THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (3), 595–619

Psychology Press
Taylor & Francis Group

Understanding the role of executive control in the Implicit Association Test: Why flexible people have small IAT effects

Karl Christoph Klauer, Florian Schmitz, Sarah Teige-Mocigemba, and Andreas Voss
Albert-Ludwigs-Universität Freiburg, Freiburg, Germany

The goal of the present research was to investigate the role of three central-executive functions—switching of mental sets, inhibition of prepotent responses, and simultaneous storage and processing (i.e., working-memory capacity)—in accounting for method variance in the Implicit Association Test (IAT). In two studies, several IATs with unrelated contents were administered along with a battery of central-executive tasks, with multiple tasks tapping each of the above executive functions. Method variance was found to be related to the switching factor, but not to the inhibition factor. There was also evidence for a small independent contribution of the working-memory capacity factor. The findings constrain process accounts of the IAT, lending support to an account in terms of task-set switching, and they have consequences for applications.

Keywords: IAT; Social cognition; Working-memory; Task-set switching; Executive functions.

In recent years, a number of response-time paradigms have been developed to measure preferences and personality traits. One of the most famous paradigms of this kind is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT has received an enormous amount of attention in many fields of psychology, among them clinical, developmental, personality, consumer, and social psychology (see Lane, Banaji, Nosek, & Greenwald, 2007, for an overview). Part of the interest in the IAT derives from the hope that it will provide alternatives to traditional questionnaire-based measures, which rely on the respondents' self-reports, for measuring a wide variety of preferences and personality traits in these fields (see Teige-Mocigemba, Klauer, & Sherman, in press).

The IAT involves two tasks, (a) a concept task, in which exemplars (e.g., male faces and female faces) of two target categories (e.g., men and women) are to be classified according to their category membership, and (b) an attribute task, in which stimuli (e.g., wonderful, terrible) are to be classified into either of two attribute categories (e.g., as either positive or negative). In the two critical phases of the IAT procedure, both tasks

---

Correspondence should be addressed to Karl Christoph Klauer, Institute for Psychology, Social Psychology and Methodology, Albert-Ludwigs-Universität Freiburg, D-79085 Freiburg, Germany. E-mail: christoph.klauer@psychologie.uni-freiburg.de

The research reported in this paper was supported by Grants Kl 617/29–2 and Kl 614/31–1 from the Deutsche Forschungsgemeinschaft to the first author.

http://www.psypress.com/qjep

DOI:10.1080/17470210903076826

alternate and are mapped onto the same response keys. For example, in one critical phase, male faces and positive stimuli share one of the two response keys, and female faces and negative stimuli the other one, a response mapping that we refer to as the male/positive–female/negative mapping. In the other critical phase, female faces and positive stimuli share the same response key, and male faces and negative stimuli the other one, a response mapping that we refer to as the female/positive–male/negative mapping. The IAT effect is the performance difference between these two critical phases. According to Greenwald et al. (1998), direction and size of the IAT effect reflect the relative association strengths between the target concepts and attribute categories. Thus, participants who find it easier to categorize stimulus items under the male/positive–female/negative mapping than under the female/positive–male/negative mapping are assumed to possess stronger associations of positive evaluation with males than with females. Conversely, participants who find it easier to sort stimulus items under the female/positive–male/negative mapping than under the male/positive–female/negative mapping are assumed to possess stronger associations of positive evaluation with females than with males. Ease of sorting is quantified in different ways, most often in terms of speed of sorting, accuracy of sorting, or indices that combine speed and accuracy data (Greenwald, Nosek, & Banaji, 2003). The mapping that leads to better performance is often called "compatible", the other one "incompatible". Using different attribute and target categories, IATs have been applied to measure prejudice, consumer preferences, personality traits, self-esteem, and many other variables (for a review, see Nosek, Greenwald, & Banaji, 2006).

This article focuses on the contribution of executive control to the IAT effect. Executive control refers to a set of interrelated mental processes that modulate and constrain thought and behaviour to reach goal-relevant ends. Executive control is involved in the switching of task sets, in the inhibition of prepotent, impulsive, and automatic responses, and in monitoring and updating the contents of working memory, among others (Miyake et al., 2000).

As elaborated in the General Discussion, there are many instances of research involving the IAT in which experimental manipulations and independent and dependent variables covary with differences in the ability to exert executive control. In these applications, the IAT is, however, typically intended to measure constructs other than executive control; for example, it is interpreted as a measure of attitudes, prejudice, self-esteem, or personality traits, and observed effects and correlations are interpreted in terms of these constructs. Any covariation of the IAT effect with executive control then constitutes a systematic confound that can compromise the construct validity of conclusions based on observed IAT effects. In other words, differences in executive control immediately suggest alternative interpretations of reported findings in these cases that do not involve the constructs of interest such as attitudes, prejudice, and so forth. Examples of this are spelt out in the General Discussion.

It is obvious that executive control is involved in performing the IAT. For example, in the critical phases, participants are required to switch between the concept task and the attribute task from trial to trial (task-set switching). It is less obvious, however, whether executive control, and individual differences in it, also affect the IAT effect. Remember that the IAT effect is the difference in performance between the two critical phases that differ in how the task categories are mapped onto the response keys. Two accounts of the IAT, the Quad model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Sherman et al., 2008) and the account by task-set switching (Klauer & Mierke, 2005; Mierke & Klauer, 2001, 2003), suggest that high control ability leads to IAT effects of reduced sizes.

Conrey et al.'s (2005; Sherman et al., 2008) Quad model comprises several processes. One of these, termed association activation, is responsible for the direction of the IAT effect. Specifically, it is assumed that the stimulus items sometimes spontaneously elicit response tendencies that may converge or conflict with the task-appropriate response. For example, in the above-described IAT, a person with a positive attitude towards women and a negative attitude towards men will sometimes spontaneously activate the response tendency to respond "positive" when seeing a female face and the response tendency to respond "negative" for a male face, according to this process. These response tendencies facilitate performance under the female/positive–male/negative mapping in which response tendency and task-appropriate response coincide, but impair performance under the male/positive–female/negative mapping in which response tendency and task-appropriate response conflict. This leads to an overall IAT effect reflecting improved performance under the former response mapping relative to the latter response mapping. Two additional processes moderate the absolute size of the effect according to the Quad model, and both of these are related to executive control. The first, termed overcoming bias, is the ability to inhibit the spontaneous response tendency; the second, termed discriminability, is the ability to determine the task-appropriate, correct response, irrespective of any response tendencies activated by association activation. The first process is by definition related to the central-executive function to inhibit impulsive response tendencies (Gonsalkorale, von Hippel, Sherman, & Klauer, 2009). An involvement of inhibitory ability is thereby predicted for the IAT. The second process, discriminability, is argued to be related to attentional cognitive resources (Conrey et al., 2005). This suggests an involvement of working-memory capacity in the IAT via the link of attentional resources with working memory (Engle, 2002).

Mierke and Klauer (2001, 2003; Klauer & Mierke, 2005) proposed a different account of the IAT by task-set switching. In the critical phases of the IAT, participants are required to switch between attribute and concept task. Although there is some amount of debate concerning the nature of task sets and the process of switching between task sets (e.g., Allport, Styles, & Hsieh, 1994; Meiran, Chorev, & Sapir, 2000; Monsell, 1996; Rubinstein, Meyer, & Evans, 2001), most analyses concur that task-set switching involves changing a complex of cognitive settings required for performing a given task. This includes which attribute of the stimulus to attend to, which response mode and value to get ready, what classification of the relevant stimulus attribute to perform, how to map those classes to response values, with what degree of caution to set one's criterion for response, and so forth. The process of switching between task sets is associated with a performance cost. In the account of the IAT by task-set switching it is assumed that the critical phases of the IAT differ in how consistently participants switch between attribute and concept task. For example, a participant with a positive attitude toward women and a negative attitude toward men does not need to perform each and every task switch between attribute and concept task under the female/positive–male/negative mapping in order to respond fast and accurately: Fast and accurate responses can still be given even if the participant neglects to switch from attribute task to concept task and classifies all stimulus items under the attribute task—that is, according to their evaluative implications. Thus, if male and female faces are responded to on the basis of the participant's attitudes instead of on the basis of their category membership as men versus women, responses are still correct, and there is no need to perform costly task shifts. In contrast, under the male/positive–female/negative mapping, each and every task switch has to be performed to respond accurately. An involvement of the ability to switch between task sets is thereby predicted for the IAT.

One of the problems with identifying which central-executive function affects the IAT is that different central-executive functions as well as the cognitive marker tasks tapping these functions tend to overlap somewhat in process requirements. For example, switching between tasks sets involves the inhibition of recently activated tasks and response tendencies still elicited by them. Overlap in these and other component processes shared by cognitive tasks (such as encoding of stimuli, response selection, response execution, and so forth) explains why most cognitive tasks tend to correlate with each other, if often only weakly.

It is possible, however, to separate the unique contributions of the different central-executive functions through the use of a latent-variable approach (Miyake et al., 2000). This rests on two ideas. The first is to have several marker tasks for each executive function and to model each executive function by a latent variable that captures the variance shared by its marker tasks. This factors out narrow, task-specific components of the marker tasks and amplifies the contribution of components shared by each group of marker tasks in the respective latent variables. Most of the components shared by a group of marker tasks (e.g., marker tasks of inhibition) will reflect the executive function (inhibition) of which the tasks are marker tasks, but a residual overlap in other executive functions (e.g., working memory) is thereby not ruled out. The second idea is therefore to model the different latent variables as correlated to capture the remaining overlap between latent variables. In the statistical analysis, this then allows one to partial out contributions of components of executive control shared by the different latent variables. This makes it possible to assess the ability of each latent variable to uniquely account for the IAT effect when the remaining overlap between the different latent variables is controlled for. The account by task-set switching predicts that there is a unique contribution of task-set switching, but not of inhibition, to the IAT effect, whereas the Quad model predicts a unique contribution of inhibition, but not of task-set switching.

The evidence for a role of individual differences in central-executive functions in the IAT is, to date, largely indirect or unspecific. A number of authors found that different IATs correlated with each other although the concepts and attributes used in each IAT made it very unlikely that the different IATs measured related associations (Back, Schmukle, & Egloff, 2005; Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; McFarland & Crouch, 2002; Mierke & Klauer, 2003). Such findings demonstrate that some of the systematic variance in IAT effects is due to factors that affect different IATs similarly, irrespective of contents, reflecting so-called method variance, but they do not elucidate which factors underlie method variance. Two findings suggest more specifically that there is a role of inhibitory abilities in the IAT in line with the Quad model: Richeson and Shelton (2003) reported correlations between a race IAT and the Stroop task, a marker of inhibitory ability, and Payne (2005) reported a small, but significant correlation between a race IAT and the so-called antisaccade task, another marker of inhibitory ability. Because different executive functions and their markers all tend to correlate with each other (Miyake et al., 2000), small correlations between the IAT and a marker of inhibitory ability might, however, reflect variance that both share with other critical components of executive control as just explained. That is, these correlations do not demonstrate a *unique* contribution of inhibition in the IAT effect.

EXPERIMENT 1

In the present studies, IATs are placed in a broader array of executive tasks. Using a latent-variables approach, we measure three executive functions that have received empirical support as correlated, but separable, components of executive control (Miyake et al., 2000): (a) switching between tasks, (b) inhibition of dominant or prepotent responses, and (c) simultaneous storage and processing as captured by measures of working-memory capacity. We administer several tasks to tap each executive function so that the executive functions can be operationalized at the level of latent variables, factoring out narrow, task-specific contributions and taking into account that the executive functions themselves tend to correlate with each other.

Task-set switching refers to the array of processes involved in switching back and forth between multiple tasks, operations, or mental sets (Monsell, 1996). Inhibition refers to the ability to inhibit dominant, automatic, or prepotent responses intentionally when necessary. Miyake et al. (2000) investigated a third executive function, termed memory updating. Memory updating was closely linked to working-memory capacity as measured in tasks that require simultaneous storage and processing.

Some of the variance in the IAT effect reflects differences in the construct that is to be measured. For example, in the above-described IAT involving male and female faces, one's attitude towards women relative to men determines whether the IAT effect is positive or negative according to both Quad model and the account by task-set switching. Both accounts also suggest, however, that another source of systematic variance is given by executive-control functions that modulate the *absolute* size of the IAT effect (i.e., its distance from zero, ignoring the sign) for any IAT whatever it is intended to measure. To quantify this shared method variance, two unrelated IATs were administered in Experiment 1: a political-attitudes IAT and a flower–insect IAT. In the latent-variable analysis reported below, the two IATs define a latent variable capturing method variance that unrelated IATs share; for these analyses, both IATs were scored in terms of absolute size, ignoring sign.

Method

Participants

Participants were 128 volunteers (73 female, 55 male); mean age was 23.1 years ($SD = 5.8$) with a range from 17 years to 57 years. About half of the participants were University of Freiburg students with different majors; the other half comprised students from vocational schools, a few students from high schools, and nonstudents with different professions. Participants received 17.50 Euros as compensation for their participation.

Procedure

Participants were tested in individual sessions of about two hours. The tasks were administered in the following order: ratings of political attitudes, political-attitudes IAT, Stroop task, number–letter task, operation span, stop-signal task, flower–insect IAT, plus–minus task, antisaccade task, colour–size task, reading span.[1] Between stop-signal task and flower–insect IAT, there was a 10-min break. At the end of a session, participants completed a biographical questionnaire.

Switching tasks

Switching tasks were adapted from Miyake et al. (2000). They were the plus–minus task, the number–letter task, and the colour–size task.

Plus–minus task. The plus–minus task (Jersild, 1927) consisted of three lists of 30 numbers between 2 and 94, shown as columns on the computer screen. For the first list, participants were to add 3 to each number, entering the result on the number pad of the keyboard; for the second list, they were to subtract 3; for the third list, they were to alternate between adding and subtracting 3. After each response, the correct result of the operation just completed replaced the result entered by the participant on the screen. Participants received practice in entering numbers via the number pad in 10 trials preceding the plus–minus task. They were instructed to complete each list quickly and accurately. The cost of switching was calculated as the difference between the mean latencies of responses in the alternating list and the mean latencies of responses in the addition and subtraction lists.

Number–letter task. In a trial of the number–letter task (Rogers & Monsell, 1995), a number–letter pair (e.g., 4H) was presented in one of four quadrants on the computer screen. Participants were instructed to indicate whether the number was odd or even when the number–letter pair was presented in one of the upper two quadrants and whether the letter was a vowel or consonant when the pair was presented in one of the lower two quadrants. The pair was presented in the top two quadrants in the first block of 36 trials and in the bottom two quadrants for the second block of 36 trials. Whereas no task switching was therefore involved in these blocks, the number–letter pairs rotated clockwise around the four quadrants in the third

---

[1] Materials for all tasks in this paper can be obtained from the authors.

block of 136 trials, so that participants had to switch from odd–even discrimination to vowel–consonant discrimination on every second trial. The first 4 trials of the first and second blocks as well as the first 8 trials of the third block were practice trials. Responses were entered by pressing a left key ("A") or right key ("5"). The intertrial interval was 150 ms. The cost of switching was calculated as the difference between the mean latencies of responses in the third block and the mean latencies of responses in the first two blocks.

Colour–size task. The colour–size task was used as the third switching task. Stimuli were geometric shapes such as triangles, circles, or squares. They were coloured in red or blue; they were also either small (subtending between 15 and 25 mm) or large (subtending between 35 and 45 mm). Two discrimination tasks were to discriminate the colour (red vs. blue) and to discriminate the size (small vs. large). Which task was to be performed was determined by whether the shapes were completely filled with colour or whether only the boundaries of the shapes were coloured. In the first case, size was to be discriminated; in the second case, it was colour. Participants worked through three blocks. In the first block of 62 trials, all stimuli were filled; in the second block of 62 trials, only the boundaries were coloured. In the third block of 100 trials, both kinds of stimuli were mixed in a prerandomized order, and half of the trials required a task switch. This block was preceded by 24 practice trials. Responses were entered by pressing a left key ("A") or right key ("5" on the number pad). The intertrial interval was 500 ms. The cost of switching was calculated on the basis of the data from the third block as the difference between the average latencies for trials in which task set had to be switched and for trials in which task set was repeated.

Inhibition tasks

The antisaccade task, the stop-signal task, and the Stroop task were used to assess inhibition. Details of these tasks closely followed Miyake et al. (2000).

Antisaccade task. In a trial of the antisaccade task (Roberts, Hager, & Heron, 1994), a fixation point was first presented in the middle of the screen for a random amount of time (between 1,500 ms and 3,500 ms in 500-ms intervals). A visual cue, a small square subtending approximately 0.4° of visual angle, was then presented for 225 ms, shifted by 9° toward either the left or the right side of the screen. This was followed by the presentation of the target stimulus, subtending approximately 2°, shifted by the same distance from the middle as the initial visual cue, but in the opposite direction. The target was presented for 150 ms before being masked by a grey cross-hatching. The mask stayed on screen until the participant entered the response. The target consisted of an arrow inside an open square. Participants were to indicate the direction of the arrow (left, right, up) using the appropriate arrow buttons on the computer keyboard. Given that the target appeared for only 150 ms before being masked, participants were required to inhibit the reflexive response of looking at the initial visual cue because doing so would make it difficult to identify the direction of the arrow correctly. Intertrial interval was 1,500 ms. Participants worked through 30 practice trials, followed by a block of 93 trials, of which the first 3 were warm-up trials. The dependent variable was the percentage of correct responses.

Stop-signal task. The stop-signal task (Logan, 1994) consisted of two blocks of trials. The first block of 48 trials served to build up a prepotent categorization response. In each trial, participants were shown 1 of 24 words (e.g., dog, radio) and were to categorize it as animal or nonanimal as quickly as possible without making errors. The 12 animal words and the 12 nonanimal words were matched for length and frequency. Each trial began with a fixation point presented in the middle of the screen for 500 ms. This was followed by the target word, which remained on screen for 1,500 ms. In the case of a wrong response or no response after 1,500 ms, the word "Fehler" (error) appeared on the screen for 500 ms. In the second block of 192 trials, participants were required not to respond (i.e., to inhibit the categorization response) when they heard a tone (440 Hz, 100-ms duration) for 48 randomly selected stop-signal trials. The tone followed the target onset with a delay that was given by the participant's median latency in the first block minus 225 ms. On the other trials, participants were to perform the categorization task as before. Error feedback was the same as before, but on the 48 stop-signal trials, error feedback signalled that a categorization response had been entered. As recommended by Logan (1994), instructions emphasized that the participants should not slow down to wait for possible stop signals. The dependent variable for this task was the proportion of stop-signal trials for which the categorization response was successfully withheld.

Stroop task. The data from the Stroop task were lost due to a programming error. We therefore omit a description of its procedural details.

Working-memory capacity

Reading span (Daneman & Carpenter, 1980) and operation span (Turner & Engle, 1989) were used to operationalize working-memory capacity. The details of reading span and operation span followed Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000).

For each of the 15 trials of the reading span, a list of three to seven sentences was presented on the screen, one sentence after the other, for 3 seconds per sentence, followed by a 1-second pause before the onset of the next sentence. The sentences were all either trivially true or false, and participants were to rate each sentence as true or false during the 4-second interval by pressing one of two keys labelled true and false. In addition, the last word of each sentence was to be memorized. After a list of sentences, these words had to be written down in their order of presentation.

Operation span is a similar task based on numerical content. Instead of sentences, participants verified simple addition and subtraction equations with one- and two-digit numbers (e.g., $5 + 4 = 8$). The given results in the list of four to eight equations were one-digit numbers (e.g., 8), and they had to be remembered and written down in correct order after the end of the sequence. Presentation times were the same as those for reading span. For both span tasks, three lists were presented for each of the five list lengths.

List lengths were presented in ascending order from three to seven for reading span, and from four to eight for operation span. There were 2 practice lists for each span task with list lengths of two and three for reading span and operation span, respectively. For each span task, a span score was computed as the average number of items (i.e., words or numbers) recalled in correct position across the 15 lists. This score could vary between zero and five and between zero and six for reading span and operation span, respectively.

Implicit Association Tests

The format of the IATs followed the procedures introduced by Greenwald et al. (1998) with few exceptions. We administered a flower–insect IAT and a political-attitudes IAT. The IATs consisted of seven blocks of either 24 or 48 trials. In Table 1, the specifics of each block are summarized for the flower–insect IAT. Each block was preceded by additional warm-up trials using stimuli that were reserved for the warm-up trials and did not appear in later trials, one trial and stimulus for each stimulus category that appeared in the block. Single-task blocks were thus preceded by 2 warm-up trials; blocks combining both tasks were preceded by 4 warm-up trials.

The IATs used six stimuli per attribute and concept category. For the flower–insect IAT, these stimuli were taken from the word pools already used by Mierke and Klauer (2001, 2003). The political-attitudes IAT was the same as that in Experiments 2 and 3 of Klauer et al. (2007). It contrasted a red political attitude and a black political attitude. In Germany, the red political attitude is associated with the left political spectrum, including issues of social equality, preservation of the environment, and openness to other cultures. The black political attitude is associated with the right political spectrum, including issues of patriotism, authority, and conservative values. On the basis of pretests, six concept stimuli were

Table 1. *Blocks of the flower–insect IAT*

			Response key	
Block	*Trials*	*Tasks*	*Left*	*Right*
1	26[a]	Concept	Insect	Flower
2	26[a]	Attribute	Negative	Positive
3	28[b]	Combined Compatible	Insect or Negative	Flower or Positive
4	52[b]	Combined Compatible	Insect or Negative	Flower or Positive
5	26[a]	Concept	Flower	Insect
6	28[b]	Combined Incompatible	Flower or Negative	Insect or Positive
7	52[b]	Combined Incompatible	Flower or Negative	Insect or Positive

Note: IAT = Implicit Association Test.
[a]The first two trials are warm-up trials. [b]The first four trials are warm-up trials.

selected to represent the red political attitude (e.g., socialism, multicultural) and six to represent the black political attitude (e.g., conservative, fatherland) along with six attribute stimuli for the attribute "positive" (e.g., joy, love) and six for the attribute "negative" (e.g., emergency, poison). The political-attitudes IAT was otherwise constructed as the flower–insect IAT. The order in which the critical combined phases (Blocks 3 and 4 vs. Blocks 6 and 7) were presented was balanced across participants for the political-attitudes IAT, whereas the order was as shown in Table 1 for the flower–insect IAT. The dependent variable was the difference between the mean response latencies in the two kinds of critical combined phases (Blocks 3 and 4 vs. Blocks 6 and 7).

Data preprocessing, transformation, and outlier analyses

The dependent variables that were based on proportions (antisaccade task and stop-signal task) were arcsine-transformed. The latency-based measures were based on latencies of correct responses. For the central-executive measures based on response latencies (plus–minus task, number–letter task, colour–size task), response latencies that fell above or below an individual's mean latency by more than two standard deviations were left out of the analyses, excluding between 4.9% and 5.1% of the trials per task. For the IAT, response latencies below 300 ms and above 3,000 ms were set to these values (Greenwald et al., 2003).

Of the participants, 1 was excluded because the data from the span task were lost for this person. For the remaining 127 participants, we performed, following Miyake et al. (2000), bivariate outlier analyses on the within-construct correlations for the tasks designed to tap the executive functions, because analyses based on correlations are sensitive to outliers. There are five within-construct correlations, three between the three switching tasks, one between the two span tasks, and one between the two inhibition tasks. Outliers were identified by computing three outlier indices per correlation and participant: leverage, studentized t, and Cook's D values (Judd & McClelland, 1989, chap. 9), resulting in $15 = 3 \times 5$ outlier analyses per person. A total of 2 participants were excluded on the basis of these analyses, because they were identified as outliers in many of these indices; 1 person was conspicuous in 7 of 15 analyses, the other one in 4 of 15 tests.[2] The analyses were thus based on data from 125 participants.

Results

Descriptive statistics for the executive-control measures and the IATs are shown in Table 2 along with reliabilities. The correlations between the executive-control measures and the absolute sizes of the IAT effects can be found in Table 3 (in this and the subsequent analyses, all executive-control measures were signed so that higher values indicate higher ability). As can be seen, the intercorrelations of the executive-control measures were generally low, consistent with previous results (Miyake et al., 2000), but correlations

---

[2] A few more participants had unusually high outlier indices in one ($N = 10$), two ($N = 1$), and three ($N = 1$) of these tests, but given the large number of outlier indices computed ($n = 15 \times 127 = 1{,}905$), this was considered acceptable.

Table 2. *Descriptive statistics for central-executive tasks and IATs in Experiment 1*

Task family	*Task*	*Mean*	*SD*	*Range*	*Reliability*
Switching	Plus−minus	626	439	−218, 2,092	.79[a]
	Number−letter	226	166	−301, 704	.95[a]
	Colour−size	121	92	−148, 399	.66[a]
Inhibition	Antisaccade	1.15	0.16	0.63, 1.57	.88[a]
	Stop-signal	0.87	0.23	0.00, 1.37	.91[a]
Working memory	Reading span	3.66	0.75	0.47, 5.00	.85[b]
	Operation span	4.16	1.63	0.27, 6.00	.95[b]
IAT	Flower−insect IAT	203	137	−85, 945	.80[b]
	Political-attitudes IAT	322	269	−424, 1,272	.86[b]

Note: IAT = Implicit Association Test. Statistics for the switching tasks and IAT tasks are given in ms units; statistics for the inhibition tasks are based on arcsine-transformed proportions of correct responses; statistics for the working-memory tasks are based on the span scores described in the body of the paper.

[a] Split-half reliability, Spearman-Brown corrected. [b] Cronbach's α.

Table 3. *Intercorrelations of central-executive measures and (absolute sizes) of IATs in Experiment 1*

	Switching tasks			*Inhib. tasks*		*WM tasks*		*IATs*	
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1. Plus−minus	—	.18*	.05	.12	.10	.03	−.02	−.07	−.02
2. Number−letter		—	.31*	.12	−.04	.08	−.02	−.20*	−.20*
3. Colour−size			—	.13	.17	.05	.05	−.17	−.32**
4. Antisaccade				—	.26*	.21*	.11	−.13	−.17
5. Stop-signal					—	.27*	−.06	−.14	−.06
6. Reading span						—	.24*	−.12	−.10
7. Operation span							—	−.21*	−.07
8. Flower−insect IAT								—	.28**
9. Political-attitudes IAT									—

Note: IAT = Implicit Association Test. Inhib. = inhibition; WM = working-memory.
$^{*}p < .05$. $^{**}p < .01$.

between tasks believed to tap the same executive function tended to be higher than correlations between tasks tapping different functions. Moreover, as expected, the executive-control measures were generally negatively correlated with the absolute sizes of the IAT effects. The absolute sizes of the political-attitudes IAT and the flower−insect IAT were significantly correlated, $r = .28$, $p < .01$, suggesting that both share common method variance.

Preparatory analyses: The factor structure of executive tasks

As a preparatory step, we investigated the factor structure of the executive-control tasks. Following Miyake et al. (2000), we fitted a model with correlated factors switching, inhibition, and working-memory capacity to the covariance matrix of the executive-control measures. Like in Miyake et al. (2000), inhibition and working-memory capacity were most strongly correlated in this model, $r = .58$, $p = .01$, whereas the switching factor was more clearly separated from both the inhibition factor and the working-memory factor, $r = .39$, $p = .12$, and $r = .16$, $p = .38$, respectively. Model fit was good, but a more parsimonious model that merged the inhibition factor with the working-memory factor and allowed for correlated errors between the two span measures fitted just as well.

Models that merged the switching factor with

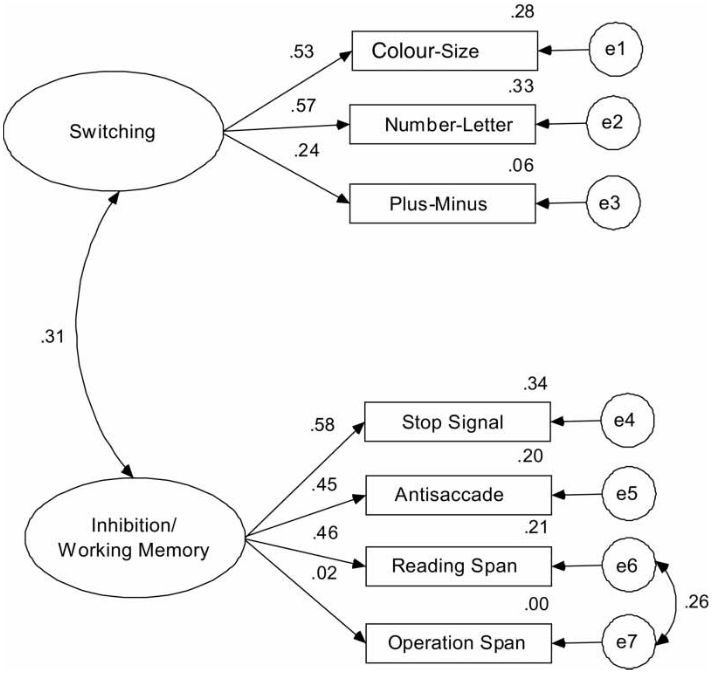

Figure 1. *The two-factor model of the central-executive tasks in Experiment 1. The numbers next to the straight single-headed arrows are the standardized factor loadings, and that next to the curved, double-headed arrow is the correlation of the two factors. The numbers next to the observed measures represent the variances explained for each executive-control measure.*

one of the other factors as well as a one-factor model were associated with substantially worse model-fit indices. Thus, the present data allow us to separate two executive-control factors, (a) switching and (b) inhibition/working-memory capacity.

For reasons of parsimony, we adopted the two-factor model shown in Figure 1 as a building block for the analyses involving the IAT. Model fit was good: $\chi^2 = 12.82$, $df = 12$, $p = .38$; $\chi^2/df = 1.07$; RMSEA = 0.02; CFI = .98, AGFI = .94.[3] As can be seen, the variances accounted for in each marker task are relatively small reflecting the fact that the intercorrelations between tasks are small to begin with, but they are in the order of magnitude of those reported by Miyake et al. (2000).[4]

---

[3] χ^2 is the chi-squared goodness-of-fit statistic; p is the associated significance level; values of χ^2/df between 0 and 2 indicate good fit, values between 2 and 3 acceptable fit. Values of p between .05 and 1 indicate good fit, values between .01 and .05 acceptable fit. RMSEA is the root mean square error of approximation; values below .05 indicate good fit, values between .05 and .08 acceptable fit. CFI is the comparative fit index; values above .97 indicate good fit, values between .95 and .97 acceptable fit. AGFI is the adjusted goodness-of-fit index; values above .90 indicate good fit, values between .85 and .90 acceptable fit (see Schermelleh-Engel, Moosbrugger, & Müller, 2003).

[4] We also conducted an exploratory factor analysis of the executive-control measures (a principal components analysis with oblimin rotation). This yielded results that converged with the confirmatory factor analysis described in the body of the text. Specifically, there were three factors with eigenvalues larger than one. One of these was a working-memory factor with highest loadings for the two span tasks; a second factor was a switching factor with highest loadings for the switching tasks; the third factor was an inhibition factor with highest loadings for the inhibition tasks. However, reading span also loaded substantially on this inhibition factor in line with the present finding that working memory and inhibition could be merged into one factor without significant loss in goodness of fit.

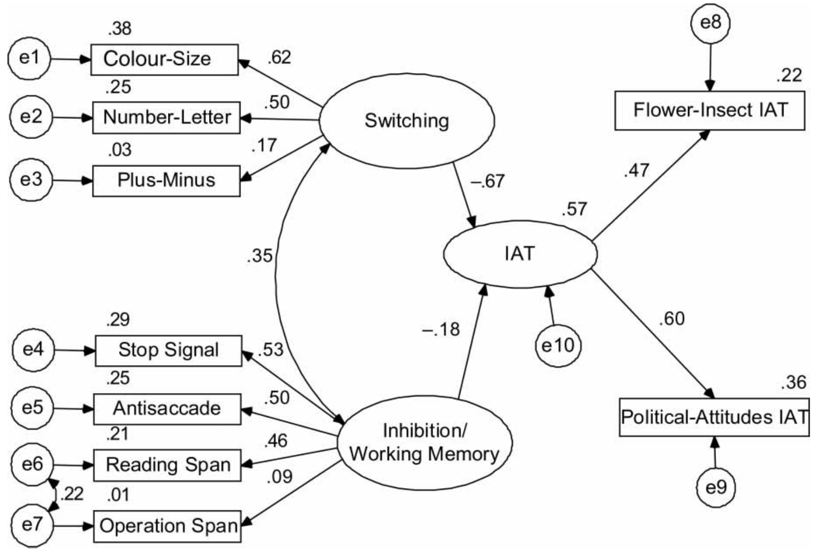

Figure 2. *Structural equation model regressing an IAT method factor on the two central-executive factors in Experiment 1. The numbers next to the single-headed arrows connecting central-executive factors and Implicit Association Test (IAT) method factor are standardized regression coefficients; the number next to the IAT method variable is the variance accounted for in that variable.*

Executive control and the size of IAT effects

How do these executive-control functions contribute to the size of IAT effects? The political-attitudes IAT and the flower−insect IAT were used to define a latent variable capturing shared method variance in the size of the IAT effect; for these analyses, the IATs were scored in terms of absolute size, ignoring the sign of the IAT effect. The model in Figure 2 includes paths from the two executive factors to this IAT factor. In addition to this full model, two reduced models were fitted. Both included only one path from executive control to IAT. The first model included only the path from switching to IAT, and the second model the path from inhibition/working memory to IAT.

The results are summarized in Table 4. As can be seen in Table 4, the coefficient for the path from switching to IAT is significant in the full model, but the coefficient for the path from inhibition/working memory to IAT is not. The χ^2−difference test indicates that the model without path from inhibition/working memory to IAT provides as good a fit as the full model, $\chi^2(1) = 0.61$, $p = .43$. Importantly, the model without path from switching to IAT fitted significantly worse than the full model, $\chi^2(1) = 8.73$, $p < .01$. Thus, the one-path model with switching as predictor for IAT method variance overall provides the best account of the pattern of covariances. It is also associated with the smallest AIC value—that is, it provides the best compromise between parsimony and fit, and its fit indices indicate a good fit (see Table 4). As expected, high ease of switching is associated with lower absolute sizes of IAT effects—that is, the standardized path coefficient is negative. This model accounts for 57% of the variance in the IAT method factor.

We repeated all analyses involving the IAT with the D_2 measure of the IAT effect proposed by Greenwald et al. (2003). The D measures have been optimized with regard to the IAT's psychometric criteria (e.g., increased internal consistency, higher correlations with explicit measures, resistance to some extraneous procedural influences). They differ from the conventional scoring algorithm in several aspects, including

Table 4. *Fit indices and standardized regression coefficients for structural equation models in Experiment 1*

	Models		
Index	*Full model*	*No path from inh./work. mem. to IAT*	*No path from switching to IAT*
$\chi^2(df)$	23.38 (23)	23.99 (24)	32.11 (24)
χ^2/df	1.02	0.99	1.34
p	.44	.46	.12
RMSEA	.01	.00	.05
CFI	.99	1.00	.87
AGFI	.93	.93	.90
AIC	67.38	65.99	74.11
Switching[a]	−.67*	−.76*	n.a.
Inhibition/working memory[a]	−.18	n.a.	−.70*

Note: IAT = Implicit Association Test. Inh./work. mem. = inhibition/working memory; n.a. = not applicable; AIC = Akaike's information criterion. RMSEA = root mean square error of approximation. CFI = comparative fit index. AGFI = adjusted goodness-of-fit index.

[a]Standardized regression coefficient from latent variable to IAT factor.

*p <.05.

modified upper and lower tail treatment of latencies, inclusion of all—that is, correct and incorrect—responses with incorrect response latencies being increased by an error penalty, and an individual standardization similar to that in Cohen's effect size measure *d* (see Greenwald et al., 2003).

Use of D_2 reduced the impact of method variance. For example, the correlation between the absolute sizes of the flower−insect IAT and the political-attitudes IAT decreased from $r = .28$, $p < .01$, for the conventional latency measures, to $r = .19$, $p = .03$, when the D_2 measures were used for both IATs. Interestingly, the structural equation model relating method variance in the D_2 measures to executive control that was analogous to the model in Figure 2 provided no evidence for significant contributions from either switching or inhibition/working memory in accounting for the IAT method factor. This null result might reflect that method variance was reduced for the D_2 measures, making it more difficult to find significant predictors of it. Alternatively, the remaining method variance in the D_2 measures might reflect more narrow or other kinds of processes than those captured in the executive-control measures.

Discussion

The results were relatively clear-cut. The space of executive-function tasks administered in the present study covered three broad executive functions: switching, inhibition, and working memory. Although we could separate only switching from a joint inhibition/working-memory factor, the switching component was found to be most clearly related to the IAT method factor, whereas there was little evidence for an independent involvement of the inhibition/working-memory factor.

As explained in the Introduction, these results are consistent with occasional findings of small correlations between the IAT and a marker of working-memory capacity or inhibitory abilities. Because different executive functions and markers of central-executive abilities all tend to correlate with each other (Miyake et al., 2000), small correlations between the IAT and such a marker can be caused by variance that both share with switching.

Use of D_2 reduced the impact of method variance in the IAT, but did not eliminate it. There was no evidence for significant contributions from either the switching or the inhibition/working-memory factor in accounting for the remaining method variance.

EXPERIMENT 2

Experiment 2 was a conceptual replication of Experiment 1 with additional controls and tasks. Specifically, Experiment 2 was to address two shortcomings of Experiment 1. First, given the important role assigned to inhibition in one major process account of the IAT, the Quad model (Conrey et al., 2005), it is unfortunate that inhibition and working-memory capacity had to be merged into one factor. The problem may have been due to the fact that only two inhibition tasks and two working-memory tasks were available for analysis. In Experiment 2, we used a broader array of four inhibition tasks and three working-memory tasks to increase the chances to separate a latent inhibition factor from a latent working-memory factor.

Second, in Experiment 1, IAT method variance was clearly related to task-set switching, but not to inhibition/working memory. This conforms well to the account of the IAT by task-set switching, but is difficult to explain in terms of the Quad model. There is, however, a potentially important confounding in Experiment 1 inasmuch as the switching measures and the IAT measures were based on response latencies, whereas the inhibition tasks were based on accuracy data. The overlap between switching tasks and IAT may therefore go back to the commonality in the metric of the measures, reflecting perhaps a general speed factor correlated with response latencies in any speeded task (Blanton, Jaccard, Gonzalez, & Christie, 2006). For this reason, inhibition measures and switching measures were all based on response latencies in Experiment 2, removing the confounding.

A number of minor changes were also implemented, most of them aiming at standardizing task features across tasks:

1. In all tasks other than the span tasks, participants used the interior keys of two computer mice positioned left and right in front of them to respond (Voss, Leonhart, & Stahl, 2007). This removed variance in the response-latency measurement introduced by the way in which the computer internally processed keypresses, and it equated switching tasks, inhibition tasks, and IATs in response modality.
2. In all tasks other than the span tasks, participants received immediate error feedback, in the form of a cross, shown below the stimulus and signalling that an error had been made. Participants had to enter the correct response to proceed to the next trial. We recorded latency to first response and latency to correct response. This allowed us to compute D_2 analogous measures for all latency-based tasks (see Greenwald et al., 2003). All latency-based tasks were also equalized in terms of intertrial interval (400 ms).
3. New versions of the span tasks operationalizing working-memory capacity were used—namely, the standardized versions recommended by Conway et al. (2005).
4. Scoring of the task-set switching tasks was standardized in that switching costs were defined as the performance decrement for task-switch trials relative to task-repetition trials in blocks with mixed tasks. Furthermore, we replaced the plus−minus task by another switching task, termed the semantic switching task, a task requiring task switches between two semantic categorization tasks. The plus−minus task had correlated little with any other task (see Table 3) and had contributed little to the measurement model defining the latent switching factor (see Figure 1).
5. Participants worked in the experimenter's presence throughout the study, whereas in Experiment 1, the experimenter had left the room when his or her presence was not required.
6. Finally, it may be that inhibition in the IAT plays a more pronounced role to the extent to which the content domain tapped by the IAT is socially sensitive, triggering efforts to inhibit socially undesirable responses. For this reason, the political-attitudes IAT was replaced by a prejudice IAT, measuring relative preference for German (ingroup) first names relative to Turkish (outgroup) first names. In addition to this prejudice IAT, we used the

flower–insect IAT already employed in Experiment 1 and an IAT with abstract contents developed by Back et al. (2005) that we refer to as the abstract IAT.

Based on Experiment 1, the hypothesis was that task-set switching would uniquely predict method variance in the IAT with little unique contribution by inhibitory abilities and working-memory capacity.

Method

Participants

Participants were 122 volunteers (79 female, 43 male); mean age was 23.5 years ($SD = 4.2$) with a range from 18 years to 42 years. About half of the participants were University of Freiburg students with different majors; the other half comprised students from vocational schools, a few students from high schools, and nonstudents with different professions. Participants received 20 euros as compensation for their participation. None of them had participated in Experiment 1.

Procedure

Participants were tested in individual sessions of about 2.5 hours. The tasks were administered in the following order: prejudice IAT, ratings of attitudes towards Germans and Turks, Stroop task, number–letter task, operation span, and antisaccade task, followed by a 5-min break; flower–insect IAT, colour–size task, reading span, and flanker task, followed by a 5-min break; abstract IAT, counting span, Simon task, and semantic switching task. At the end of a session, participants completed a biographical questionnaire.

Switching tasks

Switching tasks were the same as those in Experiment 1 with the exception that the plus–minus task was replaced by the semantic switching task.

Semantic switching task. The stimuli were nouns referring to objects that were either small (e.g., mouse) or large (e.g., elephant); orthogonally, the stimuli referred to either animate objects or inanimate objects (e.g., toaster, car). Two discrimination tasks were to discriminate the size (small vs. large) and to discriminate animacy status (animate vs. inanimate). Which task was to be performed was determined by whether the word was shown in blue or black. In the first case, size was to be discriminated; in the second case, it was animacy status. Participants worked through four blocks. In the first block of 48 trials, all stimuli were shown in blue; in the second block of 48 trials, all stimuli were shown in black. The third block was a practice block of 20 trials, in which the two tasks were mixed in a regular AABB pattern. This task sequence was also used in the fourth block of 96 trials, of which the first 4 were warm-up trials. The cost of switching was calculated on the basis of the data from the fourth block as the difference between the average latencies for trials in which task set had to be switched and those for trials in which task set was repeated.

Inhibition tasks

We developed new versions of the inhibition tasks with outcome measures based on response latency rather than error frequency. The antisaccade task, the Stroop task, the flanker task, and the Simon task were used to assess inhibition.

Antisaccade task. In a trial of the task, a fixation cue, a small square subtending approximately 0.4° of visual angle, was first presented in the middle of the screen for a random amount of time (between 500 ms and 1,250 ms in 50-ms intervals). A small square of the same size was then presented for 270 ms, shifted by 9° toward either the top or the bottom of the screen. This was followed by the presentation of the target stimulus, shifted by the same distance from the middle as the initial visual cue, either in the same direction (prosaccade task) or in the opposite direction (antisaccade task). The target consisted of either one or two thin vertical lines, separated by 0.1°, with a height of 0.4° and a width of 0.1°. Participants were to indicate whether there were one or two lines. The target stayed on screen until the participant had entered the correct response. Participants worked through 20 practice trials of the prosaccade task, followed by a block of 88 trials of that task, of which the first 8 were warm-up trials. Next, they worked through 20 practice trials of the antisaccade task followed by a block of 88 trials of that task, of which the first 8 were warm-up trials. The dependent variable was the difference between the mean latencies of responses in the antisaccade block and the mean latencies of responses in the prosaccade block.

Stroop task. A variant of the Stroop task was used that relied on comparisons of two stimuli (Kim, Kim, & Chun, 2005). The target consisted of a colour word (red, green, blue, or yellow) printed in black and a second letter string shown to the right of it. The second letter string was one of the above colour words or the letter string QQQ. It was printed in one of the above-mentioned colours, defining three consistency conditions: consistent (letter string is a colour word and is printed in the colour that it refers to); inconsistent (letter string is a colour word and is printed in a colour different from the one that it refers to); neutral (letter string is QQQ). Orthogonally, the left colour word and the print colour of the right letter string could be the same or not, defining two match conditions (match vs. mismatch). Participants were to respond "yes" and "no" according to match and mismatch, respectively. In all, there were six trial types, defined by crossing match and consistency condition. Participants first worked through 16 neutral trials, followed by four blocks in which all six trial types were equally frequent. The first of these was a practice block of 24 trials; the remaining three blocks comprised 52 trials each of which the first 4 were warm-up trials. The dependent variable was based on match trials as the difference between the mean latencies of responses in inconsistent trials and neutral trials.

Flanker task. In a trial of the flanker task (Friedman & Miyake, 2004), targets were the letters H, K, S, and C. The letters H and K were mapped on the response key of the left mouse, the letters S and C on the response key of the right mouse. Targets were flanked by letters (e.g., HHHSHHH), and participants were to respond on the basis of the middle letter, ignoring the flanking letters to the left and to the right of it. There were four trial types: identical (e.g., HHHHHHH), consistent (e.g., HHHKHHH), inconsistent (e.g., HHHSHHH), and neutral (e.g., S). Participants first worked through a practice block of 16 neutral trials, followed by five blocks in which the four trial types occurred equally frequently. The first of these was another practice block of 32 trials; the other four blocks each comprised 48 trials preceded by 4 additional warm-up trials. The cost of distracting response tendencies elicited by the irrelevant flankers was calculated, following Friedman and Miyake (2004), as the difference between the mean latencies of responses in inconsistent and neutral trials.

Simon task. A trial of the Simon task (Simon, 1990) began with the presentation of a fixation cross, subtending approximately 1° degree of visual angle for 300 ms. It was followed by an arrow pointing either right or left, shifted by 10° degree of visual angle to the right or to the left. Participants were to ignore the location of the arrows and to respond to the direction in which the arrow pointed by pressing the response key of the left mouse for arrows pointing to the left and the response key of the right mouse for arrows pointing to the right. There were two trial types: Trials in which location and pointing direction of the arrow matched and those in which they mismatched. Participants worked through a practice block of 24 trials, followed by two experimental blocks of 72 trials each, preceded by 2 additional warm-up trials. The cost of inhibiting the location of the arrows and response tendencies elicited by it was calculated as the difference between the mean latencies for mismatching trials and matching trials.

Working-memory capacity

The tasks used to operationalize working-memory capacity were reading span, operation span, and counting span in the versions recommended by Conway et al. (2005); they can be downloaded from http://psychology.gatech.edu/renglelab/tasks.htm (Engle, 2005).

Reading span and operation span were similar to the tasks used in Experiment 1. For a trial of the counting-span task, sequences of two to six pictures were presented in random order, each showing a number of targets and distractors. Targets were between three and nine dark blue circles; distractors were between one and five circles of a lighter blue, and between one and nine dark blue rectangles. Participants were to count out loud the number of targets in each picture and to repeat the total of targets for each picture. At the end of the sequence, they were to recall the totals (between two to six totals depending upon the number of pictures in the sequence) in correct order. In all, 15 sequences were presented preceded by 3 additional practice sequences. A sequence was scored by computing the proportion of items recalled in the correct serial position relative to the number of items to be recalled in the sequence. Counting span was operationalized, as recommended by Conway et al. (2005), as the sum of these proportions across items; it could range from 0 to 15. Scoring was analogous for reading span and operation span; reading span and operation span could range from 0 to 12.

Implicit association tests

The IATs followed the same format as that in Experiment 1. They were the flower–insect IAT used in Experiment 1, a prejudice IAT, and the abstract IAT (Back et al., 2005).

The prejudice IAT measured relative preference for German (ingroup) first names relative to Turkish (outgroup) first names. It was based on stimuli used by Gawronski (2002), of which we selected the six German first names (e.g., Stefan, Matthias) and the six Turkish first names (e.g., Ahmed, Mehmet) that were rated most typical of their categories in a pretest ($N = 20$).

Back et al. (2005) developed a control IAT to operationalize method variance based on abstract contents. In the abstract IAT, participants have to discriminate letter stimuli (e.g., C, N) from number stimuli (e.g., 4, 7) and words (e.g., shirt, table) from calculations (e.g., $8 - 5 = 3$, $2 + 6 = 8$). Each concept is associated with one of the attribute categories: Letters are associated with words and numbers with calculations.

Block order was constant for the three IATs; for the flower–insect IAT, prejudice IAT, and abstract IAT, the first critical blocks in the sequence mapped the following categories on the same key: in order, flowers and positive, German and positive, letters and words.

Data preprocessing, transformation, and outlier analyses

Data were preprocessed as in Experiment 1. That is, response latencies that fell above or below an individual's mean latency by more than two standard deviations were left out of the analyses, excluding between 4.4% and 6.6% of the trials per task. For the IAT, response latencies below 300 ms and above 3,000 ms were set to these values (Greenwald et al., 2003). We again performed bivariate outlier analyses. There were 12 within-construct correlations (3 between the three span tasks, 3 between the three switching tasks, and 6 between the four inhibition tasks) and 3 outlier indices (leverage, studentized t, and Cook's D values) computed per correlation and participant, resulting in 36 outlier indices computed per person. A total of 4 participants were excluded because they were conspicuous in 6 or more of the 36 outlier analyses computed per person.[5] The analyses reported below are therefore based on 118 participants.

Results

Descriptive statistics for the executive-control measures and the IATs are shown in Table 5 along with reliabilities. The correlations between the executive-control measures and the absolute

---

[5] Like in Experiment 1, a few more participants had unusually high outlier indices in one ($N = 3$), two ($N = 1$), three ($N = 5$), and four ($N = 3$) of these tests, but given the large number of outlier indices computed ($n = 36 \times 122 = 4{,}392$), this was considered acceptable.

Table 5. *Descriptive statistics for central-executive tasks and IATs in Experiment 2*

Task family	*Task*	*Mean*	*SD*	*Range*	*Reliability*[a]
Switching	Semantic	172	108	−184, 418	.58
	Number−letter	232	123	16, 681	.82
	Colour−size	134	94	−37, 459	.58
Inhibition	Antisaccade	83	40	7, 211	.94
	Stroop	135	1163	−40, 728	.67
	Flanker	58	28	3, 168	.67
	Simon	29	22	−16, 117	.69
Working memory	Reading span	9.87	1.48	4.77, 12.00	.72
	Operation span	9.54	1.41	5.75, 12.00	.68
	Counting span	11.75	1.62	6.58, 14.67	.62
IAT	Flower−insect IAT	214	132	−146, 632	.82
	Prejudice IAT	180	111	−237, 506	.82
	Abstract IAT	206	101	14, 665	.76

Note: IAT = Implicit Association Test. Statistics for switching, inhibition, and IAT tasks are given in ms units; statistics for the working-memory tasks are based on the span scores described in the body of the paper.

[a] Cronbach's α.

sizes of the IAT effects are shown in Table 6 (in this and the subsequent analyses, all executive-control measures were signed so that higher values indicate higher ability). As can be seen, the intercorrelations of the executive-control measures were generally low, consistent with previous results (Miyake et al., 2000), but correlations between tasks believed to tap the same executive function tended to be higher than correlations between tasks tapping different functions. Moreover, as expected, the executive-control measures were generally negatively correlated with the absolute sizes of the IAT effects. The absolute sizes of the three IATs were significantly correlated, $r < .21$, $p < .05$, suggesting that they shared common method variance.

Table 6. *Intercorrelations of central-executive measures and (absolute sizes) of IATs in Experiment 2*

	Switching tasks			*Inhibition tasks*				*Working-memory tasks*			*IATs*		
Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Semantic	—	.15	$.30^{**}$	.00	.07	−.16	−.01	−.09	−.16	−.04	−.15	−.02	−.14
2. Number−letter		—	.09	.17	.05	.00	.12	−.12	.01	.01	−.16	$-.21^{*}$	$-.27^{**}$
3. Colour−size			—	$.21^{*}$	.15	−.05	.12	−.02	−.06	.04	$-.26^{**}$	−.08	$-.19^{*}$
4. Antisaccade				—	$.23^{*}$	.05	$.25^{**}$	$.26^{**}$	$.22^{*}$	$.22^{*}$	$-.15^{*}$	$-.24^{**}$	$-.31^{**}$
5. Stroop					—	$.30^{**}$	.17	−.10	−.05	−.02	−.07	−.02	−.08
6. Flanker						—	$.21^{*}$	.04	.13	.13	.03	−.09	−.00
7. Simon							—	.13	.17	$.18^{*}$	−.10	−.14	$-.21^{*}$
8. Reading span								—	$.65^{**}$	$.59^{**}$	−.11	−.17	−.03
9. Operation span									—	$.51^{**}$	−.09	−.11	−.09
10. Counting span										—	−.07	−.10	$-.27^{**}$
11. Flower−insect IAT											—	$.48^{**}$	$.27^{**}$
12. Prejudice IAT												—	$.21^{*}$
13. Abstract IAT													—

Note: IAT = Implicit Association Test.

$^{*}p < .05$. $^{**}p < .01$.

Preparatory analyses: The factor structure of executive tasks

As a preparatory step, we investigated the factor structure of the executive-control tasks. Following Miyake et al. (2000), we fitted a model with correlated factors switching, inhibition, and working-memory capacity to the covariance matrix of the executive-control measures. We hoped to be able to separate these three factors more clearly than in Experiment 1. Model fit was acceptable: $\chi^2 = 41.09$, $df = 32$, $p = .13$; $\chi^2/df = 1.28$; RMSEA = 0.05; CFI = .94, AGFI = .89.

As can be seen in Figure 3, all three factors correlated only weakly with each other, largest $r = .39$, smallest $p = .15$. Importantly, it was not possible to merge any two of the three factors

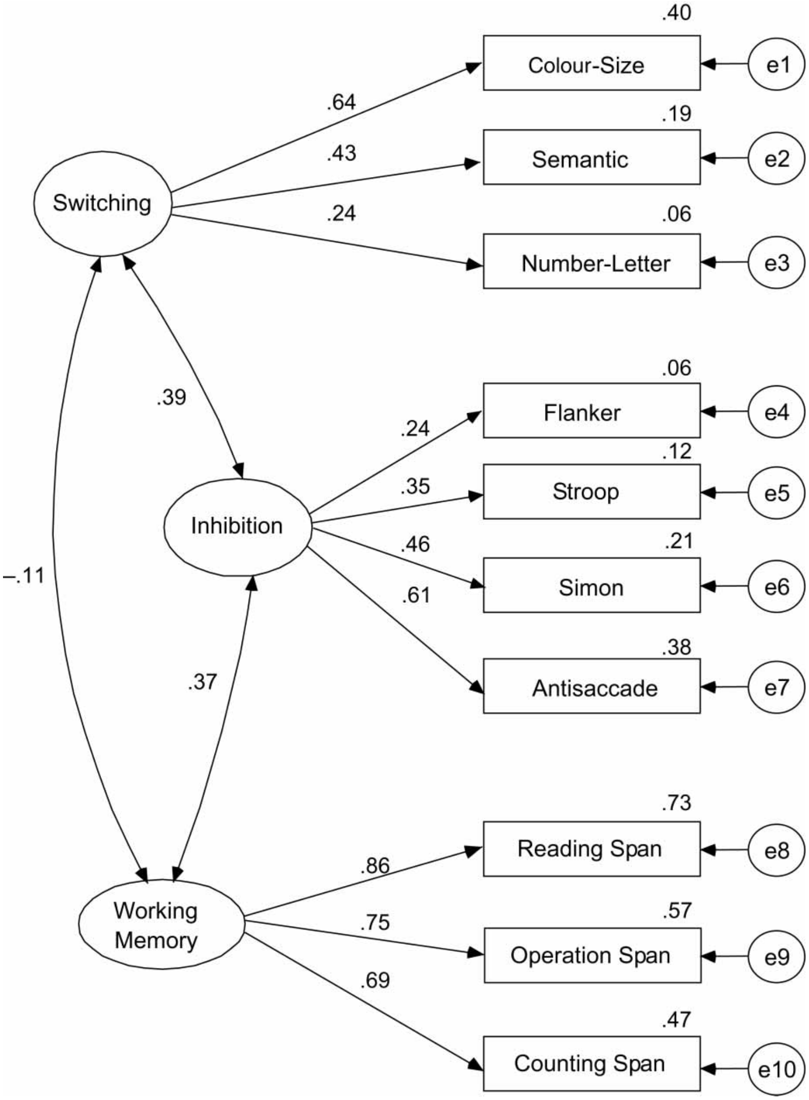

Figure 3. *The three-factor model of the central-executive tasks in Experiment 2. The numbers next to the straight single-headed arrows are the standardized factor loadings, and those next to the curved, double-headed arrows are the correlations of the three factors. The numbers next to the observed measures represent the variances explained for each executive-control measure.*

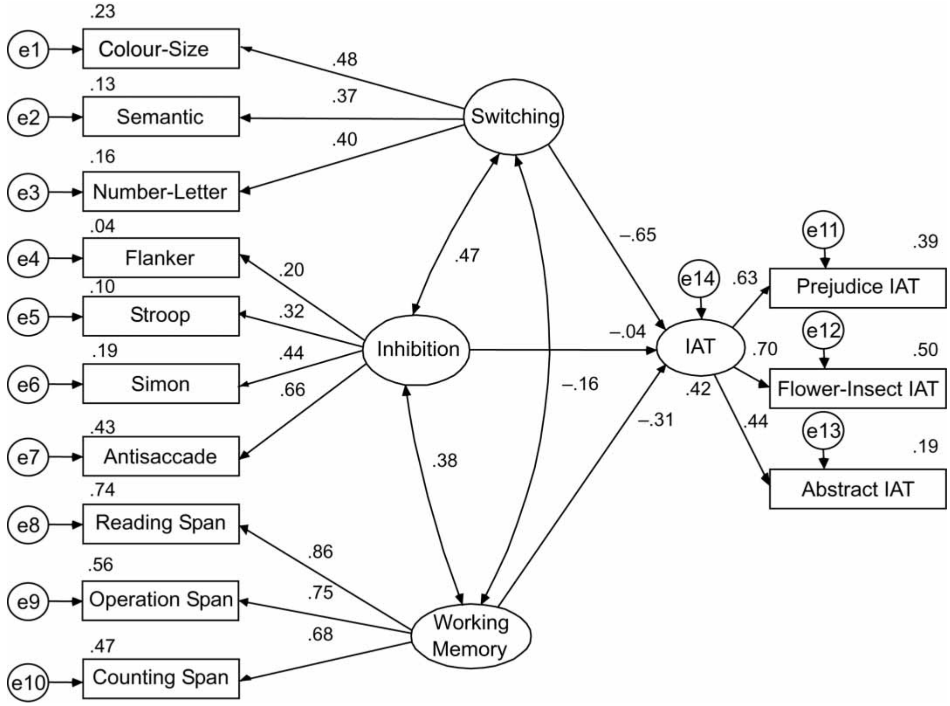

Figure 4. *Structural equation model regressing an Implicit Association Test (IAT) method factor on the three central-executive factors in Experiment 2. The numbers next to the single-headed arrows connecting central-executive factors and IAT method factor are standardized regression coefficients; the number next to the IAT method variable is the variance accounted for in that variable.*

without significant loss in model fit, smallest $\chi^2(2) = 11.11$, largest $p < .01$. The model in Figure 3 was therefore used as building block for the analyses involving the IATs.[6]

Executive control and the size of IAT effects

How do these executive-control functions contribute to the size of IAT effects? The three IATs were used to define a latent variable capturing shared method variance in the size of the IAT effect; for this purpose, the IATs were scored in terms of absolute size, ignoring the sign of the IAT effect. The model in Figure 4 includes paths from the three executive factors to this IAT factor. In addition to this full model, three reduced models were fitted. In each, one of the three paths from executive control to IAT was left out. The first model left out the path from switching to IAT, the second model the path from inhibition to IAT, the third model the path from working memory to IAT. Based on the results of Experiment 1, we expected that leaving out the path from switching to IAT would lead to a significant loss in model fit, whereas leaving out one of the other paths would not deteriorate model fit significantly.

The results are summarized in Table 7. The χ^2−difference test indicates that the model without path from switching to IAT fitted significantly worse than the full model, $\chi^2(1) = 0.01$,

---

[6] We again submitted the central-executive measures to an exploratory factor analysis (see Footnote 4). Like in Experiment 1, results converged with the confirmatory factor analysis described in the body of the text. That is, there were three factors with eigenvalues larger than one, each of them defined by one of the three groups of tasks—that is, switching tasks, inhibition tasks, and working-memory tasks, in terms of highest loadings.

Table 7. *Fit indices and standardized regression coefficients for structural equation models in Experiment 2*

	Models			
Index	*Full model*	*No path from switching to IAT*	*No path from inhibition to IAT*	*No path from work. mem. to IAT*
χ^2 (*df*)	77.51 (59)	81.93 (60)	77.52 (60)	79.62 (60)
χ^2 / *df*	1.31	1.36	1.29	1.33
p	.053	.035	.064	.046
RMSEA	.05	.06	.05	.5
CFI	.92	.90	.92	.91
AGFI	.86	.86	.86	.86
AIC	141.51	143.93	139.52	141.62
Switching[a]	−.65	n.a.	−.69*	−.34
Inhibition[a]	−.04	−.63	n.a.	−.39
Work. mem.[a]	−.31	.05	−.33*	n.a.

Note: IAT = Implicit Association Test. Work. mem. = working memory; n.a. = not applicable; AIC = Akaike's information criterion. RMSEA = root mean square error of approximation. CFI = comparative fit index. AGFI = adjusted goodness-of-fit index.
[a]Standardized regression coefficient from latent variable to IAT factor.
*$p < .05$.

$p = .91$, so that switching is shown to provide an independently significant contribution to accounting for method variance in the IAT. In contrast, the model without path from inhibition to IAT fitted as well as the full model, $\chi^2(1) = 0.01$, $p = .91$, and the model without path from working memory to IAT did not fit significantly worse than the full model, $\chi^2(1) = 2.11$, $p = .15$. Taken together, the evidence is strongest for a role of task switching in the IAT, and there is little evidence for an independent role of inhibition.

In fact, the model without path from inhibition to IAT fared best on all fit measures as shown in Table 7. All of the fit measures, except CFI, showed an acceptable or good fit for this model.[7] As can be seen, the coefficients for the paths from switching and from working memory are both significant in this model. Leaving out one of these paths significantly deteriorated model fit: smallest $\chi(1) = 5.73$, largest $p = .02$. The model accounted for 51% of the variance in the IAT method factor.[8]

We repeated all analyses using the D_2 measures for the IAT. Use of D_2 again reduced the intercorrelations between the three IATs, but the prejudice IAT still correlated significantly with the flower−insect IAT, $r = .40$, $p < .01$, and with the abstract IAT, $r = .21$, $p < .05$.

Thus, method variance was reduced, but not eliminated through the use of D_2 (see also Experiment 1; Klauer et al., 2007; Mierke & Klauer, 2003). We fitted the structural equation model of Figure 4 with the D_2 measures for the IATs in two versions: (a) with switching and inhibition measures scored as before, and (b) with switching and inhibition measures scored in D_2

---

[7] With correlated errors between flanker task and Stroop task, fit indices improved for all models tested. In particular, fit was also acceptable in terms of CFI for the full model and for the model without path from inhibition to IAT. The pattern of significant and nonsignificant results remained the same, except that the χ^2−difference test comparing the full model and the model without path from switching to IAT gave a result slightly above the conventional 5% level of significance, $\chi^2(1) = 3.616$, $p = .057$. Note that it is significant in a one-tailed test ($p = .029$). One-tailed testing is legitimate given our hypothesis that the path coefficient from switching to IAT would be negative (as it was).

[8] We also used the model from Figure 3 to predict each IAT individually to see whether inhibition might play a greater role for the prejudice IAT than for the less socially sensitive flower−insect IAT and abstract IAT (see introduction to Experiment 2). For each IAT, it was possible, however, to leave out the individual path from any latent variable to the IAT without significant loss of model fit so that the analyses were not powerful enough to demonstrate a unique contribution of any of the latent factors to the IAT.

analogous fashion (remember that we recorded both latency to first response and latency to correct response for these measures, allowing us to apply the D_2 logic to these measures as well). Analysis (a) did not reveal significant contributions from any of the executive control factors in accounting for the remaining method variance in the IATs' D_2 measures. In Analysis (b), so-called Heywood cases occurred in the estimation of the path coefficients in the measurement model for the inhibition factor (i.e., standardized path coefficients were estimated to be larger than one), possibly reflecting departures from the normal distribution for the D_2−transformed measures. This rendered the fitting of a meaningful structural equation model impossible. Instead, we computed factor scores for a switching factor and an inhibition factor as weighted sums of the D_2−transformed switching and inhibition measures, respectively, using as weights the factor loadings estimated in the structural equation model for the conventionally scored executive measures (i.e., the factor loadings as shown in Figure 3). An IAT method factor was analogously computed on the basis of the D_2−transformed IAT scores, and a working-memory factor on the basis of the working-memory span tasks. The resulting variables were entered into a multiple regression with the IAT method factor as dependent variable and switching, inhibition, and working memory as predictors. Standardized regression coefficients β and p values were for switching, inhibition, and working memory, in order, $\beta = .24$ and $p < .01$, $\beta = .09$ and $p = .30$, and $\beta = .12$ and $p = .20$. Thus, the IAT method factor was uniquely predicted by the switching factor when both predictors and dependent variables were equated in terms of scoring algorithm.[9]

Discussion

The major results of Experiment 1 could be replicated using stronger controls. First of all, the measurement model now allowed us to separate the three central executive factors that we intended to operationalize. Based on that measurement model, strongest evidence accrued for a role of switching in accounting for IAT method variance: A model without the path from switching to IAT led to a significant loss in goodness of fit, but both the paths from working memory and those from inhibition to IAT could be omitted without significant loss in goodness of fit. The evidence was weakest for a contribution of the inhibition factor. Leaving out the path from inhibition to IAT in fact did not noticeably change the χ^2 value for goodness of fit, whereas it even improved the other fit measures that incorporate the model's parsimony and robustness in different ways. The best-fitting model left out the path from inhibition to IAT, and this model also suggested a secondary contribution of working-memory capacity over and above the contribution of switching inasmuch as the path coefficient from working-memory capacity to IAT was also significant in this model.

These results defend the previous results against several alternative explanations as elaborated above. In particular, the more prominent role of switching than of inhibition can no longer be attributed to overlap between the IATs and the switching tasks in measurement domain (errors versus response latencies) and to factors such as general processing speed associated with it. In Experiment 2, both switching measures and inhibition measures were based on response latencies.

Method variance was reduced when the IATs were scored in terms of D_2, but intercorrelations of IATs were still substantial despite the use of D_2. Method variance in the D_2 metric was, however, not systematically related to the executive measures when these were scored in the conventional latency metric. In Experiment 2, we implemented procedures (i.e., the requirement to correct a wrong response) that enabled us to compute D_2 analogous scores for the switching measures as well as the inhibition measures.

---

[9] Further analyses suggested that of the many differences between the conventional scores and the D_2 scores, the person-wise standardization in D_2 is responsible for disrupting relationships between IAT method variance scored in terms of D_2 and the conventionally scored executive measures.

It turned out that use of D_2 for both predictors and IATs recovers the pattern that is evident in the latency-based analyses: Method variance in D_2 was clearly related to the switching measures, scored in terms of D_2, whereas the inhibition measures and the working-memory tasks independently contributed little to accounting for method variance. The bottom line is that IAT method variance still depends on switching even when D_2 is used to score the IATs, although the relationship is artificially masked when the nonlinear D_2 transformation is applied to only one side of the equation (i.e., only to the IAT measure and not to the executive measures).

GENERAL DISCUSSION

Across two studies, switching ability was found to be most clearly related to the size of IAT effects. This lends support to the account of the IAT by task-set switching (Klauer & Mierke, 2005; Mierke & Klauer, 2003). The predictions derived from Conrey et al.'s (2005) Quad model received only mixed support. Inhibitory abilities were not related to the IAT method factor beyond the correlation caused by the indirect path via switching (the different executive functions are intercorrelated). In Experiment 2, there was evidence for a small independent contribution of working-memory capacity as predicted by the Quad model. Note that the Quad model is a model of the accuracy data of the IAT, whereas the present analyses were based on response latencies and D_2 measures of the IAT. It is possible that inhibition and working-memory capacity play a larger role in the accuracy domain than in the latency domain. Although we recorded accuracy data for the IATs and the executive tasks, reliabilities were generally too low for a meaningful latent-variables analysis. Furthermore, as already mentioned, it is possible that inhibition plays a more prominent role in IATs that tap socially sensitive domains in which participants are motivated to conceal their attitudes although we found no evidence for this in the prejudice IAT used in Experiment 2 (see Footnote 8).

Apart from the theoretical implications for process accounts of the IAT, the present findings also bear on the use and interpretation of the IAT in research settings in which experimental manipulations and individual differences in independent and dependent variables covary with differences in executive control. In these cases, effects on, or correlations with, IAT measures may be mediated by differences in executive control.

For example, executive control declines with old age (von Hippel, Silver, & Lynch, 2000), suggesting that there may be age effects on IAT measures independently of the content domain (e.g., prejudice, self-esteem) that is to be measured by the IAT (see also Hummert, Garstka, O'Brien, Greenwald, & Mellot, 2002; Sherman et al., 2008). Similarly, Schmader and Johns (2003) argued that stereotype threat depresses executive control capacity, and thus threat effects on IATs (Frantz, Cuddy, Burnett, Ray, & Hart, 2004) might reflect the contribution of central-executive functions in the IAT. As another example, consider Richeson and Shelton's (2003) findings of correlations between a race IAT and a Stroop task. The Stroop task is seen as a central-executive task (Miyake et al., 2000), suggesting that small correlations between Stroop task and IAT might reflect a shared central-executive component. Note, however, that Richeson and Shelton (2003) found the correlation between race IAT and Stroop task to be increased when White participants had contact with a Black rather than a White confederate, a finding that cannot be explained by this central-executive hypothesis alone. Finally, Hofmann, Gschwendner, Friese, Wiers, and Schmitt (2008) recently found that working-memory capacity moderated the predictive validity of IATs measuring the implicit attitude towards temptations in predicting self-regulatory behaviour. Predictive validity was higher for participants low in working-memory capacity than for participants high in working-memory capacity. Again, this might reflect effects of working-memory capacity and related executive-control functions on the IAT scores rather than effects of working-memory capacity on self-regulatory behaviour as suggested by

Hofmann et al. (2008). Note, however, that Hofmann et al.'s results were based on the D_6 score for the IAT that is similar to the D_2 score discussed next.

The switching tasks employed here can be administered to rule out, or to evaluate the contribution made by, such alternative hypotheses. Use of D_2 reduced method variance, but did not eliminate it (see also Klauer et al., 2007). In fact, the method variance remaining in D_2 was uniquely predicted by the switching measures when these (and the inhibition measures) were also scored in D_2 analogous fashion.

The present studies contribute to our understanding of the role of the central executive in the IAT. Individual differences in the flexibility with which individuals switch between mental sets accounted for substantial amounts of the variance shared by unrelated IATs. Somewhat surprisingly, inhibitory ability independently contributed little to accounting for the IAT method factor.

Original manuscript received 12 December 2008
Accepted revision received 24 May 2009

REFERENCES

Allport, D. A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance XV* (pp. 421–452). Cambridge, MA: MIT Press.

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology*, *52*, 167–179.

Blanton, H., Jaccard, J., Gonzalez, P., & Christie, C. (2006). Decoding the Implicit Association Test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*, 192–212.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, *12*, 769–786.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Engle, R. W. (2005). *Tasks*. Retrieved September 6, 2008, from http://psychology.gatech.edu/renglelab/tasks.htm.

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*, 1611–1624.

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.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, *49*, 171–180.

Gonsalkorale, K., von Hippel, W., Sherman, J. W., & Klauer, K. C. (2009). Bias and regulation of bias in intergroup interactions: Implicit attitudes toward Muslims and interaction quality. *Journal of Experimental Social Psychology*, *45*, 410–414.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: 1. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Hofmann, W., Gschwendner, T., Friese, M., Wiers, R., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Towards an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal of Personality and Social Psychology*, *95*, 962–977.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging*, *17*, 482–495.

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *89*.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach*. San Diego, CA: Harcourt Brace Jovanovich.

Kim, S., Kim, M., & Chun, M. M. (2005). Concurrent working memory load can reduce distraction. *PNAS*, *102*, 16524–16529.

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test effect. *Personality and Social Psychology Bulletin*, *31*, 208–217.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Processing components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York: Guilford Press.

Logan, G. D. (1994). On the ability to inhibit thought and action: A user's 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.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition*, *20*, 481–506.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*, 211–253.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT. Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie*, *48*, 107–122.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*, 1180–1192.

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.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove, UK: Psychology Press.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). Hove, UK: Psychology Press.

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity: Facets of a cognitive ability construct. *Personality and Individual Differences*, *29*, 1017–1045.

Payne, B. K. (2005). Conceptualizing control in social cognition: How executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*, 488–503.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. *Psychological Science*, *14*, 287–290.

Roberts, R. J., Jr., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*, 374–393.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, *124*, 207–231.

Rubinstein, J. S., Meyer, D. E., & Evans, J. E. (2001). Executive control of cognitive processes in task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 763–797.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*. Available from Methods of Psychological Research Online, http://www.mpr-online.de

Schmader, T., & Johns, M. (2003). Convergent evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus–response compatibility: An integrated perspective* (pp. 31–86). Amsterdam: North-Holland.

Teige-Mocigemba, S., Klauer, K. C., & Sherman, J. W. (in press). Practical guide to Implicit Association Test and related tasks. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.

Von Hippel, W., Silver, L. A., & Lynch, M. E. (2000). Stereotyping against your will: The role of inhibitory ability in stereotyping and prejudice among elderly. *Personality and Social Psychology Bulletin*, *26*, 523–532.

Voss, A., Leonhart, R., & Stahl, C. (2007). How to make your own response boxes: A step-by-step guide for the construction of reliable and inexpensive parallel-port response pads from computer mice. *Behaviour Research Methods*, *39*, 797–801.