Using Implicit Association Tests in Age-Heterogeneous Samples: The Importance of Cognitive Abilities and Quad Model Processes

Cornelia Wrzus, Boris Egloff, and Michaela Riediger

Implicit association tests (IATs) are increasingly used to indirectly assess people's traits, attitudes, or other characteristics. In addition to measuring traits or attitudes, IAT scores also reflect differences in cognitive abilities because scores are based on reaction times and errors. As cognitive abilities change with age, questions arise concerning the usage and interpretation of IATs for people of different age. To address these questions, the current study examined how cognitive abilities and cognitive processes (i.e., quad model parameters) contribute to IAT results in a large age-heterogeneous sample. 549 participants (51% female) in an age-stratified sample (range 12-88 years) completed different IATs, and two tasks to assess cognitive processing speed and verbal ability. From the IAT data, D_2-scores were computed based on reaction times, and quad process parameters (*activation of associations, overcoming bias, detection, guessing*) were estimated from individual error rates. Substantial IAT scores and quad processes except *guessing* varied with age. Quad processes AC and D predicted D_2-scores of the content-specific IAT. Importantly, the effects of cognitive abilities and quad processes on IAT scores were not significantly moderated by participants' age. These findings suggest that IATs seem suitable for age-heterogeneous studies from adolescence to old age when IATs are constructed and analyzed appropriately, for example with D-scores and process parameters. We offer further insight into how D-scoring controls for method effects in IATs and what IAT scores capture in addition to implicit characteristics.

Keywords: implicit association test, lifespan, cognitive abilities, quad model, multinomial modelling

Author note.

Authors: Cornelia Wrzus, Psychological Institute, Johannes Gutenberg-University, Mainz, Germany; Boris Egloff, Psychological Institute, Johannes Gutenberg-University, Mainz, Germany; Michaela Riediger, Heisenberg Research Group Socio-emotional Development and Health Across the Lifespan, Freie Universität Berlin, and Max Planck Institute for Human Development, Berlin, Germany.
We are grateful to Christoph Stahl for providing the software HMMTree for the estimation of individual multinomial parameters. Work on this paper was supported in part by a Heisenberg stipend awarded to Michaela Riediger by the German Research Foundation [Deutsche Forschungsgemeinschaft, DFG], RI 1797/3-1, and by the German Federal Ministry for Education and research grants MPI001 and 01UW0706 awarded to Michaela Riediger.
Correspondence concerning this article should be addressed to Cornelia Wrzus, Psychological Institute, Johannes Gutenberg-University, Binger Str. 14-16, 55099 Mainz, Germany. E-mail: wrzus@uni-mainz.de

Indirect tests, such as implicit association tests (IAT) and affective priming tasks, are useful complements to explicit self-reports for a variety of constructs (e.g., personality traits, social attitudes) as they aim at assessing implicit, that is, introspectively less accessible, parts of the constructs (Evans, 2008; Greenwald, McGhee, & Schwartz, 1998). Although, for example, IATs solely aim at assessing individual differences in implicit associations (e.g., implicit personality concepts, implicit attitudes), IAT results partly reflect differences in cognitive abilities, such as general intelligence or perceptual speed, because IAT results are derived from reaction times and errors (Greenwald, Banaji, & Nosek, 2014; Ito et al., 2015; Nosek, Smyth, et al., 2007). Importantly, cognitive abilities differ between individuals and vary with individuals' age raising questions about the usage and interpretation of IATs in age-heterogeneous samples.

Until now, IATs were almost exclusively applied to samples of young adults (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), except a few studies with children (Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2015) or older adults (e.g., Gonsalkorale, Sherman, & Klauer, 2014; Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002; Kunzmann & Thomas, 2014; Riediger, Wrzus, & Wagner, 2014). The current work examines the applicability and boundaries of using IATs from adolescence to late adulthood by studying how cognitive abilities and processes (i.e., quad model parameters) contribute to IAT results in a large age-heterogeneous sample. The results offer insights into processes of age-related differences in IAT methods, and thus whether implicit representations of attitudes or personality traits can be measured equally among people of different age.

Applying and Interpreting IATs and Quad Processes

IATs have proven fruitful for many research areas including personality traits, attitudes, and emotions (e.g., reviews by Greenwald, Banaji, & Nosek, 2015; Nosek, Smyth, et al., 2007). Standard IATs are computer-based categorization tasks, where participants sort two sets of dichotomous stimuli, such as flower and insect names as well as pleasant and unpleasant adjectives, as fast and as correctly as possible according to their superordinate categories (e.g., flowers, pleasant) during several trials and using a computer keyboard (Greenwald et al., 1998, see *Method* for more details). In some trials, compatible concepts share the same response key (e.g., left key for flowers and pleasant adjectives), whereas in other trials incompatible concepts share the same key (see Table A1). The central assumption is that the strength of the cognitive association among concepts facilitates faster responses in the compatible trials when the associated concepts share the same response keys, and slows responses in the incompatible trials. Accordingly, the original conventional IAT score is the difference, or latency, between the average reaction times in the incompatible trials and the compatible trials (Greenwald et al., 1998). Thus, larger positive differences approximate a stronger association of flowers with pleasant and insects with unpleasant compared to the contrary combinations.

Previous work with young adults has shown that general response speed contributes to the conventional difference score of the IAT (Greenwald, Nosek, & Banaji, 2003; McFarland & Crouch, 2002). That is, people who respond faster, in general and especially in the incompatible combined trials, receive smaller IAT scores as these are computed as differences in reaction times between incompatible and compatible trials. The smaller IAT scores would indicate a weaker implicit association, although the smaller scores were due to faster responses. To address this problem, Greenwald and colleagues (2003) proposed D-scores as an improved scoring algorithm. The D-scores are obtained by dividing the individual difference scores in reaction times by the individual standard deviations of reaction times in the combined trials. Associations with general response speed were reduced for the D-score compared to the conventional score and hence suggest that the D-score may account for some method effects—at least in young adult populations (Cai, Sriram, Greenwald, & McFarland, 2004; Greenwald et al., 2003; Nosek, Greenwald, & Banaji, 2007). Recently, D-scores were also criticized for being susceptible to individual differences in other cognitive abilities (Evans, 2008; Ito et al., 2015; Siegel, Dougherty, & Huber, 2012) and for being unable to distinguish the associative and method-related components of IAT scores (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005). To separate and quantify the components underlying IAT performance, the quad model was introduced (Calanchini, Sherman, Klauer, & Lai, 2014; Conrey et al., 2005).

The quad model builds upon multinomial models (e.g., Payne, 2008; Riefer & Batchelder, 1988) that focus on dissociating content-related (e.g., associative) and content-unrelated method processes in categorization tasks such as IATs (Calanchini et al., 2014; Conrey et al., 2005). The relative contributions of four distinct processes to the task performance (Figure 1) are estimated from individuals' errors when categorizing the different stimuli (e.g., flowers, unpleasant adjectives) in the combined trials of IATs, that is, when all stimuli are presented in turns and compatible stimuli share either the same or different response keys. The validity of the quad model and its four parameters, described next, has been repeatedly shown for IAT measures (e.g., Calanchini & Sherman, 2013; Calanchini et al., 2014; Conrey et al., 2005; Gonsalkorale, Sherman, Allen, Klauer, & Amodio, 2011; Gonsalkorale et al., 2014; Sherman, 2006a,b).

Activation of association (AC) describes how strongly associations are activated and influence the response. This parameter is estimated separately for the two values of the categories (e.g., AC1 insect—unpleasant, AC2 flower—pleasant) and approximates the core measurement intention of IATs. *Detection* (D) describes the ability to detect correct and incorrect responses, that is, how much the ability to correctly identify and categorize the stimuli to the correct category value contributes to the response. *Overcoming bias* (OB) describes the ability to inhibit an activated association to provide the correct response. Finally, *guessing* (G) describes the general tendency to respond with a certain key (left or right), which represents guessing when no association is activated or the correct answer is unknown (Calanchini et al., 2014; Conrey et al., 2005). In general, the likelihood to produce a correct (+) or incorrect (-) categorization in one of the combined trials depends on the four processes, specifically the sum of their conditional probabilities (Figure 1). Using the actual, observed errors and the model implied error rates for different stimuli in different trials, the strength of the four quad parameters is estimated for each individual (for more details see section *Analytic Strategy*, and Calanchini et al., 2014; Conrey et al., 2005; Sherman, 2006a).

While the reaction-time based D-scores are applied routinely for analyzing IAT data, the error-based estimates of quad processes are scarcely used. D-scores possess the advantage that they are easily computable, also when no errors were made. At the same time, the interpretation of D-scores is ambiguous because in addition to the associative processes of interest, non-associative method processes are assessed (Calanchini et al., 2014; Sherman, 2006). Thus, the same D-score can result from differently strong associative processes, due to the other processes involved (Conrey et al., 2005). The estimation of quad processes aims at separating the associative and method processes. Thus, D-scores and quad process parameter provide complementary information when analyzing IAT data, but both scoring algorithms combined, or IATs in general, have been scarcely used in age-comparative research.

Age Differences in IAT Effects and Quad Processes

Only a few IAT studies on age differences in attitudes or personality traits exist, perhaps partly because age-related differences in conventional IAT scores cannot be interpreted clearly as substantial age differences in the measured concept. The observed age effects also contain age-related differences in cognitive processing speed (Hummert et al., 2002; Nosek, Greenwald, et al., 2007). The D-score presumably controls influences of age-related differences in processing speed. Two studies employed method IATs, that is IATs without (strong) age differences in pre-existing associations (e.g., letters—words vs. numbers—equations) that should largely capture method-inherent effects. Both studies observed larger conventional IAT scores with age, but no significant age differences in D-scores of method IATs—yet, neither study examined cognitive abilities or process parameters as explanation (Hummert et al., 2002; Riediger et al., 2014).[1]

Few studies used the quad or related models to discern age differences in attitude-related implicit associations from age-differences in other processes, which are involved during completing attitude IATs (Gonsalkorale, Sherman, & Klauer, 2009; Gonsalkorale et al., 2014; Stewart, von Hippel, & Radvansky, 2009). The studies observed significantly larger IAT effects with older age, suggesting more pronounced racial attitudes among older individuals, but activation of association (AC) parameters estimated from quad models did not differ significantly with age (Gonsalkorale et al., 2009; Stewart et al., 2009). In contrast, processes related to the inhibition of activated associations were less pronounced with older age, presumably contributing to the larger IAT effects. Detection of correct responses (D) was higher among older individuals and guessing (G) was negligibly to slightly higher among older individuals (Gonsalkorale et al., 2009; 2014). Surprisingly, these studies did not examine associations between quad processes and IAT scores, which would offer a better understanding of the processes contributing to IAT scores, and hence the meaning of IAT scores.

Although cognitive abilities, especially processing speed, contribute to individual differences in IAT scores, none of the earlier age-comparative studies examined cognitive measures. Only Stewart et al. (2009) showed in a small sample (N=112, aged 40-91 years) that with older age, implicit racial attitudes were more pronounced (i.e., IAT scores were larger), and that cognitive control partly accounted for age effects on control processes in the IAT (e.g., inhibition). Next, we describe studies that examined the associations between cognitive abilities and IAT results among young adults.

Cognitive Abilities Contributing to IAT Scores and Quad Processes

Most previous studies have focused on components of fluid-cognitive abilities, such as cognitive speed, when examining individual differences in IAT scores (e.g., Ito et al., 2015; Klauer, Schmitz, Teige-Mocigemba, & Voss, 2010; Siegel et al., 2012). In general, cognitive processing speed contributes to all scores based on reaction-time data (e.g., Nosek, Greenwald, et al., 2007). Thus also in IATs, individuals who work more slowly on the task exhibit larger conventional IAT scores than individuals with faster average reaction times (McFarland & Crouch, 2002).

The D-scoring algorithm largely reduced the association between IAT scores and general processing speed in some studies (Cai et al., 2004; Greenwald et al., 2003), and specifically between IAT scores and task switching costs as well as working memory (Back, Schmukle, & Egloff, 2005; Klauer et al., 2010; Mierke & Klauer, 2003). However, other studies still found a significant correlation between D-scores of a racial attitudes IAT and inhibitory control or shifting ability (Ito et al., 2015; Siegel et al., 2012).

Associations between cognitive abilities and quad processes in IATs have not yet been tested. Stewart and colleagues (2009) focused on inhibitory control ability and associative vs. control processes (i.e., attempts to control prejudicial responses) of a racial IAT, and found that higher inhibitory control predicted more pronounced control process in the IATs. Ito and colleagues (2015) also distinguished between associative vs. control processes in a racial

IAT and reported associations between control processes and cognitive abilities (i.e., specifically memory updating and task shifting abilities), but no associations between cognitive abilities and associative processes among young adults. Further indirect evidence on the assumption that general cognitive abilities contribute to quad processes comes from research showing that D and OB (an inhibition process) seemed to be domain-general processes, where individual differences are relatively stable across IATs with diverse content (Calanchini et al., 2014). This suggests that D and OB processes may depend more on fluid-cognitive abilities compared to AC and G, which were both content-specific processes, that is, their strength depended on the content of the specific IATs (Calanchini et al., 2014). Notably, the classification of domain-general and domain-specific processes stems from age-homogeneous samples with young adults and may not generalize across the lifespan.

In addition to being reaction-time based, which makes IAT scores susceptible for individual differences in processing speed, many IATs utilize verbal stimuli. Although the chosen words are simple and familiar (i.e., common in the specific language, Lane, Banaji, Nosek, & Greenwald, 2007; Riediger et al., 2014), differences in familiarity may arise from individual differences in word knowledge, or generally crystalline cognitive abilities. In a study with young students, word knowledge was not significantly related to IAT scores (von Stülpnagel & Steffens, 2010), which might result from reduced variance in word knowledge in the age-homogeneous, well-educated sample.

Fluid and crystalline cognitive abilities show distinct age-related trajectories. Most fluid cognitive abilities (e.g., processing speed, inhibition) peak in young adulthood and decline continuously afterwards, whereas most crystalline cognitive abilities (e.g., word knowledge) increase continuously during adolescence and young adulthood and stabilize during further adulthood, without or with small declines in late adulthood (Salthouse, 2010). Thus, with higher word knowledge with older age, even simple words in IATs may be more familiar and thus more easily categorized to the target or attribute category. Empirical tests of this assumption are still lacking and such tests need to ensure that stimuli are selected that are equally familiar to people of different age. Furthermore, no previous study has examined associations between crystalline or fluid-cognitive abilities and quad processes underlying IAT scores.

Quad Processes Contributing to IAT Scores

Thorough examinations on how quad processes relate to IAT scores are largely missing. Such analyses would facilitate understanding which and how strongly content-related associative and method-related processes contribute to reaction-time based IAT scores, which are mainly used. Among young students, both conventional and D-scores of a racial IAT were larger with larger AC and smaller abilities to overcome biases (OB), whereas D and G were not significantly associated with IAT scores (Conrey et al., 2005, Study 4). Stewart and colleagues (2009) observed larger scores on a racial IAT and weaker inhibitory control processes (conceptually similar to OB) with older age, but did not test associations between processes and IAT scores directly. Ito and colleagues (2015) found larger D-scores on a racial IAT with stronger associative and weaker control processes among undergraduate students. Since previous studies were restricted in age range, sample size, and/or examined processes, more thorough and comprehensive examinations on associations between quad processes and IAT scores are necessary. In addition, explorations on whether associations between quad processes and IAT scores vary with age would help to evaluate if IAT scores reflect different things for people of various ages—that is, if IAT scores capture associative and non-associative, method-related processes to a different degree for people of various ages.

The Present Study and Predictions

We applied a method and a topic IAT to examine how cognitive abilities and different quad processes contribute to IAT scores in age-heterogeneous samples due to cognition-related method effects. The application of two distinct IATs allowed for comparisons whether effects of cognitive abilities and quad processes are similar in both IATs presumably due to shared method effects.

For the method IAT, we chose the task switching ability IAT (Back et al., 2005) because it utilizes existing associations (i.e., letters with words, number with equations). Method IATs that utilize newly formed associations (e.g., between colors and size of abstract objects, Mierke & Klauer, 2003), seemed unsuitable because forming new associations becomes more difficult with older age (Salthouse, Schroeder, & Ferrer, 2004).

For the topic IAT, we chose the affect valence IAT (Riediger et al., 2014) because of the strong theoretical background on substantial age-related increases in evaluating positive affect (i.e., happy) as pleasant and negative affect (i.e., unhappy) as unpleasant (see Riediger et al., 2014 for theoretical reasoning). In addition, the affect valence IAT was developed with age-fair stimuli, which were selected in pilot studies to possess similar familiarity and meaning for different age groups (Riediger et al., 2014). Table A1 provides an overview on the design and procedure of both IATs.

First, we expected to replicate the age effects in the method and the topic IAT scores found for a subsample of the current study (Riediger et al., 2014). Specifically, we assumed to find no significant age differences in D-scores of the method IAT, where method-related age effect should be removed through the scoring algorithm; yet we assumed to observe larger D-scores of the topic IAT with older age, indicating substantive age effects in the target construct implicit affect valence (H1a). We further assumed that AC parameters are uncorrelated with age in the method IAT and larger with older age in the topic IAT (H1b), the latter again reflecting the assumed substantial age effects in the association of happy—pleasant and unhappy—unpleasant (Riediger et al., 2014). In both IATs, detection should be more pronounced, and overcoming bias should be less pronounced with age due to age-related increases in accuracy and decreases in inhibitory control (H1c, H1d).

Second, we examined the effects of processing speed and verbal ability on D-scores, and quad parameters of IATs. We assumed that processing speed is unrelated to D-scores of both method and topic IATs (H2a) because the D-scoring algorithm reduced speed effects in earlier studies with young adults (Cai et al., 2004; Greenwald et al., 2003). Furthermore, faster processing speed should predict better detection (D) and better overcoming bias (OB) in both IATs (H2b, H2c) because the faster processing speed indicates higher fluid cognitive abilities, which facilitate more correct detection and better inhibition (Stewart et al., 2009). In contrast, processing speed should not predict the AC parameters in both IATs because the AC parameter presumably reflects the substantial associative processes, which are independent from fluid-cognitive abilities (H2d). Concerning verbal abilities, we assumed that greater verbal abilities would predict a larger detection parameter D (but no other quad parameter) because even simpler words can be categorized correctly more easily with higher verbal abilities (H2e). Finally, we explored whether age moderated the associations between (a) cognitive abilities and IAT scores and (b) cognitive abilities and quad parameters to address whether method-related cognitive effects contribute to IAT performance to a different degree at various ages.

Third, we examined the associations between quad model parameters and IAT scores. For the method IAT, we assumed that no quad process parameter would predict the D-score because presumably all method effects were controlled (H3a). For the topic IAT, we assumed that AC and OB would predict the D-score because substantial differences should remain in the D-score of the topic IAT (H3b).We explored whether age moderates the associations to examine if IAT scores capture associative and non-associative, method-related processes to a different degree for people of various ages and because no previous study has addressed the age moderation.

Method

Participants

Participants included 549 individuals (51% female) from two studies (n_1=400, age range 12-88 years; n_2 = 149, age range 12-75 years).[2] The samples were combined to create a larger sample because samples greater than roughly 250 to 350 are needed to obtain stable correlational estimates for small and moderate effects (Schönbrodt & Perugini, 2013). The larger sample possessed a power of .93 to obtain significant effects *r* >. 15, α = .05 (Erdfelder, Faul, & Buchner, 1996). The subsamples were comparable in socio-demography and cognitive abilities[3], the materials and age ranges were identical in both studies, and additional analyses showed that result patterns were highly similar across subsamples (supplementary Tables S1-S3). On average, participants were 39.4 years old (*SD* = 20.7 years, range 11.6-88.1 years). The sample was distributed across adolescence (11-17 years: 21%), young (18-39 years: 31%), middle (40-59 years: 25%), and late adulthood (60-88 years: 23%). The highest level of education among the adult participants (> 18 years, n = 434) was: college or university degree: 28.6%; university-matriculation degree (Abitur): 20.0%; vocational degree: 39.6%; high-school degree (10 years): 7.8%; school degree (8 years): 3.2%; unreported: 0.7%. Among school students (n=115), 51.3% followed the university-matriculation track, 37.4% attended the 10-years school track, 5.2% attended the 8-years school track, and 6.1% did not report the school track. Highest attained education served as control variable in the analyses (1=8-years school degree, 2=10-years school degree, 3= vocational degree, 4= university-matriculation degree (Abitur); 5= college or university degree. All participants were native German speakers. Six participants (16 to 59 years, 1 woman) did not provide data for the IAT measures because of technical problems (n=4) or drop-out (n=2).

Procedure

Participants received information about the study procedures and provided written consent for participation. Then participants answered demographic questions, received standardized instructions from trained research assistants, and worked on the cognitive tasks and the IATs, with unrelated questionnaires in between. In study 1, all measures were presented on identical laptop computers (Fujitsu ESPRIMO Mobile D9510 with 14-in. monitors, 1024 x 768 pixels) at participants' homes. The IATs were administered in *DMDX* (Forster & Forster, 2003). In study 2, all measures were presented with *E-Prime* (Schneider, Eschmann, & Zuccolotto, 2007) on identical desktop computers (Wacom PL–521, 15 in. monitors, 1024 x 768 pixels) at the institute's laboratory. The ethics committee of the Max Planck Institute for Human Development, Berlin, Germany, approved both studies.

Measures

Method IAT. The task-switching ability IAT (Back et al., 2005) is supposed to be a content-free IAT that assesses method effects of the IAT (i.e., task switching costs). The IAT uses existing associations of letters with words and numbers with equations. The test follows the standard IAT procedure (e.g., Greenwald et al., 2003) and consists of five blocks, with 20 trials in the practice blocks 1, 2, and 4, and 80 trials in the combined blocks 3 and 5 (Table A1). Five stimuli represent each of the target categories labeled *letter* and *number* (e.g., N, K, 7, 3), as well as the attribute categories labeled *word* and *equation* (e.g., shirt, wall, 2 + 2 = 4, 8 – 6 = 2). Within each block, stimuli were presented in a randomized order in the middle of the screen (interstimulus interval = 33.5 ms) and repeated after all stimuli had been presented (i.e., repeated sampling without replacement). In the combined blocks 3 and 5, the target and the attribute stimuli alternated. The participants sorted the stimuli to the left or right categories using the answering keys "Q" and "Ü" which are on the far left and far right side of the German keyboard. During the entire task, participants kept their left and right index fingers on the keys. When an error occurred, a red "X" appeared on the middle of the screen and the next item only appeared when participants pressed the correct key.

Topic IAT. The topic IAT measured the implicit associations of the concepts happy-unhappy with pleasant-unpleasant (see Riediger et al., 2014 for development and validation). This IAT follows the same structure as the method IAT and also uses five stimuli for each of the target categories labeled *happy* and *unhappy* (e.g., joyful, lighthearted, sad, distressed) as well as for the attribute categories labeled *pleasant* and *unpleasant* (e.g., appealing, favored, repulsive, undesirable; see Riediger et al., 2014 for complete German and English stimuli). The stimulus words and category labels were presented in the same color to facilitate categorization, that is, target category labels and stimuli (e.g., happy, unhappy) were in blue font; attribute category labels and stimuli (e.g., pleasant, unpleasant) were in black font.

Perceptual speed. The Symbol-Digit Test (Lang, Weiss, Stocker, & Rosenbladt, 2007)—a modified version of the Digit Symbol Substitution Test of the Wechsler Adult Intelligence Scale-Revised (WAIS-R, (Wechsler, 1981)—was used to assess perceptual speed, an indicator of processing speed. The computerized version consisted of a look-up table with ten pairs of hieroglyphic-like symbols and digits (0 to 9). Participants typed the digits into consecutive boxes underneath the look-up table using the computers' keyboard (Lang et al., 2007). After a brief practice phase, participants worked on the task for 90 seconds, and entered on average 33.7 digits correctly (*SD* = 10.8). With older age, perceptual speed was significantly lower ($\beta_{age} = -.51$, $p < .01$, $\beta_{age^2} = -.08$, $p = .05$, $F(2, 544) = 110.50$, $p < .01$, $R^2 = .29$). Among adolescents (i.e., 11-20 years old, n=148), perceptual speed was higher with older age ($r = .24$, $p < .01$), as expected.

Word knowledge. A verbal test (MWT-A; Lehrl, Merz, Burkhard, & Fischer, 1991) was used to assess word knowledge as an indicator of verbal ability. Participants had to identify the real word among four fake words per item (e.g., balon, banoon, balloon, malloon, babboon). Without time pressure, participants worked on 40 items with increasing task difficulty. On average, participants identified 30.6 words correctly (*SD* = 3.3). With older age, word knowledge was significantly larger ($\beta_{age} = .66$, $p < .01$, $\beta_{age^2} = -.35$, $p < .01$, $F(2, 544) = 173.62$, $p < .01$, $R^2 = .39$). The age-related increase in word knowledge was less pronounced with higher age, as the quadratic age effect indicated.

Analytic Strategy

IAT scoring algorithms. We computed conventional scores and D_2-scores for both the method and the topic IATs. The computations used the reaction times in the combined blocks 3 and 5 and followed the prevailing recommendations (Greenwald et al., 2003; Richetin, Costantini, Perugini, & Schönbrodt, 2015): (a) For each trial the reaction time was determined until the correct response was provided (built-in error penalty); (b) reaction times < 300 ms and > 10,000 ms were winsorized; (c) for each block, the reaction times were summed across all trials; (d) the average reaction time of block 3 was subtracted from the average reaction time of block 5. These raw mean differences represent the conventional IAT score. The conventional IAT score was divided by the pooled standard deviation of blocks 3 and 5 to compute the D_2-score (Greenwald et al., 2003). Thirteen outliers ($> M + 3SD$) in the method IAT and four outliers in the topic IAT were winsorized to values representing the mean plus 3*SD*. Results for the conventional IAT scores are reported in Appendix B because it is widely accepted that the conventional IAT score contains method variance in addition to the intended content, and should not be used (Cai et al., 2004; Greenwald et al., 2003; Nosek et al., 2007).

Estimation of quad process parameters in IATs. We estimated the process parameters activation of associations (AC1, AC2), detection of correct response

(D), overcoming bias (OB), and guessing (G) for each participant separately using *HMMTree* (Stahl & Klauer, 2007). The program utilizes individual error rates and number of correct trials for each target and attribute category from the combined blocks 3 and 5 (e.g., Calanchini et al., 2014, p. 1288). Specifically, multinominal models are estimated that fit parameter-predicted error rates to observed error rates. For example, in the method IAT, whether the stimuli *shirt*, *wall*, etc. are correctly categorized to the category *word*, which shares the same response key with *number* in the incompatible block, depends on the four quad processes (Figure 1): the individual strength of the activated association *letters-words* (AC2), the detection of the correct category *word* (D), overcoming the bias for the other answer key (OB), and guessing (G). These parameters are estimated simultaneously based on equations describing the likelihood of a correct response or error for each stimulus in the combined blocks when all parameters work together to create a correct or false response. All correct responses and error rates from the combined blocks are used simultaneously (see Calanchini et al., 2014 for equations of the estimation approach).

The overall model fits were for the method IAT $\chi^2(3) = 894.30$, $p < .01$, $w = .10$, and for the topic IAT $\chi^2(3) = 717.88$, $p < .01$, $w = .09$—with significance being due to the large sample (see Calanchini et al., 2014, Study 2). Since quad parameters were estimated for each individual separately, parameters could not be computed when a participant made too few errors. This applied to 126 participants for the method IAT, and 37 participants for the topic IAT. In addition, three outliers (< M - 3SD) were identified for both method and topic IATs and winsorized to values representing the mean minus 3SD. Since the OB parameter showed bimodal distributions for the method and topic IATs (i.e., people received rather low or high estimates on overcoming bias), the OB parameters were dichotomized (0 thru 0.40 = 0; 0.60 thru 1 = 1; 15 values between .40 and .60 were set as missing). All analyses including the OB parameter were repeated with the original parameters, and yielded the same result pattern.

We computed multiple regression analyses to test hypotheses H2a-e (i.e., perceptual speed, verbal ability and interactions with age predicting D-scores and quad parameters of IATs) and H3a-b (i.e., quad process parameters and interactions with age predicting D-scores). To test whether the effects of the predictors differed between the method and the topic IAT, we specified multilevel models because IAT parameters were nested within participants. Weinfurt (2000) explained the appropriateness of multilevel modelling for repeated measures data.

Results

We first present age differences in D-scores and in quad process parameters estimated for both the method and topic IATs. Second, we show that cognitive abilities, indicated by perceptual speed and word knowledge, were associated with the predicted IAT effects and quad process parameters, but that the association did not differ significantly with age. Third, we demonstrate that quad process parameters predicted IAT D-scores, but the associations were again not significantly different with age. Table 1 shows descriptive information for the method and the topic IATs (D-score, error rates, quad parameters) and zero-order correlations with age, perceptual speed, word knowledge, and education as control variable.

Age Differences in IAT Effects and Quad Processes

IAT D-Scores. As expected (H1a), age was not significantly related to D-scores of the method IAT—despite the large sample size, which provides sufficient power to detect even relatively small effects (Table 1). No significant quadratic age effect occurred. Yet for the topic IAT, D-scores were significantly larger with older age with a slight levelling off in late adulthood as indicated by the significant quadratic effect ($\beta_{age} = .26$, $p < .01$, $\beta_{age^2} = -.10$, $p = .03$, $F(2, 539) = 16.49$, $p < .01$, $R^2 = .05$)—these effects indicated more pronounced implicit associations of happy—pleasant and unhappy—unpleasant compared to the inverted associations the older people were. As expected from previous studies, association with age were significantly larger for the conventional method IAT and topic IAT scores compared to the respective D-scores (Table B1; Method IAT: $z = 6.00$, $p < .01$; Topic IAT: $z = 5.35$, $p < .01$).

Quad process parameters. As predicted for the method IAT (H1b), no significant correlations between participants' age and the strength of the association (AC) for number—equation, the strength of the associations letter—word, the strength of the tendencies to overcome biases (OB) and to guess (G) were observed. With older age, detection process parameters (D) for the method IAT were larger (Table 1). In line with our predictions for the topic IAT (H1b,c), the estimates of happy—pleasant AC, unhappy—unpleasant AC, and D were larger with older age. Unexpectedly (H1d), OB was also significantly larger with older age. G estimates were not significantly related to participants' age (Table 1).

Cognitive Abilities Predict IAT Effects and Quad Processes

We computed regression models to predict D-scores or quad model parameters of both the method and the topic IAT (Table 2). Predictors in all models were age, perceptual speed, word knowledge (all sample-mean centered) and the interactions between age and perceptual speed or word knowledge. Attained education was included as control variable, but did not change the effects (supplementary tables S4). We used the bias-corrected, accelerated method (BCa) to compute coefficients and 95% confidence intervals based on 1000 bootstrap replications. Age effects from Table 2 are not interpreted because they are conditional effects and would refer to age differences in IAT effects among people with the same sample-average values in perceptual speed and word knowledge, which is unlikely since perceptual speed and word knowledge varied strongly with participants' ages.

IAT D-Scores. (H2a) Although the D-score algorithm should control for cognition-related methodological effects, D-scores of the method IAT were higher with greater perceptual speed, but this was the case irrespective of participants' age (i.e., age did not significantly moderate the association; Table 2). We explain in the discussion that higher method IAT D-scores with greater perceptual speed might be partly an artifact: faster responses often mean smaller standard deviations of reaction times, and since the standard deviation is the divisor for computing D-scores, dividing by smaller standard deviations leads to larger D-scores. Word knowledge showed no significant association with the D-score of the method IAT. In contrast, D-scores of the topic IAT were higher with slower perceptual speed and larger word knowledge.

Quad process parameters. In the method IAT, no significant model fit and no significant effects of perceptual speed and word knowledge were observed for the quad processes, except for detection (Table 2, left half). Regarding detection, people with faster perceptual speed or larger word knowledge showed better detection of correct and incorrect responses in the method IAT (H2b, H2e). In the topic IAT (H2d), the activated association between happy and pleasant was less pronounced with larger word knowledge, but not significantly associated with perceptual speed (Table 2, right half). Again, detection was predicted by perceptual speed and word knowledge, which mirrored the effects from the method IAT: With faster perceptual speed or larger word knowledge, detection of correct and incorrect responses was higher in the topic IAT (H2b, H2e). No significant age moderation occurred (Table 2, right half). In addition, neither perceptual speed nor word knowledge predicted differences in overcoming bias (OB) in the method or the topic IAT (H2c).

Quad Processes Predict IAT Scores

We computed regression analyses similar to the analyses in the previous section. We predicted IAT D-scores of the method and the topic IAT by age, the quad model parameters AC1, AC2, D, OB, G, and the respective interactions with age to examine which processes contribute to IAT scores, when method have been removed through the D-scoring algorithm (Table 3). Again, attained education served as control variable, but did not change the effects (supplementary tables S5).

Higher D-scores in the method IAT were predicted only by greater ability to detect correct and incorrect responses (Table 3, left column). This might be explained by the association between detection and the standard deviation of reaction times in the combined blocks, which is the divisor for computing D-scores: With better detection, that is, knowing the correct answer consistently faster, standard deviations were smaller (r = -.11, p =. 02). And smaller standard deviations (i.e., smaller divisors in computing D-scores) were associated with larger D-Scores (r = -.48, p <. 01). No further significant predictors or interactions with age were observed for the D-score of the method IAT (Table 3, left column).

The topic IAT was assumed to elicit a different picture because it should contain valid content variance after controlling method effects through using the D-score algorithm. Accordingly, stronger associations of happy—pleasant predicted larger D-scores of the topic IAT (Table 3, right column). The effect for the strength of the unhappy—unpleasant association was reduced from r = .13, p <. 01 for the conventional IAT score (Table B1) to b =.08, p = .09 for the D-score (Table 3, right column). In addition, greater ability to detect correct and incorrect responses and less guessing predicted larger D-scores in the topic IAT. Again, no significant age interactions occurred (Table 3, right column).

Discussion

The current study demonstrated that individual differences in cognitive abilities and quad process parameters contribute to IAT scores in age-heterogeneous samples, however not significantly differently with age. We elaborate below how the findings suggest that IATs seem suitable for age-heterogeneous studies from adolescence to old age, when IATs are constructed and analyzed appropriately. We first discuss age differences in IAT effects and quad processes before addressing effects of cognitive abilities. We also reflect on the meaning of D-scores and the role of errors when interpreting IAT effects.

Age Differences in IAT Processes and IAT Scores and Associations Among IAT Processes and IAT Scores

Our results suggest that D-scores effectively control age-related methodological effects of IATs because D-scores of the method IAT did not differ significantly with age, whereas the conventional scores were significantly larger, as in previous studies (Hummert et al., 2002). Consequently, age effects in D-scores of topic IATs can be interpreted largely as substantial age differences in implicitly measured concepts (e.g., Kunzmann & Thomas, 2014; Riediger et al., 2011). The more pronounced activated associations (AC) happy-pleasant and unhappy-unpleasant with older age further substantiate this conclusion. The conclusion that D-scores capture substantial content-related variance of implicit associations is also supported by regression results: Individual differences in activation of association (AC; marginally significant for unhappy-unpleasant association, significant zero-order correlation) predicted D-scores of the topic IAT, but not of the method IAT because no substantial variance is left after D-scoring the method IAT.

The ability to detect (D) the correct category for the stimuli was more pronounced with older age in both IATs, which agrees with studies using a racial IAT (Gonsalkorale et al., 2009; 2014). Also with better detection, D-scores were higher in both IATs, which was also reported in a study using a racial IAT (Conrey et al., 2005). This counterintuitive effect might be a partial artifact as lower detection (D) was associated with larger standard deviations—since some reactions take longer when the correct answer is difficult to detect—and thus to larger denominators which create smaller scores when computing D-scores. A similar effect was observed in an age-homogeneous sample, where D-scores were unexpectedly smaller under cognitive load compared to a control condition (Schmitz, Teige-Mocigemba, Voss, & Klauer, 2013). Schmitz et al. (2013) equally explained this unexpected effect as artifact due to larger standard deviations, and thus larger denominators under cognitive load.

The findings on how quad process parameters (based on errors) contribute to IAT-scores (based on reaction times) are especially relevant as previous studies have rarely examined the interrelations (e.g., Gosalkorale et al., 2009; 2014). Only a small study (n=42) with students showed that larger AC and smaller OB contributed to larger scores in a racial stereotypes IAT (Conrey et al., 2005). We observed similar associations for AC with a much larger, age-heterogeneous sample and concerning a different domain, that is, affective attitudes compared to racial stereotypes.

Similarly, previous studies have not examined how individual differences in quad parameters derived from different IATs co-vary (Appendix C). Theoretically, the detection (D) parameter should be domain-general, i.e., strongly correlated across different IATs (Calanchini et al., 2014). As predicted, we observed a strong positive correlation: people who had better detection of correct responses in the method IAT also showed better detection in the topic IAT. The positive associations among detection parameters suggest a common influence of cognitive abilities, as observed in the current study.

Cognitive Abilities Predict D-Scored IAT Effects and Quad Processes Similarly Across Age

We examined effects of perceptual speed and word knowledge because IATs are reaction-time based measures that often utilize verbal stimuli. Consistent with previous studies (Klauer et al., 2010; von Stülpnagel & Steffens, 2010), greater perceptual or cognitive speed was related to lower conventional IAT scores (Table B1). Appendix B additionally showed that this association was more pronounced with older age. Thus, individual differences in perceptual speed contributed more to conventional IAT score among older people compared to younger people. The finding substantiates previous skepticism about the usage of the conventional scoring of IATs (Greenwald et al., 2003; Hummert et al., 2002). The D-score presumably controls method effects and individual differences in cognitive abilities (Cai et al., 2004; Nosek et al., 2007). Yet, greater perceptual speed predicted higher D-scores in the method IAT. We return to this unexpected finding when reflecting on *A Better Understanding of IAT Scores*.

In contrast to the method IAT, slower perceptual speed and greater word knowledge predicted larger D-scores in the topic IAT (i.e., more strongly associating happy with pleasant and unhappy with unpleasant compared to the contrary associations). This could indicate that implicit affect valence differs with cognitive abilities. Yet, neither perceptual speed nor word knowledge predicted individual differences in the AC or OB parameters in the topic IAT (except for a small correlation between lower word-knowledge and stronger unhappy-unpleasant associations). These findings correspond with a previous study, which also showed no significant associations between cognitive abilities and associative process parameters in a racial IAT, but also larger D-scores with lower cognitive abilities (i.e., task shifting ability, Ito et al., 2015). An alternative interpretation is that D-scores do not fully control method effects—yet this assumption awaits further examination specifically regarding the impact of cognitive control (Hilgard, Bartholow, Dickter, & Blanton, 2015; Ito et al., 2015).

Finally, faster perceptual speed and greater word knowledge predicted better detection of correct and incorrect responses in both method and topic IATs. Thus, although the stimuli in both IATs were selected for simplicity and familiarity, faster and more literate people were still better at categorizing the stimuli correctly.

In summary, although individual differences in cognitive abilities contributed to IAT D-scores, as in age-homogenous samples (Ito et al., 2015; Klauer et al., 2010;

Siegel, 2012), and to the detection process parameter, the associations were not significantly moderated by age. This suggests that the effects of cognitive abilities on IAT scores are similar for people of different ages—making D-score IAT effects age-fair. And as discussed in the previous section, the associations between quad processes and D-scores were also not significantly different with age. If the IAT D-scores were more "saturated" with individual differences in cognitive abilities or in some of the quad processes among certain age groups, usage in age-heterogeneous studies would be questionable.

A Better Understanding of IAT Scores

Previous studies noted that whereas conventional IAT scores cannot be interpreted as solely measuring individual differences in association strength due to containing method effects related to processing speed or general intelligence (e.g., Fiedler, Messner & Bluemke, 2006; Klauer et al., 2010), D-scores of IATs should largely control such method effects. The question remains how D-scoring controls method effects and to what extent. Quad process analyses showed that activated associations (AC) contribute as expected to individual differences in a substantial IAT (i.e., on implicit affect valence), but not the D-scored method IAT. Unexpectedly, individual differences in detection (D) contributed to D-scores in both the method and the topic IATs. Detection is assumed to be a domain-general cognitive process that correlates positively across IATs, and we observed this positive association (Appendix C) in accordance with previous studies (Calanchini et al., 2014). We thus speculate that the D-score does not fully control for method effects (see Back et al., 2005 and Ito et al., 2015 for similar conclusions in age-homogeneous samples) and may even introduce method effects through the division by the standard deviation, which varies with cognitive abilities, as discussed before.

At the same time, individual differences in overcoming biases of activated associations (OB) predicted conventional scores of the topic IAT, but not D-scores. Previous studies assumed that OB could be an integral part of implicit attitudes because lower inhibition of activated associations (i.e., OB) contributes to stronger implicit attitudes (Gonsalkorale et al., 2009; 2014)—yet these studies did not directly test associations between OB and IAT-scores. Originally, OB is conceptualized as an inhibitory process that depends on the strength of the activated bias (i.e. association), cognitive abilities, and motivation (Conrey & Sherman, 2006a, b). Accordingly, we observed no significant correlations between OB parameters of the method and the topic IAT (Appendix C). The insignificant correlation is consistent with a previous study with equally diverging content of IATs (Calanchini et al., 2014, study 1c), and suggests that OB parameters reflect domain-specific processes in addition to domain-general inhibitory processes.

In sum, the reduced associations between D-scores and OB (i.e., inhibition of activated associations) as well as the more pronounced association between D-scores and D (i.e., detection of correct responses)—compared to conventional scores—suggest that the D-score perhaps controls valid variance (i.e., OB) and also method variance more strongly than intended. Clearly future research with additional heterogeneous samples and further IATs is needed to better understand the associations among reaction-time based IAT scores and error-based process estimates.

Future Directions

Although the current research studied age differences in IAT performance using a large, age-and gender-stratified sample and sophisticated multinomial process models, limitations and future directions need to be discussed. First, we assessed fluid and crystalline intelligence using perceptual speed and word knowledge to capture domains of cognitive functioning that show strong age differences (Salthouse, 2010). Future studies might want to assess further, domain-specific cognitive variables, such as working memory capacity or inhibitory control, either behaviorally (Ito et al., 2015) or based on cortical activity (Hilgard et al., 2015). Such studies might then be able to explain which cognitive abilities contribute to individual differences in OB and G because these parameters were not significantly predicted by perceptual speed and word knowledge in the current study.

Second, we applied the quad process model because it was developed specifically to distinguish processes underlying IAT performances (Sherman, 2006a, b). Further process-dissociation models such as the ReAL model (Meissner & Rothermund, 2013) or the diffusion model (Klauer et al., 2007) differ in the modeled processes and necessitate even larger numbers of errors to achieve reliable estimates. In the current study, the ReAL and the diffusion model were not applied because too few errors occurred for estimating these models. The number of errors, however, was comparable with previous studies using IATs (e.g., Calanchini et al., 2014; Gonsalkorale et al., 2011, 2014).

Third, a generalization that other indirect tests, such as affective priming (deHouwer, Teige-Mocigemba, Spruyt, & Moors, 2009) or dot-probe (Isaacowith & Choi, 2011), are equally suited for age-heterogeneous samples is premature, especially since implicit measures are often only weakly correlated (Ito et al., 2015). Clearly, the applicability of these tasks in heterogeneous samples and their associations with cognitive abilities need to be shown, before age differences in other indirect tests can be interpreted as substantial age differences in the measured constructs.

Conclusion

IATs are increasingly used in various domains. The current findings revealed that IATs can be meaningfully used and interpreted in age-heterogeneous samples from adolescence to old age since content-related quad process parameters were not significantly related to cognitive processing speed and further associations with cognitive abilities did not vary significantly with age. At the same time, the findings underscore that the conventional IAT score should not be used, and process parameters should be estimated in addition to the D-score because the D-score does not control for all method effects and individual differences in cognitive abilities. The findings suggest that IATs can be applied successfully in age-heterogeneous samples to examine developmental changes of implicit characteristics.

References

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the implicit association test. *Experimental Psychology, 52,* 167-179. doi: 10.1027/1618-3169.52.3.167

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*, 53-58. doi: 10.1111/j.1467-9280.2005.01664.x

Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's D measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition, 22*(6), 673-684. doi: 10.1521/soco.22.6.673.54821

Calanchini, J., Sherman, J. W., Klauer, K. C., & Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. P*ersonality and Social Psychology Bulletin, 40,* 1285-1296. doi: 10.1177/0146167214540723

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89*, 469-487. doi: 10.1037/0022-3514.89.4.469

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*, 347-368. doi: 10.1037/a0014211

Dunham, Y., Baron, A. S., & Banaji, M. R. (2015). The development of implicit gender attitudes. *Developmental Science, n*, nn. doi: 10.1111/desc.12321

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers, 28,* 1-11.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278. doi: 10.1146/annurev.psych.59.103006.093629

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the “I”, the “A”, and the “T”: A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology, 17*, 74-147. doi: 10.1080/10463280600681248

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, *35*, 116-124. doi: 10.3758/BF03195503

Gonsalkorale, K., Sherman, J. W., Allen, T. J., Klauer, K. C., & Amodio, D. M. (2011). Accounting for successful control of implicit racial bias: The roles of association activation, response monitoring, and overcoming bias. *Personality and Social Psychology Bulletin, 37,* 1534-1545. doi:10.1177/0146167211414064

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2009). Aging and prejudice: Diminished regulation of automatic race bias among older adults. *Journal of Experimental Social Psychology*, *45*, 410-414. doi: 10.1016/j.jesp.2008.11.004

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science, 5*, 271-278. doi: 10.1177/1948550613499239

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology*. doi: 10.1037/pspa0000016

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: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216.

Hilgard, J., Bartholow, B. D., Dickter, C. L., & Blanton, H. (2015). Characterizing switching and congruency effects in the Implicit Association Test as reactive and proactive cognitive control. *Social Cognitive and Affective Neuroscience, 10*, 381-388. doi:10.1093/scan/nsu060

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369-1385.

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.

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive function in implicit racial bias. *Journal of Personality and Social Psychology, 108*, 187-218. doi:10.1037/a0038557

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. *The Quarterly Journal of Experimental Psychology, 63*, 595-619. doi: 10.1080/17470210903076826

Kunzmann, U., & Thomas, S. (2014). Multidirectional age differences in anger and sadness. *Psychology and Aging, 29*, 16-27. doi:10.1037/a0035751

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

Lang, F. R., Weiss, D., Stocker, A., & Rosenbladt, B. V. (2007). Assessing cognitive capacities in computer-assisted survey research: Two ultra-short tests of intellectual ability in the Germany Socio-Economic Panel (SOEP). *Schmollers Jahrbuch: Journal of Applied Social Science Studies, 127*, 183–192.

Lehrl, S., Merz, J., Burkhard, G., & Fischer, S. (1991). Mehrfachwahl – Wortschatz – Intelligenztest, MWT-A [Multiple choice vocabulary intelligence test] Göttingen: Hogrefe.

Luong, G., Wrzus, C., Wagner, G. G., & Riediger, M. (2016). When bad moods may not be so bad: Valuing negative affect is associated with weakened affect-health links. *Emotion, 16*, 387-401. doi: 10.1037/emo0000132

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483-510.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of Personality and Social Psychology, 104*, 45-69. doi: 10.1037/a0030734

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the implicit association test. *Journal of Personality and Social Psychology, 85*, 1180. doi: 10.1037/0022-3514.85.6.1180

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In NN (Ed.), *Automatic processes in social thinking and behavior* (pp. 265-292). nn: nn.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36-88. doi: 10.1080/10463280701489053

Payne, B. K. (2008). What mistakes disclose: A process dissociation approach to automatic and controlled processes in social psychology. *Social and Personality Psychology Compass*, *2*, 1073-1092. doi: 10.1111/j.1751-9004.2008.00091.x

Richetin, J., Costantini, G., Perugini, M., & Schönbrodt, F. D. (2015). Should we stop looking for a better scoring algorithm for handling Implicit Association Test data? Test of the role of errors, extreme latencies treatment, scoring formula, and practice trials on reliability and validity. *PLoS ONE, 10*(6), e0129601. doi: 10.1371/journal.pone.0129601

Riediger, M., Wrzus, C., & Wagner, G. G. (2014). Happiness is pleasant, or is it? Implicit representations of affect valence are associated with contrahedonic motivation and mixed affect in daily life. *Emotion, 14*, 950-961. doi: 10.1037/a0037711

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*, 318-339. doi: 10.1037/0033-295X.95.3.318

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International neuropsychological Society, 16*(05), 754-760. doi: 10.1017/S1355617710000706

Salthouse, T. A., Schroeder, D. H., & Ferrer, E. (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. *Developmental Psychology, 40*, 813-822.

Schmitz, F., Teige-Mocigemba, S., Voss, A., & Klauer, K. C. (2013). When scoring algorithms matter: Effects of working memory load on different IAT scores. *British Journal of Social Psychology, 52*, 103-121. doi: 0.1111/j.2044-8309.2011.02057.x

Schneider, W., Eschmann, A., & Zuccolotto, A. (2007). E-Prime. Pittsburg, PA: Psychology Software Tools, Inc.

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

Sherman, J. W. (2006a). On building a better process model: It's not only how many, but which ones and by which means? *Psychological Inquiry, 17*, 173-184. doi: 10.1207/s15327965pli1703_3

Sherman, J. W. (2006b). Clearing up some misconceptions about the quad model. *Psychological Inquiry, 17*, 269-276.

Siegel, E. F., Dougherty, M. R., & Huber, D. E. (2012). Manipulating the role of cognitive control while taking the implicit association test. *Journal of Experimental Social Psychology, 48*, 1057-1068. doi: 10.1016/j.jesp.2012.04.011

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods, 39*, 267-273.

Stewart, B. D., von Hippel, W., & Radvansky, G. A. (2009). Age, race, and implicit prejudice using process dissociation to separate the underlying components. *Psychological Science, 20*, 164-168. doi: 10.1111/j.1467-9280.2009.02274.x

von Stülpnagel, R., & Steffens, M. C. (2010). Prejudiced or just smart? *Zeitschrift für Psychologie/Journal of Psychology, 218*, 51-53. doi: 10.1027/0044-3409/a000008

Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: A meta-analysis. *Psychology and Aging, 26*, 15-20. doi: 10.1037/a0020912

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale—Revised. New York: Psychological Corporation.

Weinfurt, K. P. (2000). Repeated measures analysis: Anova, Manova, and HLM. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 65-98). Washington, DC: American Psychological Association.

Footnotes

[1] The studies by Hummert and colleagues (2002) were published before Greenwald and colleagues (2003) proposed new scoring algorithms. Thus, Hummert and colleagues did not use the D_2-scores, but a conceptually similar scoring algorithm because reaction times were z-standardized, i.e., divided by the standard deviation.

[2] Parts of the data of the first sample are published in Riediger et al. (2014). The article focused on a different research question, that is, associations between implicit affect valence and affective experiences in daily life. It reported age-related effects in conventional and D-scores of both IATs used in the current study. In contrast, the current study focuses on associations among age, cognitive abilities, process parameters and D-scores of IAT, and if the associations vary with age. The current study includes cognitive abilities as new variables, quad models as new analytic approaches, and new data from another sample.

[3] Samples did not differ significantly by gender, $\chi^2 = 0.39$, $p = .53$; age, $t(1,547) = 0.83$, $p = .41$; or word knowledge, $t(1,546) = 1.69$, $p = .09$. The participants in study 2 had faster perceptual speed, $t(1,547) = 9.24$, $p < .01$, Cohen's $d = 0.90$ (see Table S1 for descriptive statistics). We tested, but mainly observed no significant differences between samples regarding the central analyses (see supplement Tables S2-S3). Of 82 *sample-by-predictor* interaction terms in eight regression models with significant model fit, only two interaction effects with *sample* were statistically significant ($p < .01$). Once, an additional significant effect of word-knowledge on the AC2 parameter of the method IAT emerged in the smaller sample, and once the effect of guessing on D-scores of the topic IAT was only significant in the larger sample. We therefore conclude that results from the combined sample are a conservative estimate of the effects.

Table 1

Descriptive statistics of IAT parameters and correlation with age and cognitive abilities

	M (*SD*)	Age *r*	Education *r*	Perceptual speed *r*	Word knowledge *r*
Method IAT					
IAT D_2-score	1.03 (0.50)	.07	.19**	.17**	.07
Number of errors combined block 3	2.94 (3.26)	-.37**	-.14**	.12**	-.33**
Number of errors combined block 5	7.72 (6.83)	-.16**	-.21**	-.01	-.30**
Activation of associations *number & equation* (AC1)	0.25 (0.30)	.07	-.01	-.09	.03
Activation of associations *letter & word* (AC2)	0.24 (0.32)	.08	.01	-.09	.06
Detection (D)	0.87 (0.11)	.27**	.19**	-.01	.30**
Overcoming bias[a] (OB)	0.62 (0.45)	.04	.06	-.05	.09
Guessing (G)	0.73 (0.29)	-.06	.00	-.01	.02
Topic IAT					
IAT D_2-score	0.97 (0.41)	.22**	-.04	-.20**	.21**
Number of errors combined block 3	4.08 (4.69)	-.37**	-.17**	.05	-.34**
Number of errors combined block 5	11.24 (10.16)	-.05	-.25**	-.15**	-.22**
Activation of associations *unhappy & unpleasant* (AC1)	0.16 (0.26)	.14**	-.05	-.03	.03
Activation of associations *happy & pleasant* (AC2)	0.34 (0.35)	.18**	-.00	-.15*	.07
Detection (D)	0.83 (0.16)	.21**	.20**	.06	.29**
Overcoming bias[a] (OB)	0.58 (0.46)	.10*	.07	-.03	.07
Guessing (G)	0.34 (0.29)	-.01	-.04	.05	-.06

Note. [a] Correlations are nearly identical if modeled with original, continuous variable. [b]Based on individuals (method IAT n = 417, topic IAT n = 505), for whom parameters were estimated to be trustworthy (no fail code, see method section). * $p < .05$, ** $p < .01$.

Table 2

IAT D-Scores and Process Parameter Predicted by Age, Cognitive Ability, and the Interactions (Standardized coefficients from multiple regression analyses)

	Method IAT						Topic IAT					
	D_2-score	Number-equation (AC1)[a]	Letter-word (AC2)[a]	Detection (D)[a]	Overcoming bias (OB)[a,b]	Guessing (G)[a]	D_2-score	Unhappy-unpleasant (AC1)[a]	Happy-pleasant (AC2)[a]	Detection (D)[a]	Overcoming bias (OB)[a,b]	Guessing (G)[a]
Age	$.24_a$**	$.05_a$	$.05_a$	$.28_a$**	$-.002_a$	$-.18_a$*	$.05_b$	$.22_a$**	$.18_a$**	$.18_a$**	$.02_a$*	$.08_b$
	[.10;.38]	[-.10;.19]	[-.09;.20]	[.14;.42]	[-.02;.01]	[-.35;-.01]	[-.09;.19]	[.10;.37]	[.05;.32]	[.04;.33]	[.01;.03]	[-.08;.22]
Perceptual speed	$.30_a$**	$-.07_a$	$-.06_a$	$.17_a$**	$-.01_a$	$-.09_a$	$-.15_b$**	$.07_a$	$-.05_a$	$.20_a$**	$.01_a$	$.07_a$
	[.20;.40]	[-.19;.06]	[-.18;.07]	[.03;.31]	[-.03;.01]	[-.22;.04]	[-.25;-.05]	[-.04;.17]	[-.17;.06]	[.09;.31]	[-.02;.03]	[-.05;.18]
Word knowledge	$.01_a$	$-.01_a$	$.03_a$	$.18_a$**	$.05_a$	$.13_a$	$.14_a$**	$-.13_a$*	$-.03_a$	$.22_a$**	$-.01_a$	$-.11_b$
	[-.10;.12]	[-.16;.13]	[-.10;.17]	[.03;.34]	[-.04;.15]	[-.02;.27]	[.02;.26]	[-.26;.01]	[-.14;.09]	[.08;.36]	[-.09;.07]	[-.25;.02]
Age × speed	$.06_a$	$-.001_a$	$.08_a$	$.10_a$	$.04e^{-3}{}_a$	$-.07_a$	$-.04_a$	$.02_a$	$.08_a$	$-.05_b$	$.02e^{-2}{}_a$	$-.03_a$
	[-.03;.14]	[-.12;.12]	[-.02;.18]	[-.03;.22]	[-.00;.001]	[-.18;.05]	[-.12;.05]	[-.06;.11]	[-.02;.18]	[-.13;.04]	[-.00;.001]	[-.13;.07]
Age × knowledge	$.03_a$	$-.02_a$	$.02_a$	$.01_a$	$.01e^{-2}{}_a$	$.05_a$	$-.04_a$	$-.10_a$	$.02_a$	$.02_a$	$-.08e^{-2}{}_a$	$-.05_a$
	[-.07;.12]	[-.13;.09]	[-.08;.12]	[-.12;.13]	[-.01;.01]	[-.07;.18]	[-.15;.06]	[-.20;.01]	[-.08;.12]	[-.10;.14]	[-.00;.003]	[-.16;.07]
Model fit												
F(5,536)	7.57**	0.73	1.41	11.76**	3.52	1.42	9.40**	3.30**	4.38**	12.36**	8.47	1.03
R^2	.06	.01	.01	.13	.01	.01	.07	.03	.04	.12	.03	.01

Note. All predictors, including multipliers in interaction terms, were centered on the sample mean. Parameters with different subscripts for the same predictor and outcome variables differ between both IATs with $p < .05$. Confidence intervals were computed based on Bias corrected and accelerated method (BCa) with 1000 repetitions. [a]Models predicting quad parameter possessed df = 417 for method IAT and df = 505 for topic IAT due to persons for whom parameters could not be estimated (see *Analytic Strategy* section). [b]For parameter overcoming bias, logistic regression were computed and unstandardized coefficients, χ^2, as well as Nagelkerke R^2 reported; the results are the same when computed with linear OLS regression. * $p < .05$, ** $p < .01$.

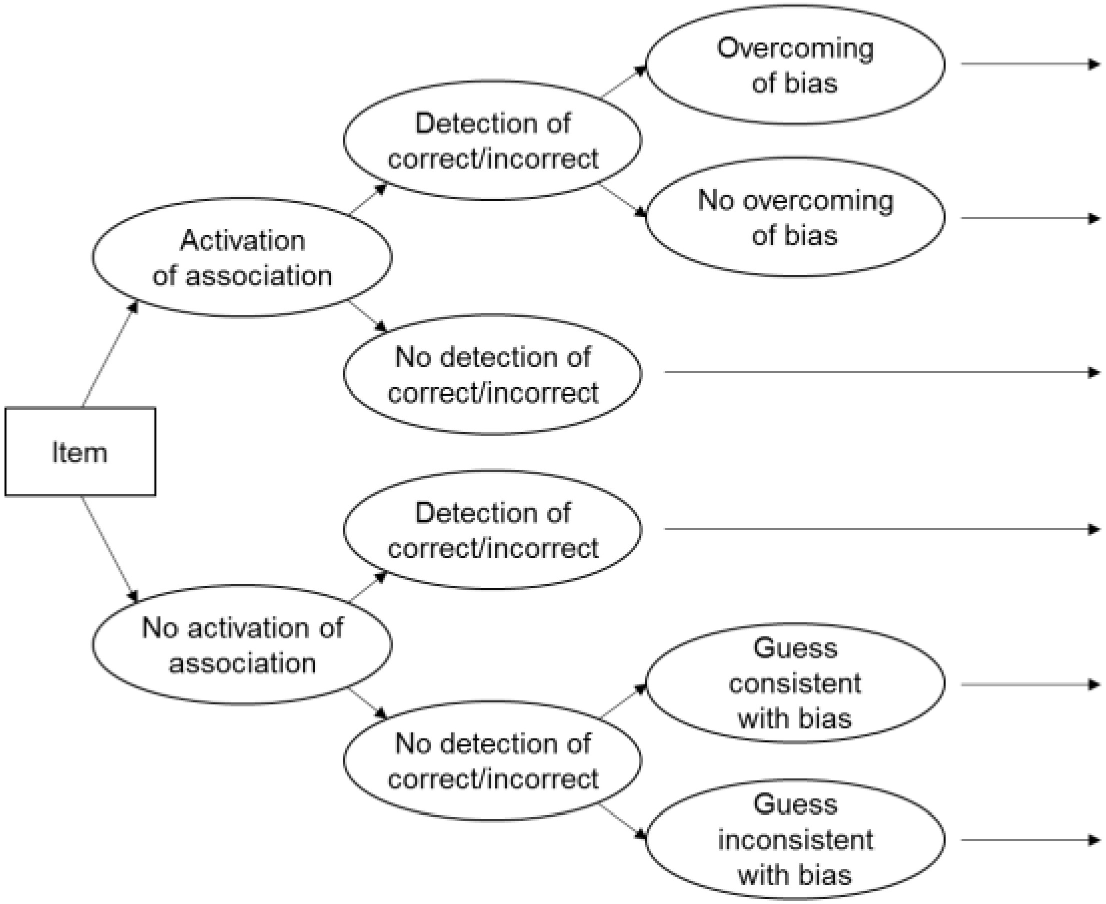
Table 3

IAT D_2- Scores Predicted by Age, IAT Process Parameter, and the Interactions (Standardized coefficients from multiple regression analyses)

	Method IAT		Topic IAT	
	β	[95% CI]	β	[95% CI]
Age	$.11_a$	[-.06;.27]	$.11_a$	[-.09;.31]
Association 1 (AC1)[a]	$.03_a$	[-.10;.15]	$.08_a$	[-.04;.18]
Association 2 (AC2)[a]	$-.06_a$	[-.17;.05]	$.19_b$**	[.08;.31]
Detection (D)	$.21_a$**	[.11;.31]	$.27_a$**	[.14;.41]
Overcoming bias (OB)[b]	$.02_a$	[-.12;.15]	$-.02_a$	[-.13;.09]
Guessing (G)	$.06_a$	[-.04;.16]	$-.12_b$*	[-.21;-.02]
Age × AC1	$-.07_a$	[-.19;.05]	$-.06_a$	[-.14;.03]
Age × AC2	$.01_a$	[-.10;.10]	$.01_a$	[-.09;.11]
Age × D	$-.06_a$	[-.16;.04]	$-.03_a$	[-.17;.11]
Age × OB[b]	$-.11_a$	[-.31;.08]	$-.001_a$	[-.16;.16]
Age × G	$-.01_a$	[-.10;.10]	$.06_a$	[-.03;.15]
Model fit				
$F(11,292)$	2.52**		8.00**	
R^2	.07		.17	

Note. All predictors, including multipliers in interaction terms, were centered on the sample mean. Parameters with different subscripts for the same predictor differ between both IATs with $p < .05$. Confidence intervals were computed based on Bias corrected and accelerated method (BCa) with 1000 repetitions. [a]For method IAT: association 1 = number equation, association 2 = letter-word; For topic IAT: association 1 = unhappy-unpleasant, association 2

= unhappy-unpleasant. [b] the same pattern of results was obtained when computed with original continuous variable. * $p < .05$, ** $p < .01$.



Compatible		Incompatible	
Left	Right	Left	Right
+	+	+	+
+	+	-	-
+	+	-	-
+	+	+	+
-	+	-	+
+	-	+	-

Figure 1. Graphical representation of the quad process model (Conrey et al., 2005; Sherman, 2006), where each path represents a likelihood. Parameters with paths leading to them are conditional upon all preceding parameters. The table presents correct (+) and incorrect responses when categorizing stimuli as result of different quad processes in the compatible and the incompatible block.

Appendix A

Overview of the applied IATs

Table A1

Design of Method and Topic IAT used in the Current Study

			Method IAT (Back et al., 2005)		Topic IAT (Riediger et al., 2014)	
Block	*N* trials	Task	Left key	Right key	Left key	Right key
1	20	Target categorization (practice)	Letter	Number	Happy	Unhappy
2	20	Attribute categorization (practice)	Word	Equation	Pleasant	Unpleasant
3	80	Combined compatible categorization (test)	Letter and word	Number and equation	Happy and pleasant	Unhappy and unpleasant
4	20	Reversed target categorization (practice)	Number	Letter	Unhappy	Happy
5	80	Combined incompatible categorization (test)	Number and word	Letter and equation	Unhappy and pleasant	Happy and unpleasant

Note. See Back et al., 2005 and Riediger et al., 2014 for complete German and English stimuli.

Appendix B

Conventional IAT score

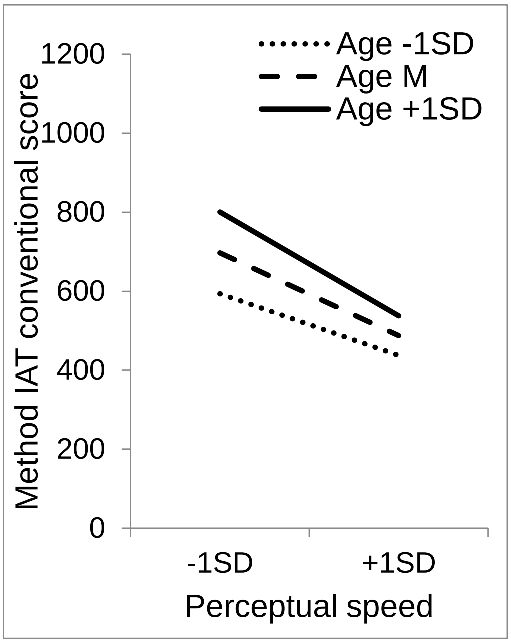
Since the conventional IAT score has been criticized repeatedly for containing method effects in addition to substantial individual differences in attitudes or traits (e.g., (Greenwald et al., 2003; McFarland & Crouch, 2002; Nosek et al., 2007), we included the suggested D_2-score in our main analyses. Here, we show that we replicated previous findings regarding the convential IAT scores correlating strongly with age and cognitive abilities, especially cognitive speed (Klauer et al., 2010; Siegel et al., 2012). In addition, all quad parameters contribute to the conventional score of the topic IAT, but not of the method IAT, which aims at assessing task switching costs, and thus should not show meaninful associations with the quad parameters.

Table B1

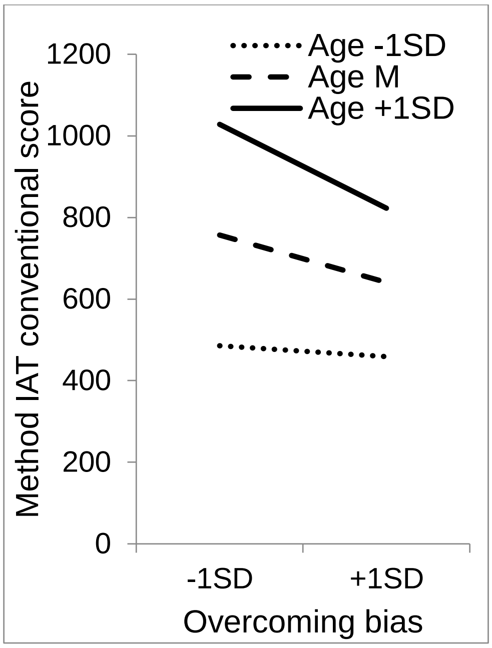
Zero-order Correlations of Conventional IAT Scores with Age, Cognitive Abilities, and IAT Process Parameter

	Conventional Score Method IAT	Conventional Score Topic IAT
Age	.41**	.50**
Perceptual speed	-.43**	-.46**
Word knowledge	.16**	.22**
IAT D_2-score	.06	.46**
Errors in combined block 3	-.24**	-.32**
Errors in combined block 5	.11*	.07
Association 1 (AC1)	.12*	.13*
Association 2 (AC2)	.02	.29**
Detection (D)	.07	.14**
Overcoming bias (OB)	-.06	.14**
Guessing (G)	-.06	-.10*
M (*SD*)	595.81 (306.74)	1055.85 (651.70)

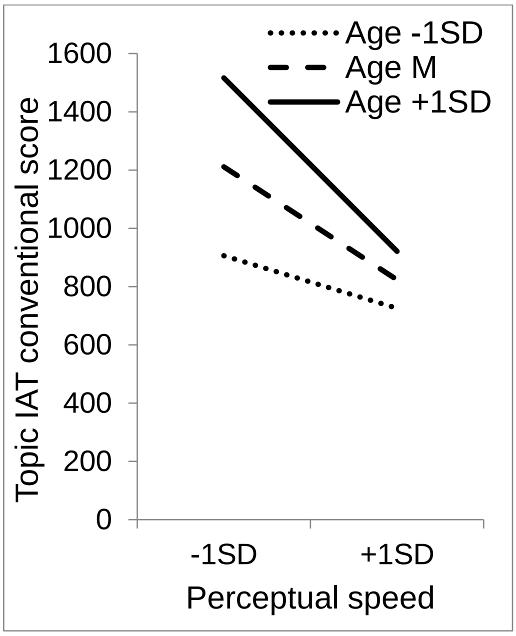
Note. Method IAT: letter-word, number-equation (Back et al., 2005); Topic IAT: happy-pleasant, unhappy-unpleasant (Riediger et al., 2014). Multiple regression results were nearly identical with the zero-order correlations presented in the table, except that associations between word knowledge and IAT scores were diminished and statistically non-significant, when age was controlled statistically. In addition, age moderated the associations between the method IAT scores and perceptual speed (β = -.08, $p < .05$) or overcoming bias (β = -.20, $p < .01$), respectively (see Figures B1, B2). Similarly, age moderated the association between the topic IAT scores and perceptual speed (β = -.15, $p < .01$, Figure B3). * $p < .05$, ** $p < .01$.



B1



B2



B3

Appendix C

Associations between the Method and the Topic IAT

Table C1

Correlations among Method and Topic IATs regarding Conventional scores, D-scores, Errors, and IAT Process Parameter

	Zero-order correlation among method and topic IAT *r*
Conventional IAT Score	.56**
IAT D_2-score	-.10*
Errors in combined block 3	.60**
Errors in combined block 5	.46**
Association 1 (AC1)	.04
Association 2 (AC2)	-.07
Detection (D)	.49**
Overcoming bias (OB)	-.04
Guessing (G)	-.02

Note. Method IAT: letter-word, number-equation (Back et al., 2005); Topic IAT: happy-pleasant, unhappy-unpleasant (Riediger et al., 2014). * $p < .05$, ** $p < .01$.