suppress irrelevant information. At the same time, it has been subdivided into inhibition of prepotent responses (i.e., the ability to stop dominant responses) and resistance to distracter interference (i.e., the ability to ignore distracting information). The present study investigated the unity and diversity of inhibition as a psychometric construct, and tested the hypothesis of an inhibition deficit in older age. We measured inhibition in young and old adults with 11 established laboratory tasks: antisaccade, stop-signal, color Stroop, number Stroop, arrow flanker, letter flanker, Simon, global-local, positive and negative compatibility tasks, and n-2 repetition costs in task switching. In both age groups, the inhibition measures from individual tasks had good reliabilities, but correlated only weakly among each other. Structural equation modeling identified a 2-factor model with factors for inhibition of prepotent responses and resistance to distracter interference. Older adults scored worse in the inhibition of prepotent response, but better in the resistance to distracter interference. However, the model had low explanatory power. Together, these findings call into question inhibition as a psychometric construct and the hypothesis of an inhibition deficit in older age.

Keywords: executive functions, cognitive control, adult aging, structural equation modeling, Bayesian hypothesis testing

Supplemental materials: http://dx.doi.org/10.1037/xlm0000450.supp

In many areas of psychology, cognitive inhibition has been used as an explanation to account for the human ability to select relevant information in face of interference. More precisely, cognitive inhibition has been described as the ability to ignore and to suppress irrelevant information (Logan, 1985). For at least 15 years, inhibition has been considered as a critical component of executive functions (i.e., the set of cognitive processes supervising and controlling ongoing thoughts and actions in working memory; Miyake et al., 2000). It was found to be separable from other components of executive functions (but see Friedman & Miyake, 2016; Miyake & Friedman, 2012), such as switching (i.e., the ability to shift the attention to other tasks or perceptual dimension; Kiesel et al., 2010; Vandierendonck, Liefooghe, & Verbruggen, 2010) and updating of working memory (i.e., the ability to replace the currently stored material in working memory by new information; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Ecker, Oberauer, & Lewandowsky, 2014; Kessler & Oberauer, 2014). At the same time, cognitive inhibition has been subdivided into several different forms (e.g., Friedman & Miyake, 2004; Nigg, 2000). The present study has two goals. First, we ask whether a psychometric construct of inhibition can be identified among a broad set of measures typically assumed to assess inhibition in the laboratory, and which subdivision of inhibition can be empirically supported. Second, we aimed to test the hypothesis of an inhibition deficit in old age (Hasher & Zacks, 1988).

The Unity and Diversity of Inhibition

Inhibition is often conceptualized as a unitary construct, which is used to explain group differences in cognition. For instance, the cognitive decline observed in adult aging has been proposed to result from a deficit in inhibition (Hasher & Zacks, 1988). According to this hypothesis, older adults show performance decrements in cognitive tasks compared with young adults because they are less able to inhibit irrelevant thoughts and actions in order to achieve their current goals. Similarly, the cognitive decline observed in depression has been explained by an inhibition deficit

This article was published Online First September 28, 2017.

Alodie Rey-Mermet and Miriam Gade, Department of Psychology, Cognitive Psychology Unit, University Research Priority Program (URPP) "Dynamics of Healthy Aging", University of Zurich, and Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt; Klaus Oberauer, Department of Psychology, Cognitive Psychology Unit, University of Zurich.

This work was supported by the University Research Priority Program (URPP) Dynamics of Healthy Aging, University of Zurich, and by a grant from the Velux Foundation. We thank Sofia Fäh, Seraina Kleussner, Jan Nussbaumer, Lena Pisarzewski, and Sarah Schoch for their help in data collection. The test battery, data, scripts analyses and results presented in this work can be found at https://osf.io/rygex/.

Correspondence concerning this article should be addressed to Alodie Rey-Mermet, Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt, Ostenstr. 27, D-85072 Eichstätt, Germany. E-mail: alodie.rey-mermet@ku.de

(Gotlib & Joormann, 2010). Efficient inhibition has also been argued to yield general cognitive advantages. For example, bilinguals have been claimed to perform better than monolinguals in nonverbal cognitive tasks because they constantly practice to inhibit the language currently not in use (Green, 1998; but see von Bastian, Souza, & Gade, 2015).

At the same time, various authors have proposed distinctions between inhibitory processes along several conceptual dimensions. The different taxonomies are summarized in Table 1. The older taxonomies are motivated by dissociations through neuropsychiatric and developmental disorders. The first fractionation of the inhibition construct through individual differences in healthy young adults has been presented by Friedman and Miyake (2004). They tested a factorial model distinguishing three forms of inhibition, namely the *inhibition of prepotent responses* (i.e., the ability to suppress dominant responses), the *resistance to distracter interference* (i.e., the ability to ignore distracting information), and the *resistance to proactive interference* (i.e., the ability to resist memory intrusions from information that was previously relevant to the task but has since become irrelevant). Each form of inhibition was measured with three tasks (see Appendix A for a presentation of each task). Using a structural equation modeling (SEM) approach, Friedman and Miyake (2004) found that inhibition of prepotent responses and resistance to distracter interference were very closely correlated. They combined them into a single form of inhibition, namely the *response-distracter inhibition*. This response-distracter inhibition factor was unrelated to the factor measuring resistance to proactive interference.

Using mostly the same tasks as Friedman and Miyake (2004; see Appendix A), Pettigrew and Martin (2014) confirmed the strong relation between the inhibition of prepotent responses and the resistance to distracter inference, and also considered both forms of inhibition as a single factor. In contrast, Stahl et al. (2014), who created new tasks to measure inhibition (see Appendix A), found empirical support for the separability of the three forms of inhibition (i.e., the inhibition of prepotent response, the resistance to distracter interference, and the resistance to proactive interference). Moreover, Stahl et al. (2014) emphasized the necessity of distinguishing within the inhibition of prepotent response between *behavioral inhibition* (i.e., the withholding or stopping of already initiated responses) and *response interference* (i.e., the suppression of competing response tendencies).

This overview shows that more fine-grained forms of inhibition have been empirically supported across the years (see Table 1). Specifically, resistance to proactive interference was found to be separable from the inhibition of prepotent response and the resistance to distracter interference; it remains, however, unclear whether the latter two are distinguishable from each other. The diversity of inhibition does not preclude the unity of a general inhibition construct. It is still possible that there is shared variance across the different inhibitory processes and the different tasks used to assess them.

Recent results, however, have called into question whether inhibition can be established as a psychometric construct. First, inhibition has been found to be not separable from the basic executive ability that accounts for the common variance among all measures of executive functions (Friedman & Miyake, 2016; Miyake & Friedman, 2012). Second, some studies did not identify an

Table 1
Overview of the Different Taxonomies About Inhibition

Dempster (1993)	Kornblum (1994)	Harnishfeger (1995)	Nigg (2000)	Friedman and Miyake (2004)*	Hasher et al. (2007)	Cyders and Coskunpinar (2011)	Pettigrew and Martin (2014)*	Stahl et al. (2014)*
Control of perceptual interference	Stimulus-stimulus interference	Resistance to interference	Interference control	Resistance to distracter interference	Access	Resistance to distracter interference	Resistance to distracter interference	Stimulus interference
Control of motor interference	Stimulus-response interference	Behavioral inhibition	Behavioral inhibition	Inhibition of prepotent response	Restraint	Inhibition of prepotent response	Inhibition of prepotent response	Response interference and behavioral inhibition
Control of verbal-linguistic interference		Cognitive inhibition	Cognitive inhibition	Resistance to proactive interference	Deletion	Resistance to proactive interference	Resistance to proactive interference	Proactive interference

* In these studies, inhibition was measured as a psychometric construct using a latent variable analysis.

inhibition factor (Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010, Exp. 1; Krumm et al., 2009; van der Sluis, de Jong, & van der Leij, 2007). Third, other studies did not find that inhibition measures have common variance (Keye, Wilhelm, Oberauer, & Ravenzwaaij, 2009; Wilhelm, Oberauer, & Hildebrandt, 2013). Fourth, in most studies that established an inhibition factor, that factor was often dominated by a single measure with a high loading on the factor, whereas other measures had very low loadings on the inhibition factor, implying that the factor does not represent much common variance among multiple measures of inhibition (see Appendix B for an overview of the factor loadings in previous studies). Therefore, it is not clear whether a general ability of cognitive inhibition even exists.

Inhibition and Aging

A large number of studies have been dedicated to test the hypothesis of an inhibition deficit in older age (see Rey-Mermet & Gade, 2017; Verhaeghen, 2011, 2014, for meta-analyses), and no consensus has been reached yet. Typically, these studies consisted of the comparison of two age groups (a younger sample and an older one) in a single experimental task intended to measure inhibition ability. The results so far are mixed. An age-related inhibition deficit was found in some studies (e.g., Andrés, Guerrini, Phillips, & Perfect, 2008; Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010), but not in others (Borella, Delaloye, Lecerf, Renaud, & de Ribaupierre, 2009; Hsieh & Fang, 2012; Kubo-Kawai & Kawai, 2010). More surprisingly, in a few studies, older adults were found to better inhibit irrelevant information than younger adults (Fernandez-Duque & Black, 2006; Madden & Gottlob, 1997; Sullivan, 1999).

So far, only Pettigrew and Martin (2014) have tested the inhibition deficit in older age at the latent variable level, and they have found empirical support for it. However, they did not differentiate between inhibition of prepotent response and resistance to distracter interference. Therefore, it remains unclear whether both forms of inhibition decline in older age.

The Present Study

The present study had two purposes. The first goal is to determine the psychometric structure of inhibition among a large set of tasks used to measure inhibition. In particular, we ask two questions: (a) Is there sufficient common variance between all (or most) measures of inhibition to establish a general inhibition factor? (b) Is there evidence for more specific factors reflecting distinct forms of inhibition (i.e., the inhibition of prepotent responses and the resistance to distractor interference)? The second purpose of the present study is to test the hypothesis of an inhibition deficit in older age (Hasher & Zacks, 1988) on the latent variables level—in case a psychometric construct of inhibition could be identified—or at least across a broad range of indicators of inhibition.

To this end, we measured inhibition with a comprehensive battery of 11 laboratory tasks. We examined several structural equation models to account for the correlational structure among these inhibition measures. Furthermore, we included two age groups—young adults (aged between 18 and 28) and older adults (aged between 65 and 75). The inclusion of the older age group provided a cognitively more heterogeneous sample than the typical college-student sample of young adults, reducing the risk that correlations between inhibition measures are underestimated due to a restriction of range. More importantly, the present design allowed us to test the hypothesis of an inhibition deficit in older age (Hasher & Zacks, 1988) on the basis of a broad and comprehensive set of inhibition measures.

How to Measure Inhibition?

In the present study, we opted for tasks for which there is broad (though not necessarily universal) agreement that inhibition plays a role for performance. Hence, we included most tasks of Friedman and Miyake (2004) because these tasks played a key role in defining inhibition as an individual-differences construct. These were the color Stroop, antisaccade, stop-signal, and letter flanker tasks (see Appendix A for a description).[1] Moreover, in order to extend the measure of inhibition as a psychometric construct, we also included tasks from experimental psychology that have been argued to show an effect of inhibition. These were the number Stroop, arrow flanker, Simon, global-local, positive and negative compatibility tasks as well as the task assessing n-2 repetition costs (see the bottom part of Appendix A). Following previous individual differences studies (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014), we assumed that color Stroop, antisaccade, and stop-signal tasks measure the inhibition of prepotent response, whereas the flanker tasks measure the resistance to distracter interference. For the other tasks, it is more speculative which form of inhibition they measure as, so far, they have not been used in individual-differences studies investigating the different facets of inhibition (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014). However, some evidence from experimental studies suggests that the inhibition measures from the Simon, global-local, and negative compatibility tasks should be associated to the inhibition of prepotent response, whereas the inhibition measures from the number Stroop and positive compatibility tasks, as well as the task assessing n-2 repetition costs, should be associated to the resistance to distracter interference (e.g., Gade, Schuch, Druey, & Koch, 2014; Hommel, 2011; Hübner & Malinowski, 2002; Salthouse & Meinz, 1995; Schlaghecken, Birak, & Maylor, 2012; Shilling, Chetwynd, & Rabbitt, 2002).

In most of these tasks (e.g., color and number Stroop, arrow and letter flanker, Simon, global-local, and positive compatibility tasks), inhibition is reflected in a difference score, such as the congruency effect. Specifically, the congruency effect is computed by subtracting reaction times (RTs) on *congruent* trials from RTs on *incongruent* trials. Congruent trials are those

[1] We deliberately excluded word naming and shape matching tasks in the present study because these tasks were used to assess two different effects: a distractor interference effect (i.e., the performance decrement occurring when a stimulus is presented with a distractor compared with when no distractor is presented) and the negative priming effect (i.e., the performance decrement of an item, which has previously been ignored). Therefore, it is not clear which effect is of primary interest in these tasks. Moreover, there is a long debate whether the negative priming effect measures inhibition or memory retrieval (Tipper, 2001).

without conflict between stimulus features or responses, such as the word "red" printed in "red" for the color Stroop task. Incongruent trials are those that involve some form of conflict between relevant and irrelevant information, such as the word "green" printed in red. Inhibition is assumed to be involved in incongruent trials to ignore or suppress the irrelevant information that creates the conflict (e.g., the meaning of the word for the color Stroop task). Computation of a difference score is necessary to isolate the contribution of inhibition from the contribution of other processes (e.g., word-reading or color-naming processes for the color Stroop task). A smaller congruency effect is thought to reflect stronger inhibition, because strong inhibition of the irrelevant stimulus feature or the incorrect response tendency should reduce the performance cost arising from the conflict in incongruent trials.

There is a long debate about whether congruent trials represent an appropriate baseline. On congruent trials, a facilitation process is assumed to occur because both the relevant and the irrelevant features activate the correct response (Glaser & Glaser, 1982; Tzelgov, Henik, & Berger, 1992). In order to avoid this facilitation process, *neutral* trials (i.e., trials with only one response-relevant feature, such as a row of five X's printed in red) are sometimes used as baseline. In the present study, we included both congruent and neutral trials. Thus, we were able to compute three difference scores: the *interference effect I* measured as the difference between incongruent and congruent trials, the *interference effect II* measured as the difference between incongruent and neutral trials, and the *facilitation effect* measured as the difference between congruent and neutral trials. The third set of difference scores is assumed to measure facilitation, while both interference effects are assumed to measure the strength of inhibition, with larger difference scores reflecting worse inhibitory abilities.

All these difference scores are based on RTs. This can be problematic regarding the older age group because older adults' RTs are slower than young adults' RTs by a constant proportion across many tasks and conditions (e.g., Cerella & Hale, 1994; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Verhaeghen & Cerella, 2002). For example, absolute RT differences between congruent and incongruent trials are larger for older adults simply because their RTs in both trial types are proportionally slowed relative to young adults. A fictitious example might help here. In a Stroop task, if young adults take 700 ms to perform a congruent trial and 1,000 ms to perform an incongruent trial, older adults would take—according to a proportional slowing—1,050 ms and 1,500 ms, respectively. Thus, the congruency effect would be 300 ms for young adults but 450 ms for older adults, although the underlying age effect (a proportional slowing of 1.5) is identical in congruent and incongruent trials. This could lead to the erroneous impression of a specific inhibition deficit in older age when in fact age differences are merely caused by a general proportional slowing of all RTs (Verhaeghen, 2011, 2014). To avoid this misinterpretation, we applied a natural logarithm (log) transformation to all RTs before computing the difference scores. In that way, the difference scores of log RTs reflect proportional difference scores. If RTs are increased in older age only due to general proportional slowing, then the differences of log RTs between congruent and incongruent trials should not differ between age groups.

Method

Participants

In total, 130 young adults and 159 older adults were tested. All participants reported Swiss German or German as native language, normal or corrected-to-normal vision, and no colorblindness. Participants were screened for dementia with the Mini-Mental State (MMS; Folstein, Folstein, & McHugh, 1975), and for depression with the Beck Depression Inventory II (BDI-II; Hautzinger, Keller, & Kühner, 2006). In addition to the BDI-II, older adults performed the Geriatric Depression Scale (GDS; Yesavage et al., 1982). Following the standard criteria, we excluded seven participants because of a score below 27 in the MMS, and 10 participants because of a score above 13 in the BDI-II, or above 5 in the GDS. In addition, we removed four participants because their age did not correspond to our criteria, three participants because they did not complete the second session, and two participants because they got the same tasks in both sessions.

Seven participants had missing data (six in the antisaccade tasks because they dropped out and one in the number Stroop task because s/he gave no response at all). Moreover, a visual screening of the data for the tasks in which RTs were the dependent measure revealed that some participants performed some blocks in some tasks at chance level. Thus, for each participant and each task, if the mean accuracy of a block was smaller than 2.5 standard deviations (*SD*) than the corresponding mean accuracy averaged across all participants, the block was removed from the dataset. If two blocks or more were removed, the task was removed from the dataset. If a task was missing, the participant was removed from the dataset. Following this procedure, 24 participants were removed.

The final sample consisted of 108 young adults (aged between 18 and 28) and 124 older adults (aged between 65 and 75). Demographic characteristics are summarized in Table 2 (top part). All participants performed the Consortium to Establish a Registry for Alzheimer's Disease-Plus (CERAD-Plus) test battery (Satzger et al., 2001) to assess their overall cognitive status, and the SF-36 questionnaire (Bullinger & Kirchberger, 1998) to assess their health status. Results on these background measures are presented in Table 2 (bottom part).

The study was carried out according to the guidelines of the ethics committee of the Faculty of Arts and Social Sciences at the University of Zurich, and all participants gave informed consent. At study completion, participants received CHF 45 (about $45) or course credits.

Material and Task Description

Tasks were programmed using Tscope5 (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006) and run on IBM compatible computers. Instructions equally stressed speed and accuracy of responding. For each task, unless specified otherwise, a feedback about the accuracy of the given response was given. This feedback was a smiling face (2.54 cm × 2.54 cm) after a correct response and a frowning face (2.54 cm × 2. 54 cm) after an error. The same pseudorandom trial sequence was administered to all participants. To reduce the impact of episodic memory as well as associative learning on performance, we applied three constraints. First, the

Table 2
Sample Characteristics and Background Measures

Measure	Young adults	Older adults	Differences
Demographics			
Age (years)	22.5 (2.6)	69.4 (2.9)	$t(230) = 128.85$, $p < .001$
Gender (female/male)	78/30	51/73	$\chi^2(1, N = 230) = 21.36$, $p < .001$
Education (years)	14.1 (3.4)	14.1 (3.4)	$t(227) = .11$, $p = .91$
Socioeconomical status[a]	5.3 (1.2)	5.0 (1.7)	$t(227) = -1.52$, $p = .13$
CERAD-Plus			
Boston Naming Test (accuracy)	.97 (.06)	.97 (.05)	$t(229) = -.11$, $p = .91$
Figure recall (accuracy)	.97 (.07)	.90 (.16)	$t(229) = -3.59$, $p < .001$
Word list recall (accuracy)	.89 (.08)	.76 (.11)	$t(229) = -9.92$, $p < .001$
Word list recognition (discriminability)[b]	1 (.01)	.99 (.03)	$t(229) = -3.55$, $p < .001$
Verbal fluency (number of words)	24.58 (5.99)	23.20 (5.65)	$t(229) = -1.80$, $p = .07$
S-words fluency (number of words)	17.36 (4.83)	14.14 (4.56)	$t(229) = -5.20$, $p < .001$
TMT score (in sec)[c]	2.02 (.80)	2.18 (.69)	$t(229) = 1.65$, $p = .10$
Health			
Physical index (standardized score)[d]	56.08 (5.04)	52.06 (6.34)	$t(229) = -5.28$, $p < .001$
Mental index (standardized score)[d]	47.01 (8.75)	56.72 (4.57)	$t(229) = 10.77$, $p < .001$

Note. Standard deviations (*SD*) are given in parentheses. For the characteristics "education (years)" and "socioeconomical status", data from one young adult and two older adults were missing because of a technical error. For the Consortium to Establish a Registry for Alzheimer's Disease-Plus (CERAD-Plus) test and the health questionnaires, data from one young adult was missing due to an experimenter's error.
[a] Socioeconomical status ranged from 1 (no or less than 9 school years) to 8 (PhD). [b] The discriminability score was computed as the difference $\{1 - [(10 - \text{Hits}) + (10 - \text{Correct Rejections})]/20\}$. [c] Trail Making Test (TMT) score was calculated as the time required to connect alternatively numbers and letters (TMT-B) divided by the time required to connect numbers only (TMT-A). [d] Higher scores indicate better health status.

trial sequence did not contain any complete stimulus repetitions (see description of the Simon task for an exception). Second, trial types (e.g., incongruent, congruent and neutral), response keys, presentation location, and individual stimulus exemplars, were balanced as far as possible. Third, stimulus material did not overlap across tasks. Furthermore, to minimize demands on procedural working memory, stimulus-response mappings were presented on the lower part of the screen during the complete trial sequence. The numbers of trials and of blocks for each task are summarized in Appendix C. The dependent variables used to measure inhibition are summarized in Table 3.

Congruency tasks. First, we described the congruency tasks (i.e., the color and number Stroop, the arrow and letter flanker, the Simon, the global-local, as well as the positive and negative compatibility tasks). For these tasks, incongruent, congruent and neutral trials were presented randomly and they occurred with equal frequency. First-order transitions between incongruent, congruent and neutral trials also occurred with equal frequency. For

Table 3
Dependent Measures for Each Task

Task	Interference effect I	Interference effect II	Facilitation effect
Color Stroop	ic—c	ic—n	n—c
Number Stroop	ic—c	ic—n	n—c
Arrow flanker	ic—c	ic—n	n—c
Letter flanker	ic—c	ic—n	n—c
Simon	ic—c	ic—n	n—c
Local	ic—c	ic—n	n—c
Positive compatibility	ic—c	ic—n	n—c
Negative compatibility[a]	–(c—ic)	–(c—n)	–(n—ic)
n-2 repetition cost[b]	–(n-2 repetition—n-2 switch)	—	—
Stop signal	SSRT computed with the integration method[c]	—	—
Antisaccade	error rates in antisaccades	—	—

Note. ic = incongruent; c = congruent; n = neutral; SSRT = stop signal reaction time.
[a] For the negative compatibility task, a larger RT cost for congruent relative to incongruent and neutral trials reflects stronger inhibition of the prime, which needs to be overcome on congruent trials (e.g., Eimer and Schlaghecken, 1998). We reversed the sign of these inhibition measures so that, in agreement with the other inhibition measures, larger values reflect worse inhibitory abilities. [b] For the n-2 repetition cost, we reversed the sign of this effect so that, in agreement with the other inhibition measures, larger values reflect worse inhibitory abilities. [c] According to the integration method (Schachar et al., 2007), go reaction times were rank ordered and the SSRT was computed by subtracting the mean SSD from the go RT that corresponded to the probability of inhibition.

each task, the three difference scores were computed as presented in Table 3, and larger interference effects reflect worse inhibitory abilities (except for the negative compatibility task, see below).

Color Stroop task. In the color Stroop task (MacLeod, 1991), participants were asked to indicate the color of color words while ignoring the meaning of the words. To this end, they were instructed to press with the index and middle fingers of the left and right hands the colored keys (namely the keys 6, 7, 8, and 9, on which a color disk—red, blue, green, or yellow, respectively—was stuck). The stimuli were the four German words red, blue, green, and yellow (i.e., "rot", "blau", "grün", and "gelb") as well as a row of five X's, displayed either in red, blue, green, or yellow. In congruent trials, the color corresponded to the word meaning (e.g., the word "red" printed in red). In incongruent trials, the color did not correspond to the word meaning (e.g., the word "red" printed in blue). In neutral trials, unrelated symbols were colored (e.g., a row of five X's displayed in blue).

During each trial (see Figure 1), a fixation cross was presented centrally for 500 ms. Then, the stimulus was presented centrally until response or 2,000 ms elapsed. After stimulus presentation, the accuracy feedback was presented in the middle of the screen for 500 ms, which was followed by a blank for 500 ms.

Number Stroop task. In the number Stroop task (Salthouse & Meinz, 1995), participants were asked to count the number of one to four centrally displayed characters while ignoring the numerical value of digit characters. To respond, they were instructed to press the keys 1, 2, 3 or 4 on the upper row of the keyboard with the index and middle fingers of the left and right hands, respectively. In congruent trials, the number of digits corresponded to the digits displayed (e.g., 22). In incongruent trials, the number of digits did not correspond to the digits displayed (e.g., 222). In neutral trials, unrelated symbols were displayed (e.g., $$). The trial sequence (see Figure 1) was the same as for the color Stroop task.

Arrow flanker task. In the arrow flanker task (e.g., Unsworth & Spillers, 2010), participants were asked to respond to the direction of the central arrow (left or right) while ignoring four flanking characters. To this end, they were instructed to press the keys *A* or *L* with the index fingers of the left or right hand, respectively. In congruent trials, the central arrow indicated the same direction of the flanking arrows (e.g., →→→→→). In incongruent trials, the central arrow indicated the opposite direction of the flanking arrows (e.g., ←←→←←). In neutral trials, lines with no arrowhead were displayed as flankers (e.g., ——→——). The trial sequence (see Figure 1) was the same as for the color Stroop task.

Letter flanker task. In the letter flanker task (e.g., Friedman & Miyake, 2004), participants were asked to decide whether a centrally presented target was a vowel (E or U) or consonant (S or H) while ignoring four flanker characters. To this end, participants were instructed to press the keys A or L with the index fingers of the left or right hand, respectively. In congruent trials, the central letter belonged to the same category as the flanker letters (e.g., UUEUU or UUUUU). In incongruent trials, the central letter belonged to the other category than the flanker letters (e.g., SSESS). In neutral trials, unrelated symbols (i.e., # or %) were displayed as flankers (e.g., ##U##). The trial sequence (see Figure 1) was the same as for the color Stroop task.

Simon task. In the Simon task (Hommel, 2011), participants were asked to decide whether a shape was a square or a circle while ignoring the position of the shape on the screen. The shape was presented on either the left or right side of the screen (i.e., 8.64 cm away from the middle of the screen). Participants were instructed to press the keys A or L with the index fingers of the left or right hand, respectively. In congruent trials, the left-right position of the response key corresponded to the position of the shape on the screen (e.g., a square presented on the left side and requiring to press the left key A). In incongruent trials, the response position did not correspond to the position of the shape on the screen (e.g., a square presented on the right side and requiring to press the left key A). In neutral trials, the shape was presented in the middle of the screen, a position associated to none of the response options. The trial sequence (see Figure 1) was the same as for the color Stroop task.

Because neutral trials are not usually employed in the Simon task (see Hommel, 2011), we presented neutrals trials in the first and last blocks of the task, and we intermixed congruent and incongruent trials in the middle blocks (see Appendix C). Thus, we were able to compare performance on incongruent and congruent trials under conditions matching those in previous research, but also to compute the same dependent measures as in the other tasks of the battery. However, presenting the stimuli in such a sandwich-like design reduced the number of stimulus exemplars per block type (two stimulus exemplars—that is, a square and a circle, each presented in the middle of the screen—for the blocks with neutral trials, and four stimulus exemplars—that is, a square or a circle presented either on the left or right side of the screen—for the blocks with congruent and incongruent trials). To make the responses in each trial unpredictable, we enabled complete stimulus repetitions for this task, but we excluded them in the data analysis.

Global-local task. In the global-local task (e.g., Hübner & Malinowski, 2002), participants saw element letters forming a large letter (4 cm × 4 cm), and they were asked to identify the letters (i.e., Y or V) by focusing either on the large letter (global task) or on the element letters (local task). Participants were instructed to press the keys Y or V with the index fingers of the left or right hand, respectively. In congruent trials, the large letter matched the identity of the small letters (e.g., a large Y made up from small Ys). In incongruent trials, the large letter did not match the identity of the small letters (e.g., a large Y made up from small Vs). In neutral trials, an unrelated letter (i.e., Z) was presented either as a large letter for the global task or as element letters for the local task. The trial sequence (see Figure 1) was the same as for the color Stroop task.

Global and local tasks were presented in a sandwich-like design (i.e., the global task occurred before and after the local task, see Appendix C). We opted for this design in order to ensure the global processing of the letter in the present study. However, congruency effects are predominantly found in the local task (e.g., Bruyer & Scailquin, 2000), and we therefore analyzed this task only.

Positive and negative compatibility tasks. In the positive and negative compatibility tasks (e.g., Eimer & Schlaghecken, 1998; Schlaghecken et al., 2012), participants first saw a prime (i.e., the double arrow "≪" or "≫", or the equal sign "=", 3.12 cm × 1.43 cm). In order to prevent conscious identification, the prime was followed by a mask after 33 ms. The mask consisted of the overlap of all prime exemplars. A target double arrow (i.e., ≪ or ≫) appeared either above or below the mask. Participants were asked to indicate the direction of the target arrow (i.e., left or right). To this end, they were instructed to press the key A and L

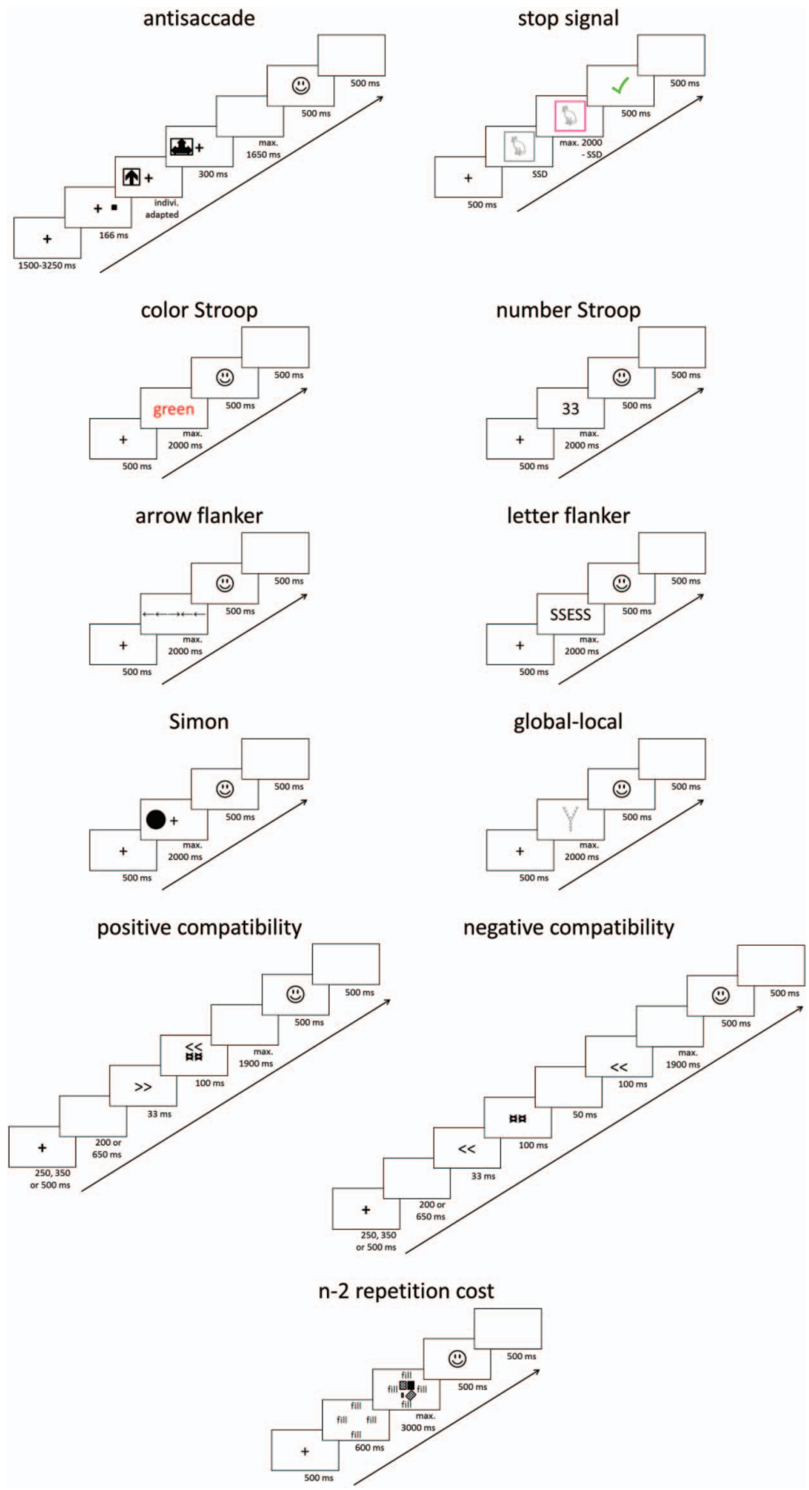

Figure 1. Example of one trial sequence for each task used in the present study. indi. = individually; SSD = stop signal delay. See the online article for the color version of this figure.

with the index fingers of the left and right hands, respectively. In congruent trials, the direction of the prime arrow corresponded to the direction of the target arrow. In incongruent trials, the direction of the prime arrow did not correspond to the direction of the target arrow. In neutral trials, an unrelated symbol (i.e., =) was presented as prime.

During each trial (see Figure 1), a fixation cross was presented centrally for either 250 ms, 350 ms, or 500 ms. This was followed by a blank lasting either 200 ms or 650 ms. Each time interval was selected randomly with equal probability. Then, the prime was presented centrally for 33 ms. For the positive compatibility task, the prime was immediately followed by the mask together with the target. Both the mask and the target were presented for 100 ms. The prime-target interval was thus 0 ms. For the negative compatibility task, the prime was followed after 33 ms by the mask, which remained for 100 ms. Then, a blank appeared for 50 ms, which was followed by the target for 100 ms. The prime-target interval was thus 150 ms. For both the positive and negative compatibility tasks, the target was followed by a blank, which was presented until response or 1,900 ms elapsed. Then, the accuracy feedback was presented in the middle of the screen for 500 ms, which was followed by a blank for 500 ms.

In contrast to the positive compatibility task, in the negative compatibility task, inhibition is assumed to incur a cost on congruent trials because the delay between the prime and target allows that the response induced by the prime is automatically suppressed, and then needs to be reactivated when responding to a congruent trial (Eimer & Schlaghecken, 1998; Schlaghecken et al., 2012). Therefore, in the negative compatibility task, RTs are slower on congruent than on incongruent trials, and the difference scores are reversed, with smaller interference effects reflecting poorer inhibitory abilities. Accordingly, the Interference Effect I was computed for the negative compatibility task by subtracting RTs on incongruent trials from RTs on congruent trials; the Interference Effect II was computed by subtracting RTs on neutral trials from RTs on congruent trials; and the facilitation effect was computed by subtracting RTs on neutral trials from RTs on incongruent trials (see Table 3).

Task assessing n-2 repetition costs. The n-2 repetition costs have been documented primarily in studies using the task-switch paradigm (Koch, Gade, Schuch, & Philipp, 2010; Mayr & Keele, 2000): When switching between three tasks (A, B, C), participants are slower when the current task repeats the task two trials back (as in sequence A-B-A) compared with when the current task has not been carried out during the last two trials (as in sequence C-B-A). This n-2 task-repetition cost is interpreted as reflecting the inhibition of task-relevant representations upon switching to another task, which is difficult to overcome when the just-inhibited representations must be reactivated two trials later. Hence, a larger n-2 repetition cost reflects stronger inhibition of task representations, which is harder to overcome when these representations have to be reactivated later.

In our version of the n-2 repetition paradigm (Mayr & Keele, 2000), participants worked through a series of deviant-detection trials (odd-item out task). On each trial, they saw four rectangles arranged in a 2 × 2 matrix (4 cm × 4 cm). Rectangles could differ in the dimensions orientation, fill, and size. For each dimension, there was a neutral state and two deviants. For orientation, the neutral state was upright while the deviants were 45-degree rotations to the left or right. For fill, the neutral state was checkered, whereas the deviant rectangles were unfilled (white) or filled (black). For size, the neutral state was an intermediate size (0.75 cm × 1. 35 cm), while the deviants were small (0.25 cm × 0.78 cm) or large (1.37 cm × 2.17 cm). In each trial, three rectangles had the neutral state and one rectangle differed in one dimension. The deviant dimension was indicated by a valid cue, that is, the word "orientation", "fill", or "size." To cue the dimension (but not the position on the screen), the word was presented in four locations, namely the left and right side of the screen (i.e., 7.94 cm away from the center) as well as the upper and lower margins of the screen (i.e., 5.29 cm away from the center). Cued dimensions occurred with equal frequency and were presented pseudorandomly so that no dimension was directly repeated in the following trial. Participants were asked to indicate which of the four rectangles was deviant in the cued dimension by pressing a spatially compatible key (i.e., the keys 5, 8, 6 and 9 in the number pad) with the index or middle finger of the left or right hand, respectively.

During each trial (see Figure 1), a fixation cross was presented centrally for 500 ms. This was followed by the cue for 600 ms. Then, in addition to the cue, the four rectangles were presented until a response or until 3,000 ms elapsed. After the four rectangles, the accuracy feedback was presented in the middle of the screen for 500 ms, which was followed by a blank for 500 ms.

In this task, the focus was on the sequences of relevant dimensions. Two types of sequences are of particular interest: n-2 repetitions (e.g., fill, size, fill) and n-2 switches (e.g., orientation, size, fill). These two types of sequences occurred with equal frequency and served to compute the dependent measure. Specifically, the n-2 repetition cost was computed by subtracting performance in n-2 dimension switch from performance in n-2 repetition (Mayr & Keele, 2000). In this task, smaller or even reversed difference scores reflect worse inhibitory abilities.

Stop-signal task. In the stop-signal task (Logan, 1994), participants were asked to classify pictures as representing living or nonliving objects by pressing the keys A and L with the index fingers of the left and right hands, respectively. Twenty-four pictures were selected from Cycowicz, Friedman, Rothstein, and Snodgrass (1997). These pictures were presented with a gray frame, which on stop trials turned pink after a variable interval (i.e., the stop signal delay [SSD]) following the onset of the picture. Participants were specifically instructed to withhold a response when the frame of the pictures turned pink (i.e., the stop-signal). In blocks intermixing go and stop trials (see Appendix C for a detailed description of the block structure), stop trials occurred randomly in 33% of the trials.

During each trial (see Figure 1), a fixation cross appeared centrally for 500 ms. Then, the target picture (9.22 cm × 9.05 cm) was presented centrally until a response was given or 2,000 ms elapsed. After the target picture, the accuracy feedback was presented centrally for 500 ms. In case of a go trial, this consisted of the smiling or frowning face depending on the correctness of the living/nonliving decision; in case of a stop-signal trial, this consisted of a green tick for correctly withheld responses or a red cross for incorrectly given responses. After the feedback, a blank appeared for 500 ms. In case of a stop-signal trial, the time between the onset of the picture and the stop-signal (i.e., the SSD) was adjusted according to a tracking procedure in order to achieve 50% accuracy in stop trials. Thus, the SSD was increased by 50 ms (to a maximum of 1,000 ms) after a correctly withheld response. After an erroneous response to a stop trial, the SSD was decreased by 50 ms (to a minimum of 50 ms). The initial SSD was set at 250 ms in the practice block as well as in the first and third experimental blocks.

In the stop-signal task, performance is modeled as a race between a go process, which is triggered by the presentation of the go stimulus, and a stop process, which is triggered by the presentation of the stop signal (Logan, 1994; Verbruggen & Logan, 2008). When the stop process finishes before the go process, the response is inhibited; when the go processes finishes before the stop process, the response is executed. Inhibition is measured by the latency of the stop process, referred to as the stop-signal RT (SSRT). Typically, longer SSRTs reflect worse inhibitory abilities. In the present study, the dependent measure was the SSRT. As our participants waited for the stop-signal delay and thus achieved more than the intended 50% accuracy in stop trials, we computed the SSRT with the integration method (Schachar et al., 2007). This method takes into account the proportion of successfully inhibited stop trials. That is, go RTs were rank ordered and the SSRT was computed by subtracting the mean SSD from the go RT at the percentile that corresponded to the probability of successful inhibition.

Antisaccade task. In the antisaccade task (adapted from Friedman & Miyake, 2004), participants were asked to indicate the direction of a small arrow (either left, up, or right). Arrows were presented briefly on the left or the right side of the screen and then masked. In prosaccade trials, the arrow was preceded by a black square on the same side of the screen, whereas on antisaccade trials, the square preceded the arrow on the opposite side of the screen. For antisaccade trials, participants were asked to inhibit a reflexive saccade toward the square and instead make a voluntary saccade to the opposite side in order to identify the briefly appearing target arrow.

During each trial (see Figure 1), a fixation cross appeared centrally for a variable amount of time (one of eight time intervals between 1,500 ms and 3,250 ms in 250-ms intervals, selected pseudorandomly with equal probability and no repetition). A black square (0.32 cm × 0.32 cm) then appeared on one side of the screen for 166 ms, followed by the target stimulus, that is, an arrow inside of an open rectangle (1.6 cm × 1.6 cm). Both the black square and the target were presented 8.64 cm away from the fixation cross. Size and eccentricity of the arrow were selected so that the arrow direction could not be identified while fixating the screen center. The target was followed by a mask (i.e., three arrows indicating left, up and down inside of an open rectangle of 4.80 cm × 3.75 cm), which was presented for 300 ms. Then, a blank was presented until a response or the deadline (i.e., 1,600 ms minus the target presentation time of the current trial). Participants responded by pressing the arrow key on the computer keypad that corresponded to the target with the index, middle and ring fingers of the right hand. The accuracy feedback appeared centrally for 500 ms, which was then followed by a blank for 500 ms.

Target presentation times (i.e., the time between onset of the arrow and onset of the mask) in prosaccade blocks was calibrated individually to achieve 80% correct responses using the weighted up-down method (Kaernbach, 1991): In case of a correct response, the target presentation time was decreased by 17 ms (to a minimum of 33 ms). In case of an error, the target presentation time was increased by 85 ms (to a maximum of 740 ms). For all participants, the initial target presentation was set at 150 ms. In antisaccade trials, target presentation time was computed individually as the median target presentation times of the prosaccades. With this procedure, individual differences in noninhibitory processes shared between prosaccade and antisaccade trials were compensated, such that accuracy in the antisaccade trials can be regarded as a relatively pure measure of those processes that are unique to antisaccade trials, of which arguably the inhibition of the prepotent tendency to saccade toward the black square is the one contributing most to performance. Thus, failure to inhibit a reflexive saccade should result in higher error rates. Accordingly, error rate on antisaccade trials was the dependent measure for the antisaccade task in the present study.

Procedure

Participants were tested in groups of up to five during two sessions of approximately 2–2.5 hr each (including breaks after every block and a longer break in the middle of each session). Both sessions were separated at least by 12 hr, and maximally by 1 week. In the first session, after informed consent was obtained, participants started with a questionnaire assessing demographics and then performed half of the tasks measuring inhibition. In the second session, participants performed the remaining tasks measuring inhibition, and then they completed the questionnaires, namely a further demographics questionnaire, the SF-36 (Bullinger & Kirchberger, 1998), the Beck Depression Inventory II (Hautzinger et al., 2006) and, in case of older adults, the Geriatric Depression Scale (Yesavage et al., 1982). The CERAD-Plus test battery was performed individually before or after one of the two sessions.

In one session, the tasks were ordered as follows: the number Stroop task, the letter Flanker task, the antisaccade task, the Simon task, as well as the positive and negative compatibility tasks. In the second session, the tasks were ordered as follows: the global-local task, the color Stroop task, the arrow Flanker task, the stop-signal task, and the task assessing n-2 repetition costs. To control for practice effects, the order of both sessions was counterbalanced across participants.

Data Preparation and Analysis

For most tasks, RTs were the primary dependent measure. In this case, RTs from errors and anticipatory responses (i.e., RTs smaller than 250 ms) were removed. It was not necessary to apply an upper limit when removing outlying RTs because a deadline was introduced experimentally in all tasks. Further, RTs falling 2.5 *SD* beyond an individual's mean per condition were removed.

Only one task—the antisaccade task—had errors as the primary dependent measure. In this case, we applied an arcsine square root transformation to the error rates in order to obtain approximately normally distributed data (Miyake et al., 2000).

In most inhibition measures (i.e., those for the color and number Stroop, the arrow and letter flanker, the Simon, the global-local, the positive compatibility, the stop-signal, and the antisaccade tasks), larger values reflect worse inhibitory abilities. Only for the negative compatibility task and the n-2 repetition cost, smaller values reflect worse inhibitory abilities. To predict positive correlations between all inhibition measures, we sign-reversed the inhibition measures for the negative compatibility task and the n-2 repetition cost (see Table 3).

To find multivariate outliers, we checked for multivariate normality using Mardia's (1970) kurtosis index. However, when multivariate outliers (i.e., cases with significant Mahalanobis's d2 values) were excluded, the pattern of results did not change. Therefore, no additional participants were excluded from the analyses reported below.

Results

Results are reported in three steps. First, we replicated previous analyses at the individual task level (i.e., analyses of variances [ANOVA] and *t* tests, respectively, for each task). Second, we examined the reliabilities as well as the correlational pattern of all tasks. Third, we investigated the relations between the tasks at the latent level using SEM.

Analyses at the Individual Level

For each task, descriptive and inferential statistics are presented in Appendixes D, E, and F. Overall, the results showed significant interference and facilitation effects for most tasks. Regarding the impact of age, the results showed no age group difference for the interference effects in some tasks (i.e., for the local task and the positive compatibility task as well as for the n-2 repetition costs). In some other tasks (i.e., the antisaccade, the stop-signal, the color Stroop, the Simon, and the negative compatibility tasks), older adults showed larger interference effects than young adults. This indicates worse inhibitory abilities for older adults, which is in line with an inhibition deficit (Hasher & Zacks, 1988). More surprisingly, smaller interference effects were found for older than for younger adults in three tasks (i.e., the number Stroop, the arrow and letter flanker tasks). This means that in these three tasks older adults scored better than young adults on indicators of inhibition, which is in line with a few studies (Fernandez-Duque & Black, 2006; Madden & Gottlob, 1997). Together, the present results replicate the diversity of the previous findings on the level of individual tasks (e.g., Andrés et al., 2008; Kramer et al., 1994; Madden & Gottlob, 1997; Verhaeghen, 2011).

Reliability and Correlations

As shown in Table 4 and Table S1 of the supplementary material, the reliability estimates of the facilitation effects across both age groups were lower than .50 for the congruency tasks (see the Simon and negative compatibility tasks for some exceptions). In contrast, the reliability estimates of both Interference Effects I and II were good or at least acceptable, ranging from .45 to .85 (see the local and letter flanker tasks for values at the lower limit). In most tasks, these estimates were slightly better for the Interference Effects I than for the Interference Effects II. For the antisaccade and stop-signal tasks, the reliability estimates were good, which replicates previous findings (Friedman & Miyake, 2004; Stahl et al., 2014). Only the reliability of the n-2 repetition costs was unacceptably low (.27). For this reason, we removed this measure in the structural equation modeling analysis.

The correlations are presented in Table 5 and Table S2 of the supplementary material. Most correlations were low (≤.20) and not significant. Moreover, the magnitudes of the correlations were similar across both age groups (see Table S2).

Structural Equation Modeling

Latent variable models were estimated in R using the lavaan package (Rosseel, 2012). Following previous research (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014), model fit was evaluated via multiple fit indices (Hu & Bentler, 1998, 1999): the χ^2 goodness-of-fit statistic, the Bentler's comparative fit index (CFI), the root mean square error of approximation (RMSEA), the standardized root-mean-square residual (SRMR), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). For the χ^2 statistic, a small, and ideally nonsignificant value indicates good fit. For CFI, values >.95 indicate good fit. RMSEA and SRMR were considered together so that RMSEA values <.06 and SRMR values <.08 indicate good fit. For the AIC and the BIC, smaller indices indicate better fit. To examine if one model fit the data reliably better than another, we performed two analyses. First, we conducted χ^2 difference ($\Delta\chi^2$) tests on nested models. If the more complex model (i.e., the model with more free parameters) yields a reduction in χ^2 that is significant given the loss of degrees of freedom, it is

Table 4
Descriptive Statistics of the Interference Effect I for Both Age Groups

Task	Mean	*SD*	Min	Max	Skew	Kurtosis	Reliability[b]
Color Stroop[a]	.05	.05	−.05	.17	.37	−.42	.84
Number Stroop	.08	.04	−.01	.24	.59	.51	.71
Arrow flanker[a]	.05	.03	−.07	.20	.17	2.83	.74
Letter flanker	.05	.03	−.03	.15	.06	−.10	.54
Simon	.06	.06	−.13	.20	−.31	.22	.73
Local	.05	.05	−.12	.21	.01	.88	.45
Positive compatibility	.05	.05	−.09	.18	−.17	−.28	.66
Negative compatibility	−.05	.07	−.24	.11	−.36	−.13	.85
n-2 repetition cost	−.04	.04	−.14	.09	.11	.07	.27
Stop signal	.67	.11	.39	1.02	−.25	.42	.72
Antisaccade	1.32	.56	.06	2.85	−.24	−.90	.97

Note. *SD* = standard deviation; min = minimum; max = maximum. Descriptive statistics for the Interference Effect II and the facilitation effect as well as for each age group separately are presented in the Table S1 in the supplementary material.
[a] For these dependent measures, a Yeo-Johnson transformation (Yeo & Johnson, 2000) was applied because skew and kurtosis were unacceptable (i.e., <−2 or >2). [b] Reliabilities were calculated by adjusting split-half correlations with the Spearman–Brown prophecy formula. Split-half correlations were computed between odd and even items, except in the stop signal task in which they were computed between the first two blocks and the last two blocks (because the stop signal delay was reset in the first and third blocks).

Table 5
Pearson Correlations Coefficients of the Interference Effect I for Both Age Groups

Task	Antisaccade	Stop signal	Color Stroop	Simon	Local	Neg. comp	Arrow flanker	Letter flanker	Number Stroop	Pos. comp.
Antisaccade	—									
Stop signal	.00 [−.12, .11]	—								
Color Stroop	**.13** [.01, .26]	**.21** [.10, .32]	—							
Simon	.09 [−.02, .21]	.01 [−.12, .13]	.08 [−.06, .22]	—						
Local	.08 [−.06, .22]	.05 [−.08, .17]	**.21** [.09, .33]	.11 [.00, .23]	—					
Neg. comp	.11 [−.02, .23]	**.16** [.02, .29]	**.44** [.35, .53]	.12 [−.02, .26]	.11 [−.02, .23]	—				
Arrow flanker	−.02 [−.14, .11]	−.02 [−.13, .10]	**−.17** [−.31, −.03]	.05 [−.08, .19]	.06 [−.06, .18]	**−.17** [−.29, −.04]	—			
Letter flanker	.07 [−.06, .20]	−.01 [−.15, .14]	−.09 [−.22, .04]	**.14** [.02, .26]	.09 [−.05, .23]	−.11 [−.24, .01]	**.14** [.00, .27]	—		
Number Stroop	−.08 [−.19, .04]	**−.15** [−.28, −.02]	**−.22** [−.33, −.10]	.03 [−.11, .16]	−.09 [−.21, .04]	**−.27** [−.40, −.15]	**.18** [.02, .33]	.12 [−.02, .25]	—	
Pos. comp.	−.01 [−.13, .11]	.05 [−.08, .17]	.05 [−.08, .17]	.11 [−.04, .26]	.04 [−.09, .16]	.09 [−.05, .23]	.08 [−.06, .23]	**.22** [.11, .33]	.06 [−.07, .19]	—
n-2 rep. cost	−.09 [−.21, .02]	−.03 [−.15, .09]	−.04 [−.16, .08]	−.02 [−.16, .12]	.02 [−.09, .14]	.01 [−.13, .15]	.00 [−.15, .14]	−.11 [−.23, .01]	−.06 [−.19, .08]	.09 [−.04, .23]

Note. Upper and lower confidence intervals from a bootstrapping procedure with 10,000 random samples are given in parentheses. Neg. comp. = negative compatibility; Pos. comp. = positive compatibility; n-2 rep. cost = n-2 repetition cost. Boldface type indicates $p < .05$. Pearson correlations coefficients for the Interference Effect II and the facilitation effect as well as for each age group separately are presented in the Table S2 in the supplementary material.

accepted as having better fit. Second, we performed a Bayesian hypothesis test using the BIC approximation (Wagenmakers, 2007). That is, we used the difference between the BIC for the null hypothesis (e.g., the single-factor model) and the BIC for the alternative hypothesis (e.g., the two-factor model) in order to compute a Bayes factor in favor of the null hypothesis (BF_{01}) and in favor the alternative hypothesis (BF_{10}). Following Raftery (1995) classification scheme, we considered a BF between 1 and 3 as weak evidence, between 3 and 20 as positive evidence, between 20 and 150 as strong evidence, and larger than 150 as very strong evidence. One advantage of using Bayesian hypothesis testing in addition to the more standard $\Delta\chi^2$ test was that we could not only reject the null hypothesis but also accept it.

For Models 1, 2, 3, and 4, we used as dependent measures the Interference Effects I because of their better reliability estimates. Moreover, we followed Pettigrew and Martin (2014) in estimating age effects by including age group as a dichotomous predictor on each latent variable. Model 1 fit the two-factor model in which the inhibition of prepotent response and resistance to distracter interference represent two distinct yet correlated latent variables, as depicted in Figure 2A. As shown in Table 6, this model provided a good fit to the data. Most tasks loaded significantly on their hypothesized latent variable, and there were significant age effects on each latent variable. Specifically, the age effect was positive for the inhibition of prepotent responses. This indicates larger interference effects for older than for young adults, which is in line with an inhibition deficit in older age (Hasher & Zacks, 1988). In contrast, the age effect was negative for the factor representing resistance to distracter interference. This indicates smaller interference effects for older adults than for young adults, which is in line with the results found at the individual task level in the present study, and in some previous studies (e.g., Fernandez-Duque & Black, 2006; Madden & Gottlob, 1997). The correlation between both latent variables—the inhibition of prepotent responses and the resistance to distracter interference—was high (.64) but did not approach the conventional level of significance ($p = .11$).

To investigate whether or not both latent variables correlate, we fit Model 2 with the constraint that both latent variables were assumed to be orthogonal (see Figure 2B). As shown in Table 6, this model also provided a good fit to the data. The difference in χ^2 between Models 1 and 2 approached significance ($p = .06$), suggesting a worse fit for Model 2 compared with Model 1, speaking for shared variance across both factors. However, neither the BF_{01} in favor of the null hypothesis (i.e., Model 2) nor the BF_{10} in favor of the alternative hypothesis (i.e., Model 1) were larger than 3, indicating ambiguous evidence about whether or not the two factors are correlated.

Model 3 fit the single-factor model, in which all tasks loaded on a single latent variable (see Figure 2C). This model also provided a good fit to the data (see Table 6). Critically, the difference in χ^2 between Models 1 and 3 was significant, indicating a worse fit for Model 3. Again, neither the BF_{01} in favor of the null hypothesis (i.e., Model 3) nor the BF_{10} in favor of the alternative hypothesis (i.e., Model 1) were larger than 3, indicating ambiguous evidence about whether two factors are needed to account for the data.

According to Gignac and Kretzschmar (2017), even if the correlated two-factor model (i.e., our Model 1) fit the data better in a $\Delta\chi^2$ test than the single-factor model (i.e., our Model 3), this is not sufficient to infer evidence for two related, but to some degree

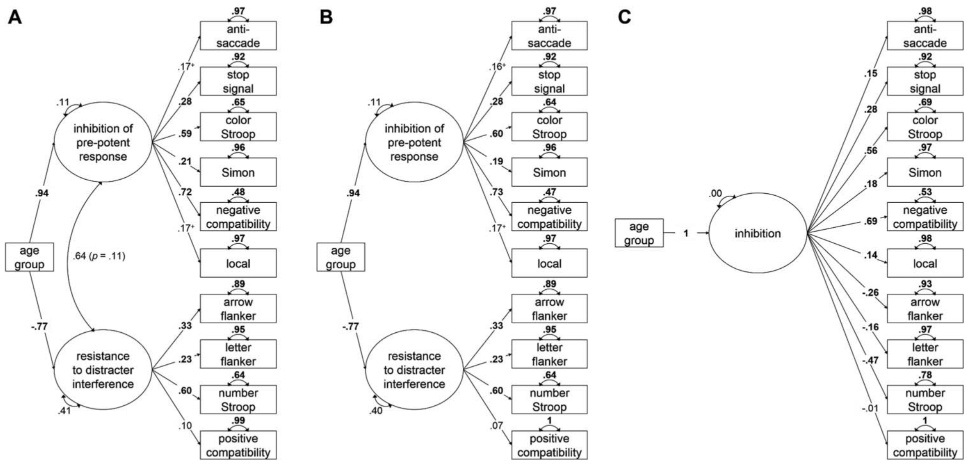

Figure 2. (A) Model 1: Two-factor model in which the inhibition of prepotent response and the resistance to distracter interference represent two correlated latent variables. (B) Model 2: Two-factor model in which the inhibition of prepotent response and resistance to distracter interference represent two distinct, uncorrelated latent variables. (C) Model 3: Single-factor model, in which all tasks loaded on a single latent variable. In all models, the effect of age was estimated on each latent variable as dichotomous predictor, and factor loadings were constrained to be positive. The numbers next to the straight, single-headed arrows are the standardized factor loadings (interpretable as standardized regression coefficients). The numbers adjacent to the curved, double-headed arrows next to each task are the error variances, attributable to idiosyncratic task requirements and measurement error. The numbers adjacent to the curved, double-headed arrows next to the latent variables are the correlations between the latent variables. For all parameters, boldface type indicates $p < .05$ and the symbol "+" indicates $p < .10$. The dependent measures were the interference effects I (with a logarithm transformation of the RTs and an arcsine square root transformation of the error rates).

distinct, factors. To do this, a bifactor model is necessary in which the *omega hierarchical* estimate of each specific latent variable (i.e., an indicator of unique latent variable strength) is larger than .20. We therefore fit Model 4 as a bifactor model so that the inhibition of prepotent response and resistance to distracter interference represent two uncorrelated specific latent variables. We also added a general inhibition factor on which all measures loaded and which was uncorrelated to the specific latent variables. This model provided a good fit to the data (see Table 6). The omega hierarchical estimate for the inhibition of prepotent response was very small at 0.07. In contrast, for the resistance to distracter interference, it was more appreciable with 0.41. Thus, the bifactor modeling approach suggests that inhibition ability is best described as a general inhibition factor together with one specific factor representing the resistance to distracter interference. However, the BF in favor of the bifactor model (i.e., BF_{01} in Table 6) was smaller than 1, indicating no evidence for this model despite its good fit statistics.

To summarize, when only the fit statistics and the $\Delta\chi^2$ tests are considered, the present results suggest that that Model 1 was better than Models 2 and 3 and thus the inhibition of prepotent response and the resistance to distracter interference represent correlated but

Table 6
Structural Equation Modeling With Age Group as a Dichotomous Predictor: Goodness-of-Fit Statistics and Model Comparison Results for the Interference Effect I

Model	χ^2	*df*	*p*	CFI	RMSEA [90% CI]	SRMR	AIC	BIC	$\Delta\chi^2$, *df*	BF01	BF10
1	*53.90*	*42*	*.103*	*.97*	*.03* [*0*, *.06*]	*.06*	*−6253.81*	*−6174.53*	—	—	—
2	*57.44*	*43*	*.069*	*.96*	*.04* [*0*, *.06*]	*.06*	*−6252.26*	*−6176.43*	*3.55, 1*$^+$	2.59	.39
3	*62.72*	*44*	*.033*	*.95*	*.04* [*.01*, *.07*]	*.06*	*−6248.99*	*−6176.61*	8.82, 2*	2.82	.35
4	*24.97*	*32*	*.807*	*1.00*	*0* [*0*, *.03*]	*.03*	*−6262.74*	*−6149.00*	—	<1	350863.06

Note. Model 1 = two-factor model; Model 2 = two-factor model with uncorrelated factors; Model 3 = single-factor model; Model 4 = bifactor model; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion. Change in χ^2 ($\Delta\chi^2$) was calculated relative to Model 1 (two-factor model). BF_{01} = Bayes factor in favor of the null hypothesis (i.e., the two-factor model with uncorrelated factors, the single-factor model, or the bifactor model). BF_{10} = Bayes factor in favor of the alternative hypothesis (i.e., the two-factor model with correlated factors, Model 1); this was computed as $1/BF_{01}$. Following Raftery (1995), we considered a BF between 1–3 as weak evidence, between 3–20 as positive evidence, between 20–150 as strong evidence, and larger than 150 as very strong evidence. Models with good fit statistics are presented in italic.
$^* p < .05$. $^+ = .06$.

distinct latent variables. However, Bayesian hypothesis testing shows that the data provide ambiguous evidence as to whether the two factors are uncorrelated, correlated, or even identical. Moreover, all four models show two problems. First, for each latent variable, one loading was substantially higher than the others (i.e., the negative compatibility task for the inhibition of prepotent responses, and the number Stroop task for the resistance to distracter interference; see Figure 2). Thus, each latent variable represents mainly the variance of one task. Second, the error variances were high for most tasks (see Figure 2), which indicates that the latent variables explained only a small fraction of the variance in the inhibition measures. Together, this suggests that irrespective of which model fits the data better, all models had low explanatory power.

For the sake of completeness, we computed multiple-groups SEM in which the model (i.e., Model 1, Model 2, Model 3, or Model 4) was fitted with different parameters in each age group, and we computed each model for each age group separately. We also estimated the models with the other dependent measures (i.e., the Interference Effects II and the facilitation effects), and we estimated bifactor models in which congruent or neutral trials of each task were forced to load on a baseline factor, and incongruent trials of each task were forced to load on an inhibition factor. Moreover, as previous individual differences studies about the diversity of inhibition (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014) used different trimming procedures, trial exclusions, and data transformations, we used several trimming procedures (e.g., including all participants or removing multivariate outliers) and data transformations (i.e., proportional scores, residuals, rate residuals—a combination of error rates and RTs—and raw RT differences; see Table 7 for a detailed description). We also reanalyzed the data by including immediate stimulus repetitions in the Simon task. The goodness-of-fit statistics of these model assessments and the parameter estimates can be found at https://osf.io/rygex/. Together, the results of these tests show that most models were either not identified, or provided a poor fit to the data. At best, their results were equivalent to those presented in this article. These additional analyses rule out the possibility that differences in trimming and data transformation procedures result in the discrepancy between the present and previous findings.

General Discussion

The present study had two purposes. The first goal was to determine the psychometric structure of cognitive inhibition: Can it be represented by a single factor, or by a structure with multiple factors reflecting different forms of inhibition? The second aim was to test whether inhibition in general, or a specific form of inhibition, declines in older age. To this end, we analyzed the data of a young and an older age group in 11 different tasks used to assess the *inhibition of prepotent responses* (i.e., the antisaccade, the stop-signal, the color Stroop, the Simon, the local, and the negative compatibility tasks) and the *resistance to distracter interference* (i.e., the arrow and letter flanker, the number Stroop, and the positive compatibility tasks as well as the n-2 repetition costs).

The Psychometric Structure of Inhibition

The inhibition measures showed reliability estimates that were in most cases as good as, or even better than those found in previous research (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014; see Table 8 for a comparison of the values found in previous research as well as in the present study). Only the n-2 repetition costs had low reliability. This in line with recent results in which n-2 repetition costs were also found to be unreliable (Kowalczyk & Grange, 2016; Pettigrew & Martin, 2016). Thus, n-2 repetition costs seem not appropriate as a measure of individual differences, probably because they represent a mixture of different cognitive processes, such as episodic retrieval and task interference (Gade, Souza, Druey, & Oberauer, 2017; Grange, Kowalczyk, & O'Loughlin, 2017). Except for the n-2 repetition costs, all remaining measures showed satisfactory reliabilities, indicating that whatever the commonly used laboratory measures of inhibition actually assess, they do it reliably when administered with the present task parameters.

Nevertheless, these measures hardly correlated among each other. A direct comparison to Friedman and Miyake (2004) revealed that our correlations between tasks were lower (see Table 8). A potential reason for these discrepancies might be that in the present study, we controlled for the impact of episodic memory as well as associative learning by excluding immediate stimulus repetitions and by avoiding stimulus overlap across tasks. Additionally, we minimized demands on procedural working memory by presenting the stimulus-response mapping at the bottom of the screen. Thus, it is possible that in the present study, we reduced the contribution of these memory processes to shared variance between inhibition measures, which reduced the correlations between them. In contrast, in previous studies (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014), these memory processes might have been common to many or all inhibition measures, and therefore have boosted the cross-task correlations. This may have occurred even when difference scores were computed as measures of inhibition, because these memory processes may contribute more to performance when responding to interference trials than when responding to baseline trials (see Verguts & Notebaert, 2008, 2009). Therefore, it is possible that memory-related variance was not subtracted out completely when difference scores were computed.

The competitive fitting of SEM models suggested a best-fitting model with two positively correlated inhibition factors, assumed to reflect the inhibition of prepotent response and the resistance to distracter interference, respectively. Thus, following previous research (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014), we could continue to assume that inhibition can be measured as a psychometric construct, and can be differentiated into two different forms.

Several observations, however, challenge this interpretation. First, a necessary consequence of the low correlations between inhibition measures was that in all models, most loadings were low, implying that most of the variance in the inhibition measures is not accounted for by the factors. Second, each factor was dominated by a single measure. Third, Bayesian hypothesis testing revealed ambiguous evidence as to whether the two factors are uncorrelated, correlated, or even identical. Fourth, further model estimations revealed that all models did not generalize across

Table 7
Description of the Different Procedures Used for Data Trimming, Data Transformation, and Analyses

Type	Description
	Data trimming
Removing data with 2.5 *SD* as criterion*	For each participant and each RT-based task, if the mean accuracy of a block was smaller than 2.5 *SD* than the corresponding mean accuracy averaged across all participants, the block was removed from the dataset. If two blocks or more were removed, the task was removed from the dataset. If a task was missing, the participant was removed from the dataset.
Removing data according to specific criteria	For each participant and each RT-based task, if the mean accuracy of a block was smaller than 75% for tasks with two response keys and 50% for tasks with four response keys, the block was removed from the dataset. If two blocks or more were removed, the task was removed from the dataset. If a task was missing, the participant was removed from the dataset.
Removing data by visual screening	For each participant and each RT-based task, if the mean accuracy of a block was considered as smaller than the mean accuracy of the other participants, the block was removed from the dataset. If two blocks or more were removed, the task was removed from the dataset. If a task was missing, the participant was removed from the dataset.
Procedure used by Friedman and Miyake (2004)	First, for each RT-based task, all RTs from errors and all RTs less than 200 ms were eliminated. Second, on the basis of visual inspections of the RT distributions, upper and lower criteria were established for each task, and any values exceeding those criteria were replaced with those values. Third, for each participant and each task, RTs farther than 3 *SD*s from the mean for each condition were replaced with values that were 3 *SD*s from the mean for that condition. Fourth, for each variable used in the models, all observations farther than 3 *SD*s from the group mean were replaced with values that were 3 *SD*s from the mean. Finally, to avoid eliminating participants with missing data, the missing data were replaced with the estimates obtained with multiple regression: the scores for each of these observations were predicted on the basis of these participants' scores on the other tasks used to assess inhibition.
Missing data were kept and the structural equation modeling was run with case-wise maximum likelihood.	
	Data transformation
log/arcsin*	This consisted of a logarithm transformation of the reaction times and an arcsine square root transformation of the error rates.
log/arcsin - shortened	Only half of the blocks were included in the analyses (because according to some accounts, measures of executive functions, such as inhibition, are most valid when they are novel; e.g., Friedman and Miyake, 2004; Rabbitt, 1997).
Proportional	These scores were computed as the differences presented in Table 3 divided by the corresponding baseline trials.
Residuals	Interference trials were regressed on baseline trials, and residuals were used as measures of inhibition.
Rate residuals	These scores—a combination of error rates and reaction times—were computed following the procedure proposed by Hughes, Linck, Bowles, Koeth, and Bunting (2014). That is, the four experimental blocks were divided into eight subsets of trials. Within each subset, the rate of correct responses per second was calculated for the interference and baseline trials by dividing the number of correct responses per trial type (interference vs. baseline) by the time taken to make all of the responses, whether accurate or inaccurate (summing the RTs for that trial type). For each subset, the interference trial rates were regressed on baseline trials rates. Then, the residuals were averaged across subsets to find each participant's final rate residual score.
Raw data	No data transformation was applied; only the RTs differences between interference and baseline trials were computed.
	Data analysis
SEM with age group as a dichotomous predictor*	
Multiple-groups SEM	SEM in which the model (Model 1, Model 2, Model 3, or Model 4) was fitted with different parameters in each age group
SEM within each age group	SEM in which the model (Model 1, Model 2, Model 3, or Model 4) was fitted separately for young and older adults

Note. SEM = structural equation modeling. To find multivariate outliers, we checked for multivariate normality using Mardia's (1970) kurtosis index. When we found multivariate outliers (i.e., cases with significant Mahalanobis's d2 values), we excluded them and we recomputed the analyses. The pattern of results did not change.
* Procedure presented in the present study. Results from the other procedures can be found at https://osf.io/rygex/.

alternative dependent variables, trimming procedures, data transformations, and multiple-groups SEM analyses. Whereas the first two problems are also present in previous research (see Appendix B), the present study is the first to use Bayesian hypothesis testing and to systematically investigate the robustness of the models across different data transformations and analyses, and we found it wanting.

In line with these observations, recent research has failed to find inhibition as a factor (Gignac & Kretzschmar, 2017; Klauer et al., 2010, Exp. 1; Krumm et al., 2009; van der Sluis et al., 2007) or to separate it factorially either from a common executive factor (Friedman et al., 2008; see also Brydges, Reid, Fox, & Anderson, 2012) or from general speed (Jewsbury, Bowden, & Strauss, 2016; see also van der Sluis et al., 2007). All these findings mandate us

Table 8
Comparison of Reliability Estimates, Correlations, and Factor Loadings for Comparable Measures of Inhibition Between Previous Studies (Friedman & Miyake, 2004; Pettigrew & Martin, 2014; Stahl et al., 2014) and the Present Study

					Two-factor model			Single-factor model
		Correlations			Inhibition of prepotent response	Resistance to distracter interference		Inhibition
Study/Task	Reliability	Antisaccade	Stop signal	Color Stroop	Loading	Loading	Correlations between both factors	Loading
Friedman and Miyake (2004)								
Antisaccade	.87	—			**.38**			**.42**
Stop signal	.72	**.16**	—		**.33**			**.36**
Color Stroop	.80	**.23**	**.15**	—	**.55**			**.45**
Letter flanker	.59	.04	**.15**	**.18**		**.42**		**.33**
—							**.66**	
Pettigrew and Martin (2014)								
Color Stroop	.91			—				**.26**
Letter flanker	.53			.03				−.09
—								
Stahl et al. (2014)								
Antisaccade	.47	—			**.39**			
Stop signal		.06	—		**.35**			
Color Stroop	.50			—				
Letter flanker	.26							
—							[a]	
The present study								
Antisaccade	.97	—			.17[b]			**.15**
Stop signal	.72	.00	—		**.28**			**.28**
Color Stroop	.84	**.13**	**.21**	—	**.59**			**.56**
Letter flanker	.54	.07	−.01	−.09		**.23**		**−.16**
—							.64[c]	

Note. For the previous studies, values were drawn from the corresponding tables and figures of each article. Boldface type indicates $p < .05$.
[a] Stahl et al. (2014) used slightly different tasks (see Appendix A) and divided the inhibition of prepotent response into two further forms (i.e., behavioral inhibition and response interference). This resulted in three correlations: the correlation between behavioral inhibition and resistance to distracter interference (.19), the correlation between behavioral inhibition and response interference (.12), and the correlation between resistance to distracter interference and response interference (**.53**). [b] $p = .06$. [c] $p = .11$.

to be more cautious in our assumptions. So far, the evidence suggests that the tasks used to assess inhibition do not measure a common underlying construct, but the highly task-specific ability to resolve the interference arising in that task (see also MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003; Miyake & Friedman, 2012). An inevitable implication of this conclusion is that studies using a single laboratory paradigm for assessing or investigating inhibition do not warrant generalization beyond the specific paradigm studied.

Inhibition and Aging

Regarding the relation between inhibition and aging, the SEM results revealed that aging was correlated with larger interference effects for the inhibition of prepotent responses factor but with smaller interference effects for the resistance to distracter interference factor. One may interpret these results by concluding that the inhibition of prepotent response, but not the resistance to distracter interference, decline in older age (Hasher & Zacks, 1988). This interpretation is based on the two-factor model. However, despite its good fit statistics, this model is of limited explanatory value, because its latent variables reflect only a small fraction of the systematic variance in the inhibition measures loading on them. This limitation of the SEM measurement model makes it difficult to interpret the impact of age on a hypothetical construct of inhibition as an ability generalizing across tasks, because there is insufficient evidence that such a general inhibition ability even exists. Moreover, the impact of age on the inhibition measures could itself have contributed to the distinction of the two factors in the model. It is possible that we found a two-factor model because aging has a positive impact on some measures loading on the resistance to distracter interference, and a negative impact on some of the measures loading on the inhibition of prepotent response.

At the individual task level, the present results replicate the diversity of previous findings (e.g., Andrés et al., 2008; Kramer et al., 1994; Madden & Gottlob, 1997; Verhaeghen, 2011, 2014) within a single study. An inhibition deficit was observed for older adults in some tasks (i.e., the color Stroop, the Simon, the negative compatibility, the stop-signal and the antisaccade tasks), but not in others (i.e., the local and the positive compatibility tasks, as well as the n-2 repetition cost). Moreover, in a few tasks (i.e., those used to assess the resistance to distracter interference, namely in the number Stroop, arrow and letter flanker tasks), older participants showed better inhibitory abilities than young adults. This "inhibition benefit" was already found in a few studies using the flanker tasks (Fernandez-Duque & Black, 2006; Madden & Gottlob, 1997) and was explained as the result of age differences in the attentional gradient over space. That is, as older adults have a steeper spatial attentional gradient than young adults, they can better exclude flankers in the flanker tasks. However, this explanation cannot account for the inhibition benefit observed in the number Stroop task. To our knowledge, the present study is the first to report such a finding, and thus this finding must be replicated before any substantive conclusion can be drawn. Nevertheless, taken together, the present findings challenge the hypothesis of a general inhibition deficit in older age (Hasher & Zacks, 1988).

Conclusion

To summarize, inhibition was assessed in the present study with a broad set of tasks typically used in experimental and individual-differences studies of inhibition. The data revealed good reliabilities of the measures used to assess inhibition but low correlations between them. Together, the present findings call into question the concept of inhibition as a psychometric construct. This is not to say that cognitive inhibition does not exist, or that individuals do not differ in its efficiency. However, individual differences in the ability to resolve interference from irrelevant stimuli or response tendencies appear to be highly task-specific. Therefore, we should perhaps stop thinking about inhibition as a general cognitive construct.

References

Andrés, P., Guerrini, C., Phillips, L. H., & Perfect, T. J. (2008). Differential effects of aging on executive and automatic inhibition. *Developmental Neuropsychology, 33,* 101–123. http://dx.doi.org/10.1080/87565640701884212

Borella, E., Delaloye, C., Lecerf, T., Renaud, O., & de Ribaupierre, A. (2009). Do age differences between young and older adults in inhibitory tasks depend on the degree of activation of information? *The European Journal of Cognitive Psychology, 21,* 445–472. http://dx.doi.org/10.1080/09541440802613997

Bruyer, R., & Scailquin, J.-C. (2000). The fate of global precedence with age. *Experimental Aging Research, 26,* 285–314. http://dx.doi.org/10.1080/036107300750015705

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

Bullinger, M. & Kirchberger, I. (1998). *SF-36 Fragebogen zum Gesundheitszustand* [SF-36 Health Survey]. Goettingen, Germany: Hogrefe.

Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life span. *Acta Psychologica, 86,* 109–197. http://dx.doi.org/10.1016/0001-6918(94)90002-7

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

Chuderski, A. (2015). The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure. *Personality and Individual Differences, 85,* 98–104. http://dx.doi.org/10.1016/j.paid.2015.04.046

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

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1,* 42–45. http://dx.doi.org/10.20982/tqmp.01.1.p042

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology, 65,* 171–237. http://dx.doi.org/10.1006/jecp.1996.2356

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review, 31,* 965–982. http://dx.doi.org/10.1016/j.cpr.2011.06.001

Dempster, F. N. (1993). Resistance to interference: developmental changes in a basic processing mechanism. In M. L. Howe & R. Pasnak (Eds.), *Emerging themes in cognitive development* (pp. 3–27). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4613-9220-0_1

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,* 170–189. http://dx.doi.org/10.1037/a0017891

Ecker, U. K. H., Oberauer, K., & Lewandowsky, S. (2014). Working memory updating involves item-specific removal. *Journal of Memory and Language, 74,* 1–15. http://dx.doi.org/10.1016/j.jml.2014.03.006

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1737–1747. http://dx.doi.org/10.1037/0096-1523.24.6.1737

Fernandez-Duque, D., & Black, S. E. (2006). Attentional networks in normal aging and Alzheimer's disease. *Neuropsychology, 20,* 133–143. http://dx.doi.org/10.1037/0894-4105.20.2.133

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,* 189–198. http://dx.doi.org/10.1016/0022-3956(75)90026-6

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

Friedman, N. P., & Miyake, A. (2016). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86,* 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17,* 172–179. http://dx.doi.org/10.1111/j.1467-9280.2006.01681.x

Friedman, N. P., Miyake, A., Young, S. E., Defries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General, 137,* 201–225. http://dx.doi.org/10.1037/0096-3445.137.2.201

Gade, M., Schuch, S., Druey, M. D., & Koch, I. (2014). Inhibitory control in task switching. In J. A. Grange & G. Houghton (Eds.), *Executive control and task switching* (pp. 137–159). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:osobl/9780199921959.003.0007

Gade, M., Souza, A. S., Druey, M. D., & Oberauer, K. (2017). Analogous selection processes in declarative and procedural working memory: N-2 list-repetition and task-repetition costs. *Memory & Cognition, 45,* 26–39. http://dx.doi.org/10.3758/s13421-016-0645-4

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence, 62,* 138–147. http://dx.doi.org/10.1016/j.intell.2017.04.001

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 875–894. http://dx.doi.org/10.1037/0096-1523.8.6.875

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology, 6,* 285–312. http://dx.doi.org/10.1146/annurev.clinpsy.121208.131305

Grange, J. A., Kowalczyk, A. W., & O'Loughlin, R. (2017). *The effect of episodic retrieval on inhibition in task switching*. Retrieved from https://papers.ssrn.com/abstract=2695998. http://dx.doi.org/10.1037/xhp0000411

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Language and Cognition, 1,* 67–81. http://dx.doi.org/10.1017/S1366728998000133

Harnishfeger, K. K. (1995). The development of cognitive inhibition. In C. J. Brainerd & F. N. Dempster (Eds.), *Interference and inhibition in cognition* (pp. 175–204). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-012208930-5/50007-6

Hasher, L., Lustig, C., & Zacks, R. T. (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.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/S0079-7421(08)60041-9

Hautzinger, M., Keller, F., & Kühner, C. (2006). *Beck Depression Inventory II (BDI 2)*. Frankfurt, Germany: Harcourt Test Service.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology, 20,* 511–528. http://dx.doi.org/10.1037/0894-4105.20.5.511

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica, 136,* 189–202. http://dx.doi.org/10.1016/j.actpsy.2010.04.011

Hsieh, S., & Fang, W. (2012). Elderly adults through compensatory responses can be just as capable as young adults in inhibiting the flanker influence. *Biological Psychology, 90,* 113–126. http://dx.doi.org/10.1016/j.biopsycho.2012.03.006

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453. http://dx.doi.org/10.1037/1082-989X.3.4.424

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

Hübner, R., & Malinowski, P. (2002). The effect of response competition on functional hemispheric asymmetries for global/local processing. *Perception & Psychophysics, 64,* 1290–1300. http://dx.doi.org/10.3758/BF03194772

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,* 702–721. http://dx.doi.org/10.3758/s13428-013-0411-5

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,* 508–522. http://dx.doi.org/10.1037/0894-4105.22.4.508

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell-Horn-Carroll model. *Journal of Experimental Psychology: General, 145,* 220–245. http://dx.doi.org/10.1037/xge0000119

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49,* 227–229. http://dx.doi.org/10.3758/BF03214307

Kessler, Y., & Oberauer, K. (2014). Working memory updating latency reflects the cost of switching between maintenance and updating modes of operation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 738–754. http://dx.doi.org/10.1037/a0035545

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

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136,* 849–874. http://dx.doi.org/10.1037/a0019842

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: Human Experimental Psychology, 63,* 595–619. http://dx.doi.org/10.1080/17470210903076826

Koch, I., Gade, M., Schuch, S., & Philipp, A. M. (2010). The role of inhibition in task switching: A review. *Psychonomic Bulletin & Review, 17,* 1–14. http://dx.doi.org/10.3758/PBR.17.1.1

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research, 56,* 130–135. http://dx.doi.org/10.1007/BF00419699

Kowalczyk, A. W., & Grange, J. A. (2016). Inhibition in task switching: The reliability of the n - 2 repetition cost. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*. Advance online publication. http://dx.doi.org/10.1080/17470218.2016.1239750

Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9,* 491–512. http://dx.doi.org/10.1037/0882-7974.9.4.491

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

Kubo-Kawai, N., & Kawai, N. (2010). Elimination of the enhanced Simon effect for older adults in a three-choice situation: Ageing and the Simon effect in a go/no-go Simon task. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 63,* 452–464. http://dx.doi.org/10.1080/17470210902990829

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica, 60,* 193–210. http://dx.doi.org/10.1016/0001-6918(85)90055-1

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language* (pp. 189–239). San Diego, CA: Academic Press.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: A comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics, 51,* 135–142. http://dx.doi.org/10.1016/j.archger.2009.09.040

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. http://dx.doi.org/10.1037/0033-2909.109.2.163

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. http://dx.doi.org/10.1016/S0079-7421(03)01014-4

Madden, D. J., & Gottlob, L. R. (1997). Adult age differences in strategic and dynamic components of focusing visual attention. *Neuropsychology, Development, and Cognition*. Section B, Aging, Neuropsychology and Cognition, *4,* 185–210. http://dx.doi.org/10.1080/13825589708256647

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57,* 519–530. http://dx.doi.org/10.1093/biomet/57.3.519

Mayr, U., & Keele, S. W. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology: General, 129,* 4–26. http://dx.doi.org/10.1037/0096-3445.129.1.4

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,* 302–320. http://dx.doi.org/10.1037/a0025250

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,* 8–14. http://dx.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. http://dx.doi.org/10.1006/cogp.1999.0734

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64. http://dx.doi.org/10.20982/tqmp.04.2.p061

Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review, 97,* 475–487. http://dx.doi.org/10.1037/0033-295X.97.4.475

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

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

Pettigrew, C., & Martin, R. C. (2016). The role of working memory capacity and interference resolution mechanisms in task switching. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 69,* 2431–2451. http://dx.doi.org/10.1080/17470218.2015.1121282

Rabbitt, P. (1997). *Methodology of frontal and executive function*. East Sussex, UK: Psychology Press.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. http://dx.doi.org/10.2307/271063

Rey-Mermet, A., & Gade, M. (2017). *Inhibition in aging: What is preserved? What declines? A meta-analysis*. Manuscript submitted for publication.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48,* 1–36. http://dx.doi.org/10.18637/jss.v048.i02

Salthouse, T. A., & Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. *The Journals of Gerontology Series B, Psychological Sciences and Social Sciences, 50,* 297–306. http://dx.doi.org/10.1093/geronb/50B.6.P297

Satzger, W., Hampel, H., Padberg, F., Bürger, K., Nolde, T., Ingrassia, G., & Engel, R. R. (2001). Zur praktischen Anwendung der CERAD-Testbatterie als neuropsychologisches Demenzscreening [Practical application of the CERAD test battery in screening for neuropsychological dementia]. *Der Nervenarzt, 72,* 196–203. http://dx.doi.org/10.1007/s001150050739

Schachar, R., Logan, G. D., Robaey, P., Chen, S., Ickowicz, A., & Barr, C. (2007). Restraint and cancellation: Multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology, 35,* 229–238. http://dx.doi.org/10.1007/s10802-006-9075-2

Schlaghecken, F., Birak, K. S., & Maylor, E. A. (2012). Age-related deficits in efficiency of low-level lateral inhibition. *Frontiers in Human Neuroscience, 6,* 102. http://dx.doi.org/10.3389/fnhum.2012.00102

Shilling, V. M., Chetwynd, A., & Rabbitt, P. M. A. (2002). Individual inconsistency across measures of inhibition: An investigation of the construct validity of inhibition in older adults. *Neuropsychologia, 40,* 605–619. http://dx.doi.org/10.1016/S0028-3932(01)00157-9

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. http://dx.doi.org/10.1016/j.jml.2014.01.004

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81,* 174–176. http://dx.doi.org/10.1037/h0027448

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

Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS windows platform. *Behavior Research Methods, 38,* 280–286. http://dx.doi.org/10.3758/BF03192779

Sullivan, M. P. (1999). The functional interaction of visual-perceptual and response mechanisms during selective attention in young adults, young-old adults, and old-old adults. *Perception & Psychophysics, 61,* 810–825. http://dx.doi.org/10.3758/BF03206899

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *The Quarterly Journal of Experimental Psychology Section A, 54,* 321–343. http://dx.doi.org/10.1080/713755969

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition, 20,* 727–735. http://dx.doi.org/10.3758/BF03202722

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

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71,* 1–26. http://dx.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. http://dx.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,* 194–200. http://dx.doi.org/10.1027/1614-0001.30.4.194

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. http://dx.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, 51,* 388–402.

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,* 427–449. http://dx.doi.org/10.1016/j.intell.2006.09.001

Vandierendonck, A., Liefooghe, B., & Verbruggen, F. (2010). Task switching: Interplay of reconfiguration and interference control. *Psychological Bulletin, 136,* 601–626. http://dx.doi.org/10.1037/a0019791

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences, 12,* 418–424. http://dx.doi.org/10.1016/j.tics.2008.07.005

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review, 115,* 518–525. http://dx.doi.org/10.1037/0033-295X.115.2.518

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

Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science, 20,* 174–180. http://dx.doi.org/10.1177/0963721411408772

Verhaeghen, P. (2014). Age-related differences in the speed of executive control. In *The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences*. Oxford, UK: Oxford University Press.

Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience and Biobehavioral Reviews, 26,* 849–857. http://dx.doi.org/10.1016/S0149-7634(02)00071-4

von Bastian, C. C., Souza, A. S., & Gade, M. (2015). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General, 145,* 246–258. http://dx.doi.org/10.1037/xge0000120

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

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

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87,* 954–959. http://dx.doi.org/10.1093/biomet/87.4.954

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17,* 37–49. http://dx.doi.org/10.1016/0022-3956(82)90033-4

Appendix A

Description of the Tasks and Their Associated Form of Inhibition

Task	Description	Associated form of inhibition
	Friedman and Miyake (2004)	
Antisaccade*	Participants were asked to identify a stimulus which is presented very briefly on the side opposite of a flashing cue. Thus, to perform this task successfully, participants had to suppress the reflexive saccade to the cue and perform a saccade in the opposite direction to identify the stimulus.	Inhibition of prepotent response
Stop signal*	Participants performed an ongoing task (e.g., a word categorization) unless the stop signal (i.e., a tone or a change in color frame) occurred. In this case, they had to withhold their responses. 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. The better participants achieve the 50% stopping criterion for longer delays of the stop signal relative to stimulus onset (reflected in shorter stop signal reaction times, SSRT).	Inhibition of prepotent response
Color Stroop*	Participants saw a color word printed in an incongruent color (i.e., the word "red" printed in green). They were asked to name the color of the word and to inhibit its meaning.	Inhibition of prepotent response
Letter flanker*	Participants were asked to identify the central letter in a letter array while ignoring the flanking letters that, on incongruent trials, were associated with the opposite response (i.e., HHSHH).	Resistance to distracter interference
Word naming	Participants were asked to name a green target word that was presented either alone or with a red distracter word.	Resistance to distracter interference
Shape naming	Participants indicated whether a white shape matched a green shape that was presented either alone or with a red distracter shape.	Resistance to distracter interference
Brown-Peterson	Participants learned and later freely recalled successive lists that were composed of words drawn from the same category.	Resistance to proactive interference
AB-AC-AD	After learning a list of cue—target word pairs to a criterion, participants learned a new list of targets that were paired with the same cues.	Resistance to proactive interference
Cued recall	Participants saw either one or two lists of four words each and had to retrieve the word on the most recent list that belongs to a cued category, ignoring any previous lists.	Resistance to proactive interference
	Pettigrew and Martin (2014)	
Color Stroop	See Friedman and Miyake (2004)	Response-distracter inhibition
Nonverbal Stroop	Participants were asked to indicate the direction of an arrow while ignoring the position of the arrow on the screen.	Response-distracter inhibition
Picture–word interference	Participants saw a picture with a superimposed distracter word and they were asked to respond by naming the picture (while ignoring the word).	Response-distracter inhibition
Letter flanker	See Friedman and Miyake (2004)	Response-distracter inhibition

(*Appendices continue*)

Appendix A (*continued*)

Task	Description	Associated form of inhibition
Recent negatives	Participants heard a list of three words followed by a probe word and indicated whether the probe word was in the previously heard list. On recent-negative trials, the probe matched a list element from a recent trial, but not the current trial. On nonrecent negative trials, the probe did not match words from the current or any recent trial. The RT difference between recent and nonrecent negative probes reflects the ability to inhibit the misleading familiarity of probes matching a list element in a recent trial.	Resistance to proactive interference
Brown-Peterson	See Friedman and Miyake (2004)	Resistance to proactive interference
Cued recall	See Friedman and Miyake (2004)	Resistance to proactive interference
	Stahl et al. (2014)	
Antisaccade	See Friedman and Miyake (2004)	Inhibition of prepotent response (behavioral inhibition)
Stop signal	See Friedman and Miyake (2004)	Inhibition of prepotent response (behavioral inhibition)
Go/no-go	Participants were asked to press the right button as soon as possible when a stimulus appeared on screen (go trials), except when an “X” was presented, in which case they should withhold a response (no-go trials).	Inhibition of prepotent response (behavioral inhibition)
Response priming	Prime and target stimuli were taken from two different sets (e.g., male or female names). Both sets served as primes and targets, and participants were asked to classify the target while ignoring the prime.	Inhibition of prepotent response (response interference)
Task switching	Participants saw digit-letter pairs, and they were asked to perform in alternation either a letter or a digit classification.	Inhibition of prepotent response (response interference)
Color Stroop	See Friedman and Miyake (2004)	Response-distracter inhibition
Letter flanker	See Friedman and Miyake (2004)	Response-distracter inhibition
Shape matching	See Friedman and Miyake (2004)	Resistance to distracter interference (stimulus interference)
Stroop matching	This task was similar to the shape matching task but used colors and color words as stimuli.	Resistance to distracter interference (stimulus interference)
Animal matching	This task was similar to the shape matching task but used animals and animal names as stimuli.	Resistance to distracter interference (stimulus interference)
Recent probes	In each trial, a set of six consonants was presented. After a retention interval, a probe stimulus appeared, and participants had to classify whether this consonant was presented as part of the current trial’s stimulus set. Then, the next trial started, following the same procedure. The probe stimulus could either be part of the current trial’s set of stimuli (match) or not (nonmatch). Orthogonally, it could have been presented in the directly preceding trial (recent) or not (nonrecent). The difficulty in recent lure trials arises from the requirement to give a nonmatch response despite a feeling of familiarity.	Resistance to proactive interference
Directed forgetting	Participants first memorized two memory sets and were then instructed to ignore one set while reporting on the basis of the other set.	Resistance to proactive interference
	Additional tasks typically used to measure inhibition	
Number Stroop*	Participants are asked to report the number of digits in a row while ignoring to read the digit value (e.g., Salthouse & Meinz, 1995)	Resistance to distracter interference[a]
Simon*	Participants are asked to respond to nonspatial features of a stimulus using manual responses. Stimuli are presented at either a left or a right position, which can be congruent or incongruent with the position (left or right) of the response keys on the keyboard (Hommel, 2011; Simon, 1969).	Inhibition of prepotent response
Global-local*	Participants are required to identify the local elements (i.e., the small “V”s building a “Y”) while suppressing the response induced by the global, prepotent letter (Hübner & Malinowski, 2002).	Inhibition of prepotent response
Positive compatibility*	Participants are asked to identify a target immediately after a masked prime is presented (Schlaghecken et al., 2012). Here, participants have to ignore the information induced by the prime.	Resistance to distracter interference
Negative compatibility*	This task is similar to the positive compatibility task, except that a delay (ca. 100 ms or 200 ms) is introduced between the prime and target (Eimer & Schlaghecken, 1998). During the delay, the prime induces a response, which is automatically suppressed, but needs to be reactivated when the target requires the same response as the prime. The negative-compatibility effect reflects the time cost of reactivating the primed response after it has been suppressed.	Inhibition of prepotent response

(*Appendices continue*)

Appendix A (*continued*)

Task	Description	Associated form of inhibition
Arrow flanker*	Participants are asked to identify the central arrow while ignoring the flanking arrows (e.g., Unsworth & Spillers, 2010).	Resistance to distracter interference
Task assessing n-2 repetition cost*	Participants are presented with arrays of four different stimuli (i.e., rectangles which vary on the feature dimensions orientation, fill or size). One of them differs in one dimension from the three other stimuli. The deviant stimulus dimension is indicated by a valid cue. Participants are asked to indicate which of the four stimuli is deviant in the cued dimension by pressing a spatially compatible key (Mayr & Keele, 2000). Here, participants have to ignore the dimensions which were target dimensions on an earlier trial but are currently irrelevant.	Resistance to distracter interference
48drop	Participants are presented with either four or eight colored squares and they are asked to remember as many colors as possible. After some time, one test colored square is presented at one of the original stimulus locations, and participants are asked to indicate if it is the same color as the original stimulus presented at that location (e.g., Unsworth et al., 2014).	Resistance to distracter interference
Disengage	In a first phase, participants are presented with a red square frame with a gap along with three more differently colored square frames. After some time, all square frames are masked with color patches, and participants are asked to report the direction of the gap. In a second phase ("attentional disengagement"), participants perform the same task, except that on one third of the trials, a colored square frame (distractor) is briefly presented prior to the target onset (e.g., Unsworth et al., 2014).	Resistance to distracter interference

[a] Previous research showed a low correlation between the interference effects of color and number Stroop tasks (Salthouse & Meinz, 1995; Shilling et al., 2002), suggesting that both tasks may not tap the same form of inhibition.
* Task used in the present study.

Appendix B

Factor Loadings for the Measures of Inhibition in Previous Studies

Study	Inhibition factor (as named in the study)	Measures used to assess inhibition	Loadings
Chuderski (2014)[a]	Antisaccade tasks	Spatial antisaccades	.96
		Letter antisaccades	.86
		Digit antisaccades	.83
Chuderski (2015)[a]	Attention control	Arrow antisaccade	.80
		Letter antisaccade	.87
Chuderski et al. (2012, Exp. 1)	Interference resolution	*Picture-word interference (RTs)*	*.76*
		Color Stroop (RTs)	.32
	Attention control	Picture-word interference (accuracy)	.42
		Color Stroop (accuracy)	.18
		Number Stroop (accuracy)	.31
		Antisaccade	*.57*
	Response inhibition	Stop signal	.36
		No-go	*.95*
Chuderski et al. (2012, Exp. 2)	Attention control	Picture-word interference (accuracy)	.27
		Number Stroop (accuracy)	.20
		Antisaccade	*.65*
Friedman and Miyake (2004)	Prepotent response inhibition	Antisaccade	.47
		Color Stroop	.45
		Stop signal	.36
		Letter flanker	.41
	Resistance to distracter interference	Word naming	.31
		Shape naming	.42
Friedman et al. (2006)	Inhibiting	Antisaccade	.61
		Stop signal	.64
		Color Stroop	.43

(*Appendices continue*)

Appendix B (*continued*)

Study	Inhibition factor (as named in the study)	Measures used to assess inhibition	Loadings
Friedman et al. (2008)	Inhibiting	Antisaccade	.44
		Stop signal	.53
		Color Stroop	.42
Hedden and Yoon (2006)	Prepotent response inhibition	*Antisaccade*	*.49*
		Color Stroop	.29
Hull et al. (2008)	Inhibition	*Color Stroop*	*.80*
		Nonverbal Stroop	.32
		Antisaccade	.19
Klauer et al. (2010, Exp. 2)	Inhibition	Letter flanker	.24
		Stroop matching	.35
		Simon	.46
		Antisaccade	*.61*
McVay and Kane (2012)	Attention control	Number Stroop	−.37
		Go/no-go (standard deviation for go trials)	−.46
		Go/no-go (signal-detection measure of performance)	.37
		Antisaccade	*.65*
Miyake et al. (2000)	Inhibition	*Antisaccade*	*.57*
		Stop signal	.33
		Color Stroop	.40
Pettigrew and Martin (2014)	Interference resolution	Recent negatives	.26
		Cued recall	.43
		Brown-Peterson	*.77*
		Letter flanker	−.09
		Picture-word interference	.20
		Nonverbal Stroop	.32
		Color Stroop	.26
Stahl et al. (2014)[b]	Stimulus interference	Stroop matching	.47
		Animal matching	.47
		Shape matching	.53
	Proactive interference	Recent probes	.31
		Directed forgetting 1	.38
		Directed forgetting 2	.36
	Response interference	Response priming	.39
		Task-switching between a number and a letter classification	.32
		Task-switching between a color and a shape classification	.32
	Behavioral inhibition	Stop signal	.35
		Antisaccade	.39
		Go/no-go	.42
Shipstead et al. (2014)	Attention control	*Antisaccade*	*.71*
		Arrow flanker	−.39
		Color Stroop	−.25
Unsworth (2015, Exp. 1)	Attention control (coefficient of variation)	Antisaccade	.50
		Arrow flanker	.74
		Color Stroop	.73
		Psychomotor vigilance	.29
Unsworth (2015, Exp. 2)	Attention control (coefficient of variation)	Antisaccade	.67
		Arrow flanker	.65
		Psychomotor vigilance	.57
Unsworth (2015, Exp. 3)	Attention control (coefficient of variation)	Antisaccade	.32
		Arrow flanker	*.71*
		Go/no-go	.34
		Color Stroop	.41
		Psychomotor vigilance	.54
Unsworth and McMillan (2014)	Attention control	Antisaccade	.54
		Arrow flanker	−.33
		Go/no-go (standard deviation for go trials)	−.45
		Go/no-go (accuracy)	.38
		Color Stroop	−.30
		Psychomotor vigilance	−.45

(*Appendices continue*)

Appendix B (*continued*)

Study	Inhibition factor (as named in the study)	Measures used to assess inhibition	Loadings
Unsworth, Miller, et al. (2009)	Response inhibition	*Antisaccade*	*.69*
		Arrow flanker	−.43
Unsworth et al. (2014)	Attention control	Disengage	−.47
		Antisaccade	*.62*
		48drop	−.40
Unsworth and Spillers (2010)	Attention control	Antisaccade	.63
		Arrow flanker	−.49
		Color Stroop	−.32
		Psychomotor vigilance	−.42
Unsworth, Spillers, et al. (2009)	Attention control	Antisaccade	.43
		Arrow flanker	−.47
		Psychomotor vigilance	−.43

Note. When the factor is dominated by a single measure, the measure with the high loading on the factor is presented in italics. As acknowledged by Unsworth and colleagues (Unsworth, 2015; Unsworth & McMillan, 2014; Unsworth & Spillers, 2010; Unsworth, Spillers, et al., 2009), the psychomotor vigilance task is rather a measure of sustained attention. In this task, participants are presented with a row of zeros on screen and after a variable amount of time the zeros begin to count up in 1-ms intervals from 0 ms. Participants are asked to press the space bar once the numbers start counting up.
[a] In these studies, only variations of antisaccade tasks were used as measures of inhibition, therefore enhancing the chance of finding a latent variable with loadings of similar range. [b] In Stahl et al. (2014), the unstandardized factor loadings were forced to be equal across the tasks of a given factor. According to the authors, this was done to ensure comparable contributions of each task (see p. 862). The fact that this unusual constraint was imposed suggests to us that it was necessary to avoid that factors were dominated by a single indicator with a uniquely high loading.

Appendix C

Block Order and Number of Trials per Block for Each Task

Block order	Trial type/task	Number of trials per block
Color Stroop		
1 practice block	Incongruent, neutral and congruent trials[a]	24
4 exp. blocks	Incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Number Stroop		
1 practice block	Incongruent, neutral and congruent trials[a]	24
4 exp. blocks	Incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Arrow flanker		
1 practice block	Incongruent, neutral and congruent trials[a]	24
4 exp. blocks	Incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Letter flanker		
1 practice block	Incongruent, neutral and congruent trials[a]	24
4 exp. blocks	Incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Simon[b]		
1 practice block	Neutral trials	4
1 exp. block	Neutral trials	50 (incl. 2 warm-up trials)
1 practice block	Incongruent and congruent trials[a]	20
1 exp. block	Incongruent and congruent trials[a]	50 (incl. 2 warm-up trials)
1 practice block	Neutral trials	4
1 exp. block	Neutral trials	50 (incl. 2 warm-up trials)
Global-lozcal[c]		
1 practice block	Global task: incongruent, neutral and congruent trials[a]	6
1 exp. block	Global task: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
1 practice block	Local task: incongruent, neutral and congruent trials[a]	12
2 exp. blocks	Local task: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
1 practice block	Global task: incongruent, neutral and congruent trials[a]	6
1 exp. block	Global task: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Positive and negative compatibility[c]		
1 practice block	Positive compatibility: incongruent, neutral and congruent trials[a]	6
1 exp. block	Positive compatibility: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)

(*Appendices continue*)

Appendix C (*continued*)

Block order	Trial type/task	Number of trials per block
1 practice block	Negative compatibility: incongruent, neutral and congruent trials[a]	12
2 exp. blocks	Negative compatibility: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
1 practice block	Positive compatibility: incongruent, neutral and congruent trials[a]	6
1 exp. block	Positive compatibility: incongruent, neutral and congruent trials[a]	74 (incl. 2 warm-up trials)
Task assessing n-2 repetition costs		
1 practice block	n-2 repetition trials and n-2 switch trials[a]	24
5 exp. blocks	n-2 repetition trials and n-2 switch trials[a]	74 (incl. 2 warm-up trials)
Stop signal[d]		
1 practice block	Go trials	24
1 exp. block	Go trials	50 (incl. 2 warm-up trials)
1 practice block	Go and stop trials (33% of the trials were stop trials)	16
4 exp. blocks	Go and stop trials (33% of the trials were stop trials)	74 (incl. 2 warm-up trials)
1 practice block	Go trials	8
1 exp. block	Go trials	50 (incl. 2 warm-up trials)
Antisaccade		
2 exp. blocks	Prosaccade	50 (incl. 10 warm-up trials)
1 practice block	Antisaccade	24
2 exp. blocks	Antisaccade	50 (incl. 2 warm-up trials)

Note. Exp. = experimental. In the analyses, only the (nonwarm-up) trials from the experimental blocks were analyzed.
[a] All trial types occurred with equal frequency. [b] Because neutral trials are not usually employed in the Simon task (Hommel, 2011), we presented neutrals trials at the beginning and the end of the task. Thus, we were able to compare performance on incongruent and congruent trials under conditions matching those in previous research, but also to compute the same dependent measures as in the other tasks of the battery. [c] To control for practice effects, we presented one condition in the middle and the other at the beginning and the end of the task. [d] By including blocks with only go trials at the beginning and the end of the task, we were able to control for practice effects and thus to compare performance on go trials in blocks with and without stop trials. This was important in order to determine whether participants strategically waited for the stop signal in blocks that included stop trials.

Appendix D

Analyses of Variances (ANOVAs) for the Log-Transformed Reaction Times (RTs) of the Congruency Tasks and n-2 Repetition Costs

For the congruency tasks (i.e., the color and number Stroop, the arrow and letter flanker, the Simon, the local, as well as the positive and negative compatibility tasks), a two-way ANOVA was conducted with congruency (incongruent, neutral, congruent) as a within-subject factor and age group (young, older) as a between-subjects factor. For the n-2 repetition cost, a two-way ANOVA was conducted with n-2 sequence (repetition, switch) as a within-subject factor and age group (young, older) as a between-subjects factor.

(*Appendices continue*)

Task/Effects	*df*	*F*	*p*	η_g^2
Color Stroop				
Age group	1, 230	438.07	<.001	.66
Congruency	1.28, 293.47	152.97	<.001	.40
Age Group × Congruency	1.28, 293.47	73.56	<.001	.24
Number Stroop				
Age group	1, 230	266.13	<.001	.54
Congruency	1.82, 417.73	717.99	<.001	.76
Age Group × Congruency	1.82, 417.73	41.46	<.001	.15
Arrow flanker				
Age group	1, 230	256.68	<.001	.53
Congruency	1.36, 313.94	458.57	<.001	.67
Age Group × Congruency	1.36, 313.94	8.52	.001	.04
Letter flanker				
Age group	1, 230	158.59	<.001	.41
Congruency	1.91, 440.39	404.49	<.001	.64
Age Group × Congruency	1.91, 440.39	4.10	.019	.02
Simon				
Age group	1, 230	157.79	<.001	.41
Congruency	1.74, 400.38	249.36	<.001	.52
Age Group × Congruency	1.74, 400.38	4.45	.016	.02
Local				
Age group	1, 230	203.60	<.001	.47
Congruency	1.95, 449.58	222.62	<.001	.49
Age Group × Congruency	1.95, 449.58	7.87	<.001	.03
Positive compatibility				
Age group	1, 230	353.26	<.001	.61
Congruency	1.89, 433.91	190.54	<.001	.45
Age Group × Congruency	1.89, 433.91	3.46	.035	.01
Negative compatibility				
Age group	1, 230	219.46	<.001	.49
Congruency	1.74, 400.89	187.06	<.001	.45
Age Group × Congruency	1.74, 400.89	138.95	<.001	.38
Task assessing n-2 repetition cost				
Age group	1, 230	171.62	<.001	.43
n-2 sequence	1, 230	182.96	<.001	.44
Age Group × n-2 Sequence	1, 230	.01	.931	.00

Note. Effect sizes are expressed as generalized η^2 values.

Appendix E

Post Hoc Contrasts Results for the Congruency Tasks and n-2 Repetition Costs

In order to disentangle the two-way interactions found in Appendix D, post hoc contrasts were conducted on the difference scores for each age group separately.

Task/Effects	Young age group			Older age group			Young vs. older		
	M (CI_w)	*t*	*p*	*M* (CI_w)	*t*	*p*	*M*	*t*	*p*
Color Stroop									
ic—c	.03 (.002)	4.39	<.001	.11 (.004)	20.05	<.001	.08	10.47	<.001
ic—n	.01 (.004)	1.26	.207	.09 (.01)	16.80	<.001	.08	10.54	<.001
n—c	.02 (.005)	3.12	.004	.02 (.01)	3.25	.004	−.001	−.07	.945
Number Stroop									
ic—c	.10 (.002)	31.85	<.001	.06 (.002)	20.79	<.001	−.04	−9.10	<.001
ic—n	.04 (.004)	12.36	<.001	.02 (.004)	7.04	<.001	−.02	−4.23	<.001
n—c	.06 (.005)	19.49	<.001	.04 (.004)	13.75	<.001	−.02	−4.87	<.001
Arrow flanker									
ic—c	.08 (.002)	18.56	<.001	.06 (.003)	15.09	<.001	−.02	−3.27	.002
ic—n	.08 (.003)	18.83	<.001	.08 (.01)	20.97	<.001	.003	.54	.587
n—c	−.001 (.004)	−.27	.791	−.02 (.01)	−5.88	<.001	−.02	−3.82	<.001

(*Appendices continue*)

Appendix E (*continued*)

Task/Effects	Young age group			Older age group			Young vs. older		
	M (CI$_w$)	*t*	*p*	*M* (CI$_w$)	*t*	*p*	*M*	*t*	*p*
Letter flanker									
ic—c	.06 (.002)	19.87	<.001	.05 (.002)	17.24	<.001	−.01	−2.76	.018
ic—n	.05 (.004)	15.92	<.001	.05 (.004)	15.99	<.001	−.003	−.73	.468
n—c	.01 (.004)	3.95	<.001	.004 (.004)	1.25	.213	−.01	−2.04	.084
Simon									
ic—c	.05 (.003)	7.17	<.001	.07 (.003)	10.93	<.001	.02	2.21	.027
ic—n	.10 (.01)	15.65	<.001	.10 (.01)	15.85	<.001	−.01	−.62	.533
n—c	−.06 (.01)	−8.48	<.001	−.03 (.01)	−4.93	<.001	.03	2.84	<.001
Local									
ic—c	.04 (.003)	9.29	<.001	.06 (.002)	13.00	<.001	.01	2.08	.076
ic—n	−.03 (.01)	−5.73	<.001	−.001 (.01)	−.33	.742	.03	3.97	<.001
n—c	.07 (.01)	15.02	<.001	.06 (.01)	13.33	<.001	−.01	−1.89	.076
Positive compatibility									
ic—c	.05 (.003)	12.43	<.001	.05 (.002)	13.14	<.001	−.001	−.12	.903
ic—n	.05 (.01)	12.17	<.001	.04 (.005)	9.62	<.001	−.01	−2.34	.060
n—c	.001 (.01)	.26	.798	.01 (.01)	3.52	<.001	.01	2.21	.060
Negative compatibility[a]									
−(c—ic)	−.10 (.003)	−24.29	<.001	−.01 (.002)	−1.60	.110	.09	16.66	<.001
−(c—n)	−.06 (.01)	−14.41	<.001	−.01 (.004)	−3.89	<.001	.04	7.88	<.001
−(n—ic)	−.04 (.01)	−9.88	<.001	.01 (.005)	2.29	.045	.05	8.78	<.001
Task assessing n-2 repetition cost[a]	−.04 (.02)	−9.19	<.001	−.04 (.02)	−9.98	<.001	−.0005	.09	.931

Note. *M* = mean; CI$_w$ = within-subject confidence intervals (see Cousineau, 2005; Morey, 2008). ic = incongruent; c = congruent; n = neutral. Larger values reflect worse inhibitory abilities.
[a] In order that larger values reflect worse inhibitory abilities for all tasks, we sign-reversed these measures.

Appendix F

t-Tests Results for the Log-Transformed Stop Signal Reaction Times (SSRT) of the Stop Signal Task and for the Arc Sine Transformed Error Rates of the Antisaccade Task

For the antisaccade and stop-signal tasks, a *t* test comparing young and older adults was performed.

Task/Effects	Young age group	Older age group	Young vs. older		
	M (CI$_w$)	*M* (CI$_w$)	*M*	*t*	*p*
Stop signal					
SSRT	.63 (.01)	.69 (.004)	.06	−4.34	<.001
Antisaccade task					
Antisaccade	1.23 (.03)	1.40 (.03)	.17	−2.35	.020
Prosaccade	.78 (.06)	.79 (.07)	.01	−1.01	.312

Note. *M* = mean; CI$_w$ = within-subject confidence intervals (see Cousineau, 2005; Morey, 2008). Larger values reflect worse inhibitory abilities.

Received February 22, 2017
Revision received May 31, 2017
Accepted June 1, 2017 ■