



# Finding an interaction between Stroop congruency and flanker congruency requires a large congruency effect: A within-trial combination of conflict tasks

Alodie Rey-Mermet<sup>1</sup>

Published online: 23 January 2020  
© The Psychonomic Society, Inc. 2020

## Abstract

Responding to a conflict is assumed to trigger attentional-control processes—that is, processes that enable us to activate goal-relevant information and to inhibit irrelevant information. Typically, conflict is induced in tasks, such as the Stroop task (which requires identifying the color of color words) or the flanker task (which requires identifying a central character among flankers). Combining the conflicts within the same trial has been found to result in an interaction in reaction times (RTs), suggesting a generalization of attentional control. However, this interaction was observed when the congruency effect was substantial—that is, when the RT difference between incongruent trials (e.g., the word “green” printed in red for the Stroop task) and congruent trials (e.g., the word “red” printed in red) was large. The purpose of the present study was to investigate whether a large congruency effect is the necessary condition for observing the interaction. To this end, Stroop and flanker tasks were combined, and participants were asked to respond to the color of the central letter/word while ignoring the flanking letters/words. The magnitude of the congruency effect was increased: (a) by testing older adults (Experiment 1), (b) by manipulating the proportion of trials in which participants were asked to respond to the word meaning (Experiment 2), and (c) by using vocal responses (Experiment 3). The results showed an interaction when the Stroop congruency effect was large. Therefore, such interactions can be used to validate or invalidate theoretical explanations only when the precondition—a large congruency effect—is fulfilled.

**Keywords** Executive functions · Cognitive control · Attentional control · Aging · Bayesian hypothesis testing approach · State-trace analysis

In cognitive psychology, responding to a task with a conflict has been assumed to induce an adjustment of attentional control. This adjustment allows us to activate goal-relevant features while inhibiting irrelevant ones (e.g., Botvinick, Braver, Barch, Carter, & Cohen, 2001). A current debate is about whether these control processes can generalize across conflicts (e.g., Boy, Husain, & Sumner, 2010; Kane et al., 2016; Miyake & Friedman, 2012; Rey-Mermet, Gade, & Oberauer, 2018; Rey-Mermet, Gade, Souza, von Bastian, & Oberauer, 2019; Stahl et al., 2014). Empirical support for the generalization of control processes comes from studies showing an interaction in reaction times (RTs) between different conflicts when the conflicts were combined within the same trial (Hommel, 1997, Experiment 2; Rey-Mermet & Gade, 2016;

Stoffels & van der Molen, 1988, Experiment 2; Wendt, Kluwe, & Peters, 2006, Experiment 2a; but see Akçay & Hazeltine, 2011; Egner, Delano, & Hirsch, 2007; Hommel, 1997, Experiment 1; Kornblum, 1994; Simon & Berbaum, 1990; Wendt et al., 2006, Experiment 2b; Stoffels & van der Molen, 1988, Experiment 1). A closer inspection of these studies revealed, however, that the interaction was mainly observed when the interference induced by the conflict was large. The purpose of the present study was to determine whether large interference effects are the necessary condition for observing an interacting pattern between conflicts.

In laboratory paradigms, a conflict is induced in tasks in which an incongruent trial is presented (i.e., a trial in which the stimulus is associated with two different response alternatives). Typically, the conflict tasks are the Stroop task (e.g., MacLeod, 1991), the Simon task (e.g., Simon & Small, 1969), or the flanker task (e.g., Eriksen & Eriksen, 1974). These tasks are described in Table 1. In all three tasks, participants are slower and more prone to errors on incongruent trials than on congruent trials (i.e., trials in which information converge

✉ Alodie Rey-Mermet  
alodie.rey-mermet@ku.de

<sup>1</sup> Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany

**Table 1** Description of the Stroop, flanker, and Simon tasks

Task	Task requirements	Incongruent trial	Congruent trial
Stroop	Participants are asked to indicate the color in which the color word is displayed while ignoring the meaning of the word.	The print color does not correspond to the meaning of the word (e.g., the word “green” printed in red).	The print color corresponds to the meaning of the word (e.g., the word “red” printed in red).
Flanker	Participants are asked to classify the central stimulus (e.g., a color decision) while ignoring the flanking stimuli.	The central stimulus requires a different response than the flanking stimuli (e.g., a red <i>X</i> flanked by two green <i>X</i> s).	The central stimulus requires the same response as the flanking stimuli (e.g., a row of five red <i>X</i> s).
Simon	Participants are asked to perform a decision (e.g., a color decision) on a stimulus presented on the left and right side of the screen by pressing left and right keys.	The stimulus location is different from the response location (e.g., a stimulus presented on the right side requiring a left key press).	The stimulus location is the same as the response location (e.g., a stimulus presented on the left side requiring a left key press).

*Note.* Incongruent trials are trials in which the stimulus is associated with two different response alternatives. Congruent trials are trials in which information converge to one response

to one response (Eriksen & Eriksen, 1974; MacLeod, 1991; Simon & Small, 1969; Stroop, 1935). This *congruency effect* occurs because in incongruent trials, the relevant feature has to be activated (i.e., the print color for the Stroop task, the central color for the flanker task, and the location of the response key for the Simon task), whereas the irrelevant feature has to be ignored (i.e., the word meaning, the color of the flanking letters, and the location on the screen, respectively).

In a few studies, these conflict tasks have been combined pairwise within the same trial (see Table 2). For example, in the Stroop–Simon combination, participants were asked to respond to the color of the color word while ignoring the position of the word on the screen (e.g., Hommel, 1997, Experiment 1). In the flanker–Simon combination, participants were asked to classify the central target while ignoring the flanking characters as well as the position of the complete stimulus on the screen (e.g., Hommel, 1997, Experiment 2). In the flanker–Stroop combination, participants were asked to respond the color of the central letter while ignoring the flanking letters (e.g., Rey-Mermet & Gade, 2016, Experiment 1). Combining such conflict tasks resulted in mixed findings. Some studies showed no interaction between the congruency conditions of both tasks (Akçay & Hazeltine, 2011; Egner et al., 2007; Hommel, 1997, Experiment 1; Kornblum, 1994; Simon & Berbaum, 1990; Stoffels & van der Molen, 1988, Experiment 1; Weissman, 2019, Experiment 2; Wendt et al., 2006, Experiment 2b). That is, the congruency effect of one task was not affected by whether the trial was incongruent or congruent for the other task. In contrast, other studies showed an underadditive interaction (Frühholz, Godde, Finke, & Herrmann, 2011; Hommel, 1997, Experiment 2; Rey-Mermet & Gade, 2016; Rey-Mermet, Gade, & Steinhauser, 2019; Stoffels & van der Molen, 1988, Experiment 2; Weissman, 2019, Experiment 1; Wendt et al., 2006, Experiment 2a). That is, the congruency effect of one task was smaller when the trials were incongruent for the other task compared with when they were congruent. For example,

when the Stroop and flanker tasks were combined, the Stroop congruency effect was reduced when the trials were flanker incongruent compared with when they were flanker congruent.

The interaction between the congruency conditions of both tasks could be explained within the conflict monitoring framework (see, e.g., Botvinick et al., 2001) by assuming that the detection and/or resolution of one conflict affects the resolution of the other conflict. That is, the conflict monitoring system estimates the current levels of conflict for the first task to be processed. For example, in the flanker–Stroop combination, if the first task is the flanker task and this is incongruent (i.e., associated with a high degree of conflict), this leads to a shift of control signal. This shift triggers an enhancement of all top-down biasing processes. This means that not only the flanking letters (i.e., the irrelevant features for the flanker task) but also the meaning of the word (i.e., the irrelevant feature for the Stroop task) would be inhibited. If the Stroop task is also incongruent, this reduces the impact of the word meaning, and thus results in a smaller Stroop congruency effect (see also Egner, 2008). According to this account, control processes generalize across both conflicts, explaining the underadditive interaction (e.g., Boy et al., 2010; Rey-Mermet & Gade, 2016). This account has been challenged by a recent study in which event-related potentials (ERPs) were measured in addition to RTs (Rey-Mermet et al., 2019). In that study, the results showed a sequential resolution of the flanker and Stroop conflicts. That is, an early ERP component (i.e., P2) was associated with the resolution of the flanker conflict, while a later ERP component (i.e., N450) was associated with the resolution of the Stroop conflict. However, although the interaction was still observed in the reaction times (RTs), there was no interaction in the ERP component associated to the Stroop congruency effect. Thus, processing flanker incongruent or congruent trials did not affect the resolution of the Stroop conflict, questioning the conflict-monitoring explanation. To explain the interaction still observed in the RTs, we

**Table 2** Studies combining Stroop, Simon, and flanker tasks

Combination of conflicts/Study	Underadditive Interaction?	Stimulus features	Congruency effect		
			Stroop	Simon	Flanker
<b>Stroop and Simon</b>					
Egner et al. (2007)	No	2 positions, 2 color words, 2 target colors	29	15	n.a.
Hommel (1997)—Exp. 1	No	2 positions, 2 color words, 2 target colors	20	27	n.a.
Kornblum (1994)	No	2 positions, 2 color words, 2 target colors	28	26	n.a.
Rey-Mermet & Gade (2016)—Exp. 2	Yes	4 positions, 4 color words, 4 target colors	209	46	n.a.
Simon & Berbaum (1990)	No	2 positions, 2 color words, 2 target colors	23	12	n.a.
Weissman (2019)—Exp. 1	Yes	4 positions, 4 color words, 4 target words	52	37	n.a.
Weissman (2019)—Exp. 2	No	2 positions, 2 color words, 2 target words	30	27	n.a.
Wendt et al. (2006)—Exp. 2b	No	2 positions, 2 color words, 2 target colors	6	27	n.a.
<b>Flanker and Simon</b>					
Akçay & Hazeltine (2011)—Exp. 1	No	4 positions, 4 flanker colors, 4 target colors	n.a.	79	43
Akçay & Hazeltine (2011)—Exp. 2	No	5 positions, 5 flanker colors, 5 target colors	n.a.	63	33
Frühholz et al. (2011)—EEG	Yes	2 positions, 2 flanker colors, 2 target colors	n.a.	-5	54
Frühholz et al. (2011)—fMRI	Yes	2 positions, 2 flanker colors, 2 target colors	n.a.	3	58
Hommel (1997)—Exp. 2	Yes	2 positions, 2 flanker colors, 2 target letters	n.a.	8	69
Rey-Mermet & Gade (2016)—Exp. 3	Yes	4 positions, 4 flanker colors, 4 target colors	n.a.	18	74
Stoffels & van der Molen (1988)—Exp. 1	No	3 positions, 3 flanker letters, 2 target letters <sup>a</sup>	n.a.	12	50
Stoffels & van der Molen (1988)—Exp. 2	Yes	3 positions, 3 flanker arrows, 2 target arrows <sup>a</sup>	n.a.	-4	70
Wendt et al. (2006)—Exp. 2a	Yes	2 positions, 2 flanker letters, 2 target letters	n.a.	2	74
<b>Stroop and flanker</b>					
Rey-Mermet & Gade (2016)—Exp. 1a	Yes	4 color words, 4 flanker colors, 4 target colors	199	n.a.	15
Rey-Mermet & Gade (2016)—Exp. 1b	Yes	4 color words, 4 flanker colors, 4 target colors	86	n.a.	46
Rey-Mermet & Gade (2016)—Exp. 1c	Yes	4 color words and 2 symbols, 6 flanker colors, 4 target colors	201	n.a.	-8
Rey-Mermet et al. (2019)—EEG	Yes	4 color words, 4 flanker colors, 4 target colors	171	n.a.	37

*Note.* Congruency effects are given in milliseconds. n.a. = not available. In studies in which the congruency effect was only presented graphically (i.e., Akçay & Hazeltine, 2011; Kornblum, 1994; Stoffels & van der Molen, 1988), WebPlotDigitizer (Version 4.1; Rohatgi, 2018) was used to estimate the data

<sup>a</sup> Simon and flanker tasks were so combined that the stimulus for the flanker task (either congruent, incongruent, or neutral) was accompanied by auditory noise presented to the ear ipsilateral to the correct response (Simon congruent), to the ear contralateral to the correct response (Simon incongruent), or to both ears (neutral)

assumed that resolving the flanker conflict and then the Stroop conflict preactivated the motor responses via direct associations. Thus, by response selection or execution, the correct response would be activated at least twice. This results in faster responding when both Stroop and flanker tasks are incongruent compared with when only the Stroop task is incongruent, explaining the reduction of the Stroop congruency effect in flanker incongruent trials.

So far, the reasons why some studies showed an interaction while others did not are unclear. For example, the discrepancy of findings did not result from combining different conflict tasks pairwise (Stroop with Simon, Simon with flanker, or Stroop with flanker), using different types of stimuli (e.g., arrows or letters) or using different stimulus set size (e.g., two vs. four). Irrespective of these manipulations, the interaction between the congruency conditions of both tasks was sometimes observed and sometimes not (see Table 2). There is, however, one key finding: The interaction was mainly observed when the congruency effect of one task was large (see Table 2, right part). For example, the interaction was found between the Stroop and Simon congruency when the Stroop congruency effect was the largest. Similarly, the interaction was observed between the flanker and Simon congruency when the flanker congruency effect was the largest. This raises the possibility that an interaction of that sort is only observed when the congruency effect is large. The reason is that a large congruency effect can be more probably reduced in its magnitude than a smaller congruency effect (see Hommel, 1997). Thus, in studies reporting no interaction (see Table 2), the small congruency effects may have worked against the possibility of finding an interaction so that it is not possible to determine whether or not the cognitive processes resulting in the interaction were required.

The purpose of the present study was to determine whether a large congruency effect is a necessary condition for observing an interaction between the congruency conditions of two conflict tasks combined within the same trial. This is important because if finding an interaction in the RTs simply requires a substantial congruency effect, the theoretical explanations used to explain the interaction in RTs can no longer be validated or invalidated by the simple presence or absence of an interaction. In the present study, three experiments were conducted in which the Stroop task was combined with the flanker task. This task combination was selected because the congruency effect—in particular, the Stroop congruency effect—was the largest (see Table 2), and thus its magnitude should be best modulated. The size of the Stroop congruency effect was increased by three experimental manipulations: (a) by testing older adults in addition to young adults (Experiment 1); (b) by manipulating the proportion of catch trials in which participants were asked to respond to the meaning of the word (instead to respond the color of the color word; Experiment 2); and (c) by asking participants to respond vocally (Experiment 3). Thus, if a large congruency effect is a necessary condition for observing an interaction between Stroop and flanker congruency (Hommel, 1997), the interaction should be expected

to be observed in older adults, in the conditions including catch trials, and with vocal responses.

In all three experiments, the data were analyzed using a null-hypothesis significance testing (NHST) approach with analyses of variance (ANOVAs) and Bayesian hypothesis testing approach with Bayesian ANOVA. Thus, the strength of evidence was assessed not only for the alternative hypothesis (e.g., the presence of the interaction), but also for the null hypothesis (e.g., the absence of the interaction). In addition, a state-trace analysis was performed (see Prince, Brown, & Heathcote, 2012, for an overview on the state-trace analysis). The goal of this analysis was to determine whether a difference due to the experimental manipulations results from additional processes or simply from large congruency effects.

## Experiment 1

In Experiment 1, young and older adults were tested in the same design as the one used in previous studies (Rey-Mermet & Gade, 2016; Rey-Mermet et al., 2019). That is, the Stroop task was so integrated with the flanker task that the central letter of the color word was congruent or incongruent to the color of the flanking letters and congruent and incongruent to the meaning of the word (see Fig. 1). Participants were asked to respond to the color of the central letter while ignoring the meaning of the word and the color of the flanking letters. Here, the Stroop conflict results from the competition between the color of the central letter and the meaning of the word. As this single-letter Stroop task was found to eliminate the Stroop congruency effect (Besner, Stolz, & Boutilier, 1997), the presence of this congruency effect was ensured by presenting catch trials. In these trials, participants were asked to respond to the meaning of word while ignoring the color of the letters.

Testing older adults has the advantage that older participants typically show larger Stroop congruency effects than young adults (Bugg, DeLosh, Davalos, & Davis, 2007; Spieler, Balota, & Faust, 1996; but see Rey-Mermet & Gade, 2018; Verhaeghen, 2014). Therefore, if a large Stroop congruency effect is the necessary condition for observing the interaction between the Stroop and flanker congruency conditions (Hommel, 1997), an interaction would be expected in both age groups. Moreover, if the size of the interaction depends from the magnitude of the Stroop congruency effect, the reduction of the Stroop congruency effect should be larger for older adults than for young adults. In contrast, if the interaction between the conflicts results from a generalization of control processes, the reduction of the Stroop congruency effect should be smaller for older adults than for young adults. The reason is that according to the account assuming age-related deficits in attentional-control processes (e.g., Hasher & Zacks, 1988; Hasher, Zacks, & May, 1999),

older adults would be less able to recruit attentional control than young adults, and thus they would have a smaller shift of control signal when the flanker task is incongruent. This should result in a smaller adjustment of attentional control and thus in a smaller reduction of the Stroop congruency effect for older adults.

Comparing the RTs of young and older adults might be problematic because older adults show slower RTs than do young adults by a constant proportion across many speeded tasks (e.g., Cerella & Hale, 1994; Myerson, Hale, Wagstaff, Poon, & Smith, 1990). To take this into account, the analyses were performed not only on raw RTs but also on RTs after a natural logarithm (log) transformation was applied.

## Method

### Participants

Sixty-eight young adults (ages between 18 and 35 years) and 59 older adults (ages between 66 and 76 years) took part in the experiment.<sup>1</sup> All older adults and 24 young adults participated in a previous study (Rey-Mermet et al., 2018). For that study, participants performed the Mini-Mental State (MMS; Folstein, Folstein, & McHugh, 1975) to screen for dementia, the Consortium to Establish a Registry for Alzheimer's Disease-Plus (CERAD-Plus) test battery (Satzger et al., 2001) to assess their overall cognitive status, the SF-36 Questionnaire (Bullinger, 1998) to assess their health status, and the Beck Depression Inventory-II (BDI-II; Hautzinger, Keller, & Kühner, 2006) to assess depression. In addition to the BDI-II, older adults performed the Geriatric Depression Scale (GDS; Yesavage et al., 1983). Demographic information as well as the results on these background measures are presented in Appendix A.

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

### Material

Similar to Rey-Mermet and Gade (2016, Experiment 1a), the experiment consisted of a stimulus–response mapping block,

flanker and Stroop pure blocks, as well as mixed blocks. The stimulus–response mapping block served as practice to learn the four-choice stimulus–response mapping. The flanker and Stroop pure blocks served to familiarize participants with each task separately and to control for the presence of congruency effects within each task. Mixed blocks were the critical blocks in which both tasks were combined. The stimuli for each block type are described separately.

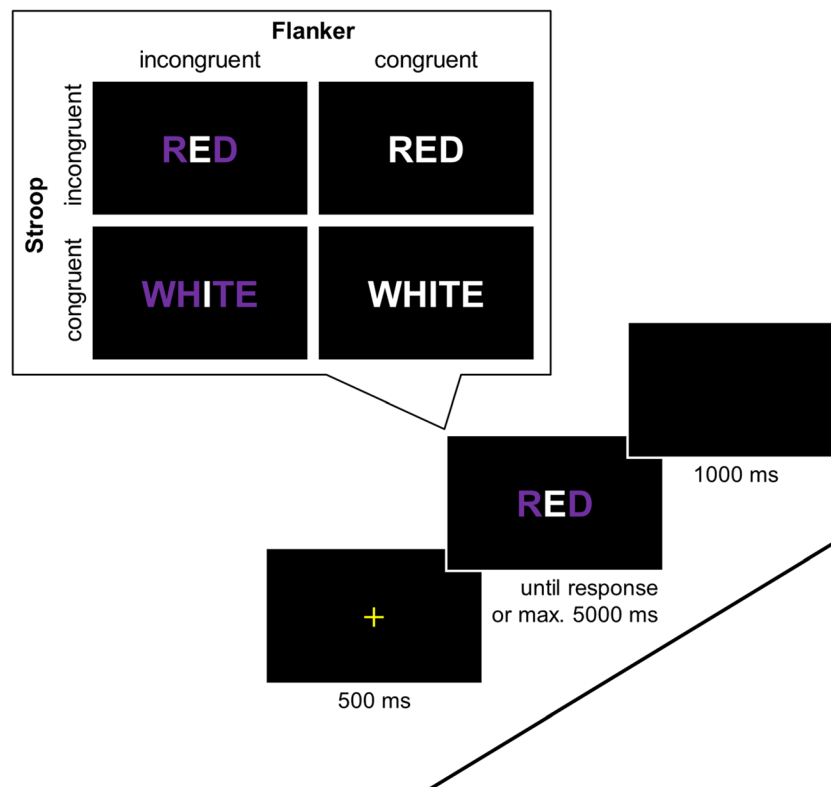
**Stimulus–response mapping block** The stimuli consisted of a row of four “\*” colored in either red, brown, violet, or white. These colors were used in order to have German color words (“rot”, “braun”, “violett”, and “weiss”, respectively) with an odd number of letters and thereby appropriate to create a flanker conflict in the mixed blocks. The stimuli were presented on a black background at the center of the computer screen in 54-point Arial font. The stimulus for each trial was determined randomly.

**Pure flanker blocks** For the flanker task, the stimuli were the middle letter of the German words for *red*, *brown*, *violet*, and *white* (i.e., *o* for “rot”, *a* for “braun”, *l* for “violett”, and *i* for “weiss”). Each letter was presented five times in a row, displayed either in red, brown, violet, or white. All stimuli were presented at the center of the computer screen in 54-point Arial Bold font. In comparison to Rey-Mermet and Gade (2016), the size of the stimuli was increased so that all stimuli were easily perceived by older adults. The stimulus was determined randomly for each trial. A trial was congruent when the color of the central letter was the same as the color of the flanking letters. A trial was incongruent when the color of the central letter was different from the color of the flanking letters. Congruent and incongruent trials were presented equally often.

**Pure Stroop blocks** The stimuli were the four German words *red*, *brown*, *violet*, and *white* (i.e., “rot”, “braun”, “violett”, and “weiss”), displayed in either red, brown, violet, or white. The stimuli were presented on a black background at the center of the computer screen in 107-point Arial Bold font. The stimulus was determined randomly for each trial. A trial was congruent when the color and the meaning of the word corresponded (e.g., the color word “red” printed in red). A trial was incongruent when the word was printed in a noncorresponding color (e.g., the color word “red” printed in brown). Congruent and incongruent trials were presented equally often. For the pure Stroop blocks, 25% of the trials were catch trials in which participants were asked to indicate the meaning of the word while ignoring the color. In these trials, the words “Wort lesen” (i.e., “read the word” in English) appeared in blue 54-point Arial Bold font 150 pixels above the centrally presented stimulus. These catch trials were randomly interspersed in the block.

<sup>1</sup> The minimum sample size to observe an interaction between the congruency conditions of the Stroop and flanker tasks was 13. This estimate was computed with G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007; Mayr, Erdfelder, Buchner, & Faul, 2007) using a power of .99, a Type I error rate of 5%, and a partial eta-square of .31. The partial eta-square was calculated with the MOTE package (Buchanan, Scofield, & Valentine, 2019) using the values of the interaction between Stroop and flanker congruency conditions,  $F(1, 23) = 10.51$ ,  $p < .01$ , observed in Rey-Mermet and Gade (2016, Experiment 1a). Therefore, the sample size in each experiment should be sufficient to observe the interaction of interest in each experimental condition.





**Fig. 1** Example of one trial sequence. Participants were asked to indicate the color of the central letter while ignoring the meaning of the word and the flanking letters. The top panel shows example of stimuli for each trial

type (Stroop incongruent–flanker incongruent, Stroop incongruent–flanker congruent, Stroop congruent–flanker incongruent, and Stroop congruent–flanker congruent)

**Mixed blocks** The material for the mixed blocks was the same as for the pure Stroop blocks, except for the following modifications. First, the middle letter of the German words for *red*, *brown*, *violet*, or *white* (i.e., o for “rot”, a for “braun”, l for “violett”, and i for “weiss”) were displayed either in the same color as in the meaning of the word (Stroop congruent) or in a different color (Stroop incongruent). Moreover, the flanking letters of the color words (e.g., “r” and “t” for “rot”) were presented either in the same color as the color of the middle letter (flanker congruent) or in a different color (flanker incongruent). Thus, there were four trial types, occurring with an equal probability of 25%: Stroop incongruent and flanker incongruent, Stroop incongruent and flanker congruent, Stroop congruent and flanker incongruent, Stroop congruent and flanker congruent (see Fig. 1, top panel). Similar to the pure Stroop blocks, 25% of the trials in the mixed blocks were catch trials in which participants were asked to indicate the meaning of the word while ignoring the colors. These trials were randomly interspersed in the block.

## Procedure

A trial sequence is illustrated in Fig. 1 (bottom panel). A trial started with the presentation of a yellow fixation cross presented centrally for 500 ms. Then, the stimulus was presented until a response or 5,000 ms elapsed. In case of an error, the

word “Fehler” (i.e., “error” in English) was displayed centrally in yellow for 500 ms. Finally, an intertrial interval of 1,000 ms elapsed. To respond, participants used four response keys (y, v, m, -), which represent every third key on the bottom row of a QWERTZ keyboard and which were mapped to the colors red, brown, violet, and white, respectively.

All participants first performed the stimulus–response mapping block for practice. After this block, half of the participants started with a pure flanker block, followed by a pure Stroop block, whereas the other half started with a pure Stroop block followed by a pure flanker block. After the first two pure blocks, all participants performed six mixed blocks. After these critical blocks, participants performed a pure flanker block and a pure Stroop block in the same order as in the first two pure blocks (e.g., participants who started with a pure flanker block, followed by a pure Stroop block, had then a pure flanker block, followed by a pure Stroop block). The stimulus–response block included 96 trials. Each pure block included 16 practice trials and 96 experimental trials. Each mixed block included 16 practice trials and 72 experimental trials. In comparison to Rey-Mermet and Gade (2016), the number of trials per mixed block was reduced in the present experiment so that large fatigue effects should be avoided in older adults. The number of blocks was, however, increased so that the number of observations was the same in the present experiment as in the previous one. An overall

feedback, which consisted of mean RTs and error rate, was displayed at the end of each block. Participants could take brief rests after each block.

At the beginning of the experiment, participants were given short instructions about the task they had to carry out. Before each change of block type (e.g., from the pure flanker block to the mixed blocks), a short instruction for the now relevant task was also given. For the stimulus-response mapping block, participants were instructed to indicate the color of the stimulus. For the pure Stroop blocks, participants were instructed that for most trials they should indicate the color of the color word while ignoring the meaning of the word. However, when “read the word” appeared above the stimulus (catch trial), they should indicate the meaning of the word while ignoring the color of the color word. For the pure flanker blocks, participants were instructed to indicate the color of the central letter while ignoring the flanking letters. For the mixed blocks, participants were instructed that for most trials they should indicate the color of the central letter while ignoring the meaning of the word and the flanking letters. However, when “read the word” appeared above the stimulus (catch trial), they should indicate the meaning of the word while ignoring the colors.

Participants were tested in group up to five during two sessions of approximately 2 hours each (including breaks after every block and a longer break in the middle of each session). Both sessions were separated at least by 12 hours, and maximally by 1 week. At the beginning of the first session, participants signed the informed consent sheet, then both sessions were structured similarly. That is, participants first performed an unrelated working-memory task, and then were given the instructions for the stimulus-response mapping block. The experiment was programmed using Tscope5 (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006) and run on IBM compatible computer.

### Data preparation

The stimulus-response mapping block, which served as a practice block, and practice trials at the beginning of the pure and mixed blocks were not analyzed. The first trial of each block was considered as a warm-up trial and thus was excluded. The dependent variables were RTs and error rates. For both dependent measures, catch trials and the first trial following a catch trial were also removed in the mixed and pure Stroop blocks. This dismissed 44.41% of the trials for the mixed blocks, and 46.03% of the trials for the pure Stroop blocks. An arcsine square root transformation was applied to the error rates for statistical analysis. For RTs, errors and one trial following an error were removed from the raw data set. Additionally, outlier RTs were excluded by removing RTs faster than three standard deviations (SD) from the mean and slower than three standard deviations from the mean for each congruency condition, participant, and session, separately. This further dismissed 8.71% of the trials for the mixed

blocks, 8% of the trials for the pure Simon blocks, and 6.53% of the trials for the pure flanker blocks. Speed differences between young and older adults were controlled for (e.g., Myerson et al., 1990; Salthouse, 1996; Salthouse & Babcock, 1991) by applying a natural logarithm (log) transformation to all RTs. In that way, if RTs were increased in older age only due to general proportional slowing, then log-transformed RTs should not differ between age groups.

### Data analysis

For the statistical analyses, the focus was on standard (no-catch) trials from the mixed blocks. For the sake of completeness, performance on pure blocks is presented in Appendix B (see Tables 11 and 12 as well as Fig. 9), and performance on the catch trials of the mixed blocks is presented in Appendix C (see Table 16 and Fig. 12). In the mixed blocks, the goal was to investigate whether both age groups differed in the interaction between Stroop and flanker congruency conditions. To this end, in the NHST approach, a three-way ANOVA was carried out, with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-subject factors, and age group (young, older) as a between-subjects factor. This was implemented in R (R Core Team, 2019) with the afex package (Singmann, Bolker, & Westfall, 2018). An alpha level of 0.05 was used for all these tests. Effect sizes are expressed as  $\eta_g^2$  values (generalized eta square; cf. Olejnik & Algina, 2003). In case of a significant three-way interaction, follow-up two-way ANOVAs were performed for each age group separately. In case of significant two-way interactions, follow-up *t* tests were performed, and effect sizes are expressed with Cohen's *d*.

In addition to these ANOVAs, Bayesian ANOVAs were computed. These were implemented with default prior scales in R using the BayesFactor package (Morey & Rouder, 2015). The Bayes factor (BF) in favor of the alternative hypothesis (i.e.,  $BF_{10}$ ) for the main effect model including either the Stroop and flanker congruency was computed by comparing the main effect model against the null model. The  $BF_{10}$  for the two-way interaction models was estimated by comparing the interaction model (i.e., the model with both main effects and the interaction) against the main-effects model (i.e., the model with both main effects, but no interaction). That is, this BF was computed as the ratio  $BF_{\text{Interaction Model}} / BF_{\text{Main Effects Model}}$ . The  $BF_{10}$  for the three-way interaction model was estimated by comparing the three-way interaction model (i.e., the model with both main effects, all two-way interactions, and the three-way interaction) against the model including the main effects and the two-way interactions. That is, this BF was computed as the ratio  $BF_{\text{Three-way Interaction Model}} / BF_{\text{Main Effects and Two-way Interactions Model}}$  (Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2017). For each model comparison, the Bayes factor in favor of the null hypothesis ( $BF_{01}$ ) was

computed as  $1/BF_{10}$ . Raftery's (1995) classification scheme was used to interpret the BF. Thus, a BF between 1 and 3 was considered as weak evidence, a BF between 3 and 20 was considered as positive evidence, a BF between 20 and 150 was considered as strong evidence, and a BF larger than 150 was considered as very strong evidence.

In a further step, a state-trace analysis was conducted (see Prince et al., 2012). The goal of this analysis was to determine whether cognitive processes are added in the Stroop congruency effect within and across age groups in flanker incongruent trials compared with flanker congruent trials. To this end, a state-trace plot was displayed in which the Stroop congruency effect for flanker congruent trials is plotted on the  $x$ -axis and the Stroop congruency effect for flanker incongruent trials is plotted on the  $y$ -axis. In this scatterplot, each participant of each age group (young and older) yields a data point. The aim of this plot was to investigate whether the data displayed in the state-trace plot can be explained with one single line or two lines (one for young adults, one for older adults). If a single line fit the data, it would imply that there was no difference in the Stroop congruency effect between both age groups in the relation between flanker congruent and incongruent trials. Therefore, no process was added in the older age group and a difference between both age groups only occurred because the Stroop congruency effect was larger in the older age group than in the young age group. If two lines were necessary, the interpretation would depend on the pattern that emerged from the two lines. If both lines were parallel to the diagonal but differ in elevation, it would imply that the relation between the Stroop congruency effect in flanker incongruent and congruent trials is additive. In this case, the intercept would be significantly different from zero. In contrast, if both lines diverge with a slope larger than 1, the impact of flanker incongruent trials on the Stroop congruency effect is considered as *multiplicative*.

To determine statistically whether one or two lines were necessary, the following linear regression model was fitted:

$$RT_{\text{Stroop effect-flanker incongruent},i} = \beta_0 + \beta_1 \text{ age group} \\ + \beta_2 RT_{\text{Stroop effect-flanker congruent},i} \\ + \beta_3 \text{ age group} \times RT_{\text{Stroop effect-flanker congruent},i} + \varepsilon_i,$$

where  $RT_{\text{Stroop effect-flanker incongruent},i}$  is the Stroop congruency effect for flanker incongruent trials from participant  $i$ ,  $RT_{\text{Stroop effect-flanker congruent},i}$  is the Stroop congruency effect for flanker congruent trials from participant  $i$ , *age group* is a dummy variable in which young and older age groups were coded with 0 and 1, respectively,  $\beta_0$  is the intercept,  $\beta_1$  is the effect of the age group (young vs. older) on the intercept,  $\beta_2$  is the slope relating flanker congruent trials to flanker incongruent trials,  $\beta_3$  is the effect of the age group on the slope, and  $\varepsilon_i$  is the residual for participant  $i$ .

As the goal was to determine the impact of age group on the Stroop congruency effect in flanker incongruent trials compared with flanker congruent trials, the primary focus was on the interaction term. Therefore, a full regression model was compared against a restricted model in which the interaction term was removed. Model selection was evaluated via multiple fit indices:  $R^2$ , the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the deviance ( $= -2 \times \text{logarithmized likelihood}$ ). Except for  $R^2$ , smaller indices indicate better fit. To examine if one model (restricted vs. full) fit the data reliably better than another, two analyses were performed. First,  $\chi^2$  difference ( $\Delta\chi^2$ ) tests were conducted on nested models. If the more complex model (i.e., the model with more free parameters) yields a reduction in  $\chi^2$  that is significant given the loss of degrees of freedom, it is accepted as having better fit. Second, a Bayesian hypothesis test using the BIC approximation was performed (Wagenmakers, 2007). That is, the difference between the BIC for the null hypothesis (i.e., the restricted model) and the BIC for the alternative hypothesis (i.e., the full model) were used to compute a Bayes factor in favor of the null hypothesis ( $BF_{01}$ ). The Bayes factor in favor of the alternative hypothesis ( $BF_{10}$ ) was computed as  $1/BF_{01}$ . Thus, if  $\Delta\chi^2$  test was not significant and the  $BF_{01}$  constituted positive to very strong evidence, it would imply that the restricted model had a better fit than the full model. Thus, only one line would be sufficient to account for the data in the state-trace plot. In contrast, if  $\Delta\chi^2$  test was significant and the  $BF_{10}$  constituted positive to very strong evidence, it would imply that the full model had a better fit. In this case, the estimates of the fixed parameters were examined. Only if the estimate of the interaction term was significant, two lines—one for the young age group and one for the older age group—were considered as necessary to account for the data in the state-trace plot.

## Results

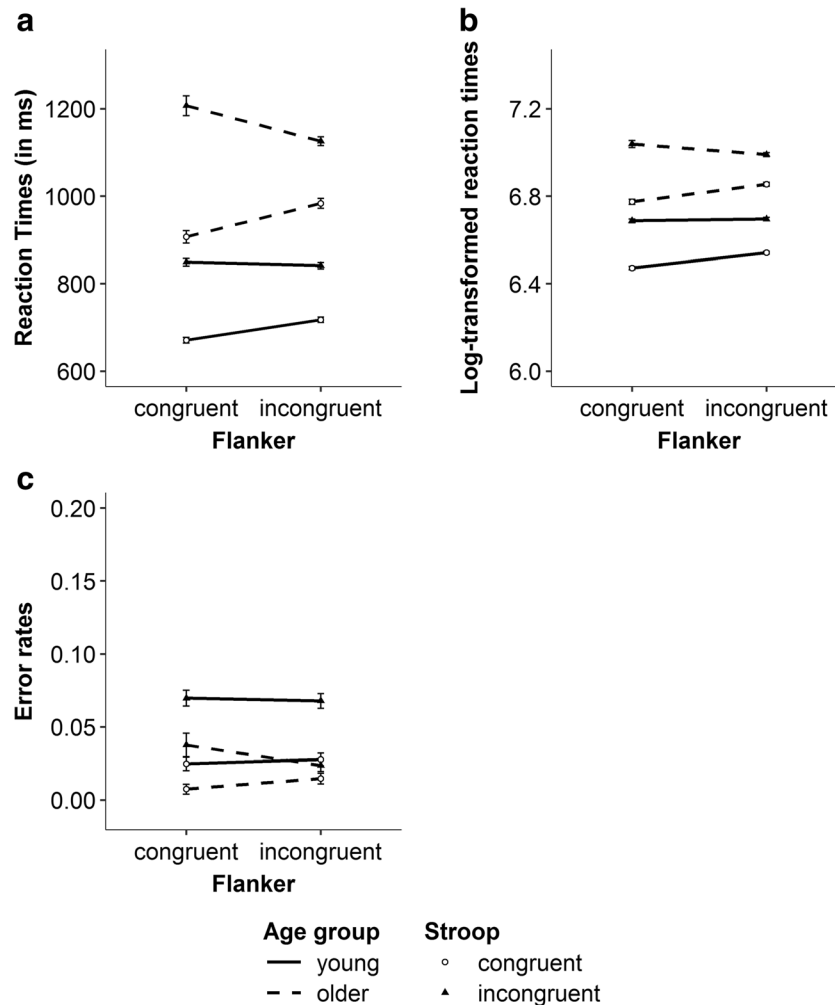
The most relevant results are whether participants in both age groups (i.e., young and older adults) differ in the modulation of the Stroop congruency effect depending on whether the trials are flanker congruent or incongruent (i.e., the interaction between both congruency variables). First, the results of the ANOVAs from the NHST and Bayesian hypothesis testing approach are reported. Then, the results for the state-trace analysis are presented.

**Analysis of variance** The descriptive results for raw and log-transformed RTs as well as for error rates are presented in Fig. 2. The results from the NHST and Bayesian analyses are presented in Table 3. For RTs and log-transformed RTs, the ANOVA showed a significant three-way interaction between Stroop congruency, flanker congruency, and age group. The Bayesian analysis suggested strong to very strong evidence in favor of the three-way interaction model. This interaction was then



disentangled by conducting follow-up two-way repeated-measures ANOVA, with the factors Stroop congruency and flanker congruency for each age group separately. The results of these follow-up ANOVAs are presented in Table 4. For the young adults, the ANOVA showed a significant two-way interaction for RTs and log-transformed RTs. The Bayesian analysis suggested very strong evidence in favor of the interaction model for both dependent measures. For RTs, the Stroop congruency effect was smaller, but still significant, when the trials were flanker incongruent (124 ms),  $t(67) = 18.25, p < .001, d = 2.21, BF_{10} = 3.12 \times 10^{24}, BF_{01} = 3.21 \times 10^{-25}$ , compared with when they were flanker congruent (178 ms),  $t(67) = 21.74, p < .001, d = 2.64, BF_{10} = 5.92 \times 10^{28}, BF_{01} = 1.69 \times 10^{-29}$  (see Fig. 2a). A similar pattern was observed for log-transformed RTs (see Fig. 2b; Stroop congruency effect for flanker incongruent trials:  $M = 0.15, t(67) = 21.57, p < .001, d = 2.62, BF_{10} = 3.7 \times 10^{28}, BF_{01} = 2.66 \times 10^{-29}$ ; Stroop congruency effect for flanker congruent trials:  $M = 0.22, t(67) = 25.90, p < .001, d = 3.14, BF_{10} = 1.78 \times$

$10^{33}, BF_{01} = 5.60 \times 10^{-34}$ ). For the older adults, the ANOVA also showed a significant two-way interaction for RTs and log-transformed RTs (see Table 4). The Bayesian analysis suggested very strong evidence in favor of the interaction model for both dependent measures (see Table 4). For RTs, the Stroop congruency effect was smaller, but still significant, when the trials were flanker incongruent (142 ms),  $t(58) = 14.85, p < .001, d = 1.93, BF_{10} = 2.27 \times 10^{18}, BF_{01} = 4.41 \times 10^{-19}$ , compared with when they were flanker congruent (300 ms),  $t(58) = 14.06, p < .001, d = 1.83, BF_{10} = 2.00 \times 10^{17}, BF_{01} = 5.01 \times 10^{-18}$  (see Fig. 2a). A similar pattern was observed for log-transformed RTs (see Fig. 2b; Stroop congruency effect for flanker incongruent trials:  $M = 0.14, t(58) = 16.54, p < .001, d = 2.15, BF_{10} = 3.19 \times 10^{20}, BF_{01} = 3.14 \times 10^{-21}$ ; Stroop congruency effect for flanker congruent trials:  $M = 0.26, t(58) = 17.40, p < .001, d = 2.27, BF_{10} = 3.52 \times 10^{21}, BF_{01} = 2.84 \times 10^{-22}$ ). Thus, the reduction of the Stroop congruency effect in flanker incongruent trials was larger for older adults (158 ms



**Fig. 2** Experiment 1. Modulation of the Stroop congruency effect by flanker congruency within a trial for both age groups (young and older). Mean reaction times (RTs) with raw RTs (a) and log-transformed RTs (b)

as well as mean raw error rates (c) for both Stroop and flanker congruency. Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008)

**Table 3** Experiment 1: Inferential statistical values for the three-way analysis of variance (ANOVA), with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-

subject factors, and age group (young, older) as a between-subjects factor, and Bayes factors (BF) from the model comparisons

Dependent measure	Effect–Model	ANOVA			Bayesian analysis	
		<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Reaction times	Stroop congruency	655.50	<.001	.28	$8.94 \times 10^{87}$	$1.12 \times 10^{-88}$
	Flanker congruency	3.16	.078	<.001	0.14	7.31
	Age group	132.71	<.001	.48	$1.63 \times 10^{18}$	$6.14 \times 10^{-19}$
	Stroop congruency $\times$ Flanker congruency	123.97	<.001	.03	$3.24 \times 10^{12}$	$3.09 \times 10^{-13}$
	Stroop congruency $\times$ Age group	23.36	<.001	.01	$9.86 \times 10^4$	$1.01 \times 10^{-5}$
	Flanker congruency $\times$ Age group	5.01	.027	.001	0.20	4.93
	3-way interaction	29.32	<.001	.007	$2.28 \times 10^3$	$4.39 \times 10^{-4}$
Log-transformed reaction times	Stroop congruency	1,033.45	<.001	.31	$3.62 \times 10^{113}$	$2.77 \times 10^{-114}$
	Flanker congruency	39.36	<.001	.009	2.09	0.48
	Age group	169.13	<.001	.55	$5.59 \times 10^{21}$	$1.79 \times 10^{-22}$
	Stroop congruency $\times$ Flanker congruency	165.38	<.001	.03	$1.88 \times 10^{16}$	$5.33 \times 10^{-17}$
	Stroop congruency $\times$ Age group	1.69	.196	<.001	0.34	2.97
	Flanker congruency $\times$ Age group	7.26	.008	.002	0.23	4.33
	Three-way interaction	19.44	<.001	.003	37.18	0.03
Arcsine transformed error rates	Stroop congruency	142.27	<.001	.20	$5.06 \times 10^{33}$	$1.98 \times 10^{-34}$
	Flanker congruency	0.08	.784	<.001	0.10	9.89
	Age group	51.06	<.001	.18	$1.01 \times 10^8$	$9.92 \times 10^{-9}$
	Stroop congruency $\times$ Flanker congruency	14.34	<.001	.01	9.72	0.10
	Stroop congruency $\times$ Age group	5.75	.018	.01	6.11	0.16
	Flanker congruency $\times$ Age group	0.00	.951	<.001	0.14	7.33
	Three-way interaction	4.97	.028	.004	0.92	1.09

Note. *dfs* = (1, 125) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

and 0.13) than for young adults (55 ms and 0.06 for RTs and for log-transformed RTs, respectively).

For the error rates, the three-way interaction between Stroop congruency, flanker congruency, and age group was significant, but the Bayesian analysis suggested inconclusive evidence regarding the interaction model including the three-way interaction. Nevertheless, the three-way interaction was disentangled by computing two-way repeated-measures ANOVA with the factors Stroop congruency and flanker congruency for each age group separately. For young adults, the two-way ANOVA showed no significant interaction, and the Bayesian analysis suggested positive evidence against the interaction model. Thus, the Stroop congruency effect was similar for flanker incongruent and congruent trials (.04 and .05, respectively). For the older adults, the two-way ANOVA showed a significant two-way interaction, and the Bayesian analysis suggested strong evidence in favor of the interaction model. In line with the RT results, the Stroop congruency effect in error rates was smaller, but still significant, when the trials were flanker incongruent (.01),  $t(58) = 3.56$ ,  $p = .001$ ,  $d = 0.46$ , BF<sub>10</sub> = 35.10, BF<sub>01</sub> = 0.03, compared with when they were flanker congruent (.03),  $t(58) = 6.47$ ,  $p < .001$ ,  $d =$

0.84, BF<sub>10</sub> =  $5.63 \times 10^5$ , BF<sub>01</sub> =  $1.78 \times 10^{-6}$  (see Fig. 2c). Again, the reduction of the Stroop congruency effect in flanker incongruent trials was larger for older adults (.02) than for young adults (.005).

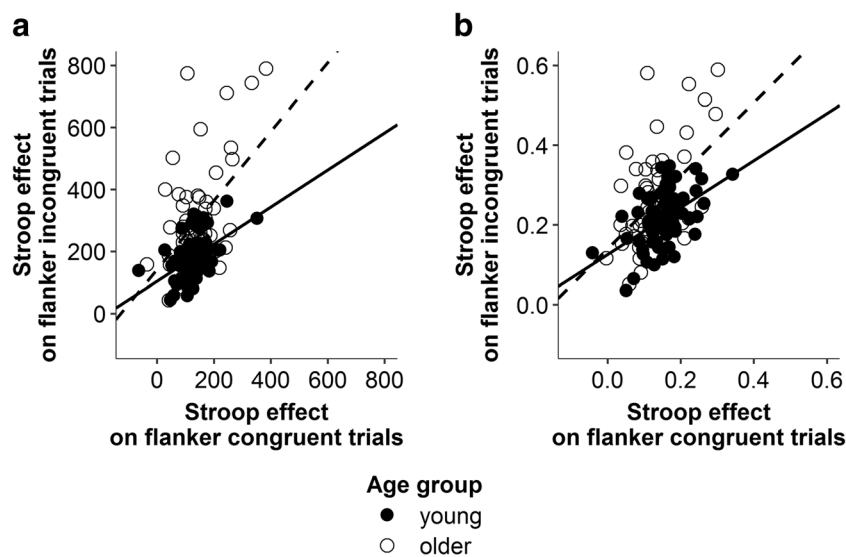
**State-trace analysis** The state-trace plot is depicted in Fig. 3. Goodness-of-fit statistics and model comparison results are presented in Table 5 (upper part). Estimates of the fixed parameters are summarized in Table 6 (upper part). As shown in Table 5, model fit was relatively similar for both models (i.e., with and without the interaction term). For both dependent measures, the  $\Delta\chi^2$  was not significant, and the Bayesian analysis suggested small to positive evidence for the model without the interaction term. Together, this indicates that the data displayed in the state-trace plot (see Fig. 3) were better explained with one single line.

In sum, the state-trace results speak against a difference in the Stroop congruency effect between both age groups in the relation between flanker congruent and incongruent trials. Therefore, no process was added in the older age group compared with the young age group.

**Table 4** Experiment 1: Inferential statistical values for the follow-up two-way repeated-measures analysis of variance (ANOVA), with the factors Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent), and Bayes factors (BF) from the model comparisons for each age group (young and older) separately

Age group/Dependent measure	Effect–Model	ANOVA				Bayesian analysis	
		<i>df</i>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
<b>Young adults</b>							
Reaction times	Stroop congruency	1, 67	542.22	<.001	.29	$9.17 \times 10^{64}$	$1.09 \times 10^{-65}$
	Flanker congruency	1, 67	17.47	<.001	.007	0.46	2.15
	Two-way interaction	1, 67	51.21	<.001	.01	$4.78 \times 10^4$	$2.09 \times 10^{-5}$
Log-transformed reaction times	Stroop congruency	1, 67	759.53	<.001	.32	$2.01 \times 10^{69}$	$4.98 \times 10^{-70}$
	Flanker congruency	1, 67	55.81	<.001	.02	4.71	0.21
	Two-way interaction	1, 67	65.13	<.001	.01	$3.71 \times 10^5$	$2.69 \times 10^{-6}$
Arcsine transformed error rates	Stroop congruency	1, 67	104.27	<.001	.26	$4.37 \times 10^{26}$	$2.29 \times 10^{-27}$
	Flanker congruency	1, 67	0.09	.769	<.001	0.13	7.50
	Two-way interaction	1, 67	1.79	.185	.002	0.29	3.46
<b>Older adults</b>							
Reaction times	Stroop congruency	1, 58	260.46	<.001	.28	$9.14 \times 10^{37}$	$1.09 \times 10^{-38}$
	Flanker congruency	1, 58	0.06	.804	<.001	0.14	7.13
	Two-way interaction	1, 58	72.73	<.001	.05	$5.33 \times 10^8$	$1.88 \times 10^{-9}$
Log-transformed reaction times	Stroop congruency	1, 58	379.52	<.001	.30	$9.49 \times 10^{45}$	$1.05 \times 10^{-46}$
	Flanker congruency	1, 58	4.75	.033	.003	0.20	4.92
	Two-way interaction	1, 58	94.50	<.001	.04	$1.13 \times 10^{10}$	$8.88 \times 10^{-11}$
Arcsine transformed error rates	Stroop congruency	1, 58	45.52	<.001	.14	$4.71 \times 10^8$	$2.12 \times 10^{-9}$
	Flanker congruency	1, 58	0.02	.900	<.001	0.14	7.05
	Two-way interaction	1, 58	12.85	<.001	.03	20.60	0.05

Note. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)



**Fig. 3** Experiment 1: State-trace plot. Solid line results from the regression analysis using the young adults only. Dashed line results from the regression analysis using the older adults only. **a** Reaction times. **b** Log-transformed reaction times

**Table 5** Experiments 1 and 2: State-trace analysis. Goodness-of-fit statistics and model comparison results between the full regression model and the restricted regression model (i.e., without the interaction term)

Experiment	Dependent measure	Model	df	R <sup>2</sup>	AIC	BIC	Deviance	$\Delta\chi^2$ , df	BF <sub>10</sub>	BF <sub>01</sub>
1	Reaction times	Full	123	0.40	1552.71	1,566.93	1,402,270.92	2.92, 1 ( <i>p</i> = .090)	0.39	2.54
		Restricted	124	0.39	1553.69	1,565.07	1,435,600.51	–	–	–
	Log-transformed reaction times	Full	123	0.30	–268.56	–254.33	0.83	1.84, 1 ( <i>p</i> = .178)	0.23	4.40
		Restricted	124	0.28	–268.67	–257.30	0.84	–	–	–
2	Reaction times	Full	91	0.53	1,079.60	1092.37	431,482.66	0.65, 1 ( <i>p</i> = .422)	0.14	6.95
		Restricted	92	0.53	1,078.28	1,088.49	434,561.06	–	–	–

Note. AIC = Akaike information criterion. BIC = Bayesian information criterion. Change in  $\chi^2$  ( $\Delta\chi^2$ ) was calculated relative to the full model. BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., the restricted model). BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., the full model); this was computed as 1/BF<sub>01</sub>

## Discussion

The results showed an underadditive interaction between Stroop and flanker congruency in both age groups for raw and log-transformed RTs. Critically, the reduction of the Stroop congruency effect in flanker incongruent trials was larger for older adults than for young adults. Thus, even after age-related differences in processing speed were removed, the size of the interaction still increased when the magnitude of the Stroop congruency effect increased. These findings cannot be explained by an account assuming age-related deficits in attentional-control processes (e.g., Hasher & Zacks, 1988; Hasher et al., 1999). According to such an account, in comparison to young adults, older adults should show a smaller—not a larger—reduction of the Stroop congruency effect. In contrast, the results of Experiment 1 are in line with the view that a large Stroop congruency effect is more reduced and thus creates an optimal condition for finding an interaction between Stroop and flanker congruency (Hommel, 1997).

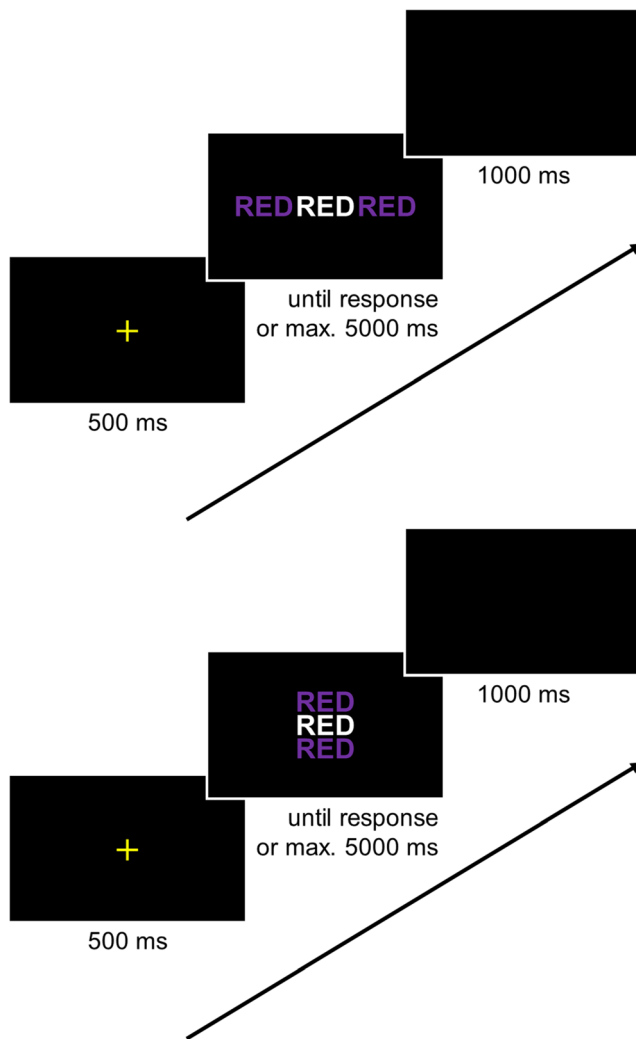
## Experiment 2

As the single-letter Stroop task was found to eliminate the congruency effect (Besner et al., 1997), the size of the Stroop congruency effect was increased in Experiment 1 as well as in the previous experiments (Rey-Mermet & Gade, 2016; Rey-Mermet et al., 2019) by including catch trials. The same logic was used in Experiment 2, so that this experiment included two conditions: a 25%-catch condition, in which 25% of the trials were catch trials, and a 0%-catch condition, in which no catch trials were presented. The presence of a substantial Stroop congruency effect was, however, ensured in both conditions by using the standard Stroop task instead of the single-letter Stroop task. That is, in Experiment 2, participants saw three colored color words, and they were asked to respond to the color of the central word while ignoring the meaning of the words and the flanking words. Because the distance between the central word and the flanker words is critical for the size of the flanker congruency effect (e.g., Eriksen & Eriksen, 1974), the position of the flanker was also manipulated. For half of participants, flanker words were presented to the left or right of the central word (see Fig. 4, top

**Table 6** Experiments 1 and 2: State-trace analysis. Estimates of the fixed parameters for the full regression model and for the restricted regression model (i.e., without the interaction term). Standard errors are presented in parentheses

Experiment	Dependent measure	Model	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$
1	Reaction times	Full	<b>104.30</b> (31.64)	37.58 (43.92)	<b>0.60</b> (0.23)	0.52 (0.30)
		Restricted	<b>66.07</b> (22.56)	<b>105.12</b> (19.34)	<b>0.91</b> (0.15)	–
	Log-transformed reaction times	Full	<b>0.13</b> (0.03)	0.01 (0.04)	<b>0.59</b> (0.17)	0.33 (0.24)
		Restricted	<b>0.10</b> (0.02)	<b>0.06</b> (0.01)	<b>0.75</b> (0.12)	–
2	Reaction times	Full	<b>44.42</b> (19.99)	37.11 (30.80)	<b>0.61</b> (0.19)	0.19 (0.24)
		Restricted	<b>32.96</b> (14.01)	<b>58.33</b> (15.94)	<b>0.74</b> (0.11)	–

Note.  $\beta_0$  is the intercept,  $\beta_1$  is the effect of age group (Experiment 1) / catch condition (Experiment 2) on the intercept,  $\beta_2$  is the slope relating the Stroop congruency effect for the flanker congruent trials to the Stroop congruency effect for the flanker incongruent trials,  $\beta_3$  is the effect of age group (Experiment 1) / catch condition (Experiment 2) on the slope. Boldface type indicates *p* < .05



**Fig. 4** Experiment 2. Example of one trial sequence presented when flanking words were displayed horizontally (horizontal flanker position, top part) and when flanking words were displayed vertically (vertical flanker position, bottom part). Participants were asked to indicate the color of the central word while ignoring the meaning of the words and the flanking words

part). As this horizontal alignment might create some distance between the central and the flanking words, the flankers were also presented vertically (see Fig. 4, bottom part). That is, for the other half of participants, flanking words were presented above or below the central word. As in Experiment 1, if a large Stroop congruency effect is critical (Hommel, 1997), participants in the 25%-catch condition—compared with those in the 0%-catch condition—would show a *larger reduction* of the Stroop congruency effect in flanker incongruent trials than in flanker congruent trials.

## Method

**Participants** In total, 97 new young participants were tested. Two participants were removed (one because of

accuracy rates at chance level, and one because of a technical error). The final sample consisted of 95 participants (74 women, mean age = 23.5 years,  $SD = 3.5$ ). Twenty-three participants were assigned to the 25%-catch condition with vertical flankers; 24 were assigned to each of the remaining conditions (i.e., 25%-catch condition with horizontal flankers, 0%-catch condition with vertical flankers, and 0%-catch condition with horizontal flankers). At study completion, participants received CHF 15 (about US\$15) or course credits.

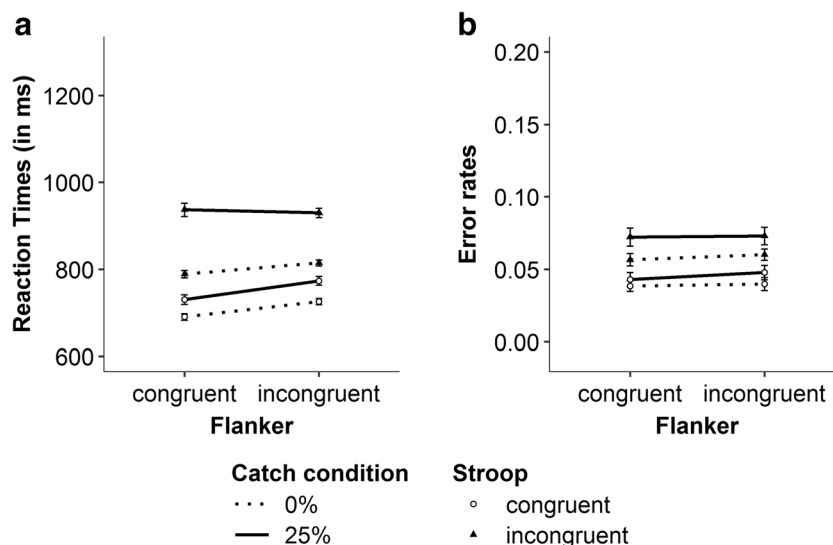
**Material** Material was the same as in Experiment 1, except for the following modifications. First, the stimuli were presented in 40-point Arial Bold font for the stimulus–response mapping block, and 28-point Arial Bold font for the pure Stroop and flanker blocks. Second, for the pure flanker blocks, each letter was presented three times in a row for the horizontal flanker position, and in a column for the vertical flanker position. The distance between the letters was 5 pixels in the horizontal flanker position and 25 pixels in the vertical flanker position. Third, for the mixed blocks, the same three color words were presented either in a row for the horizontal flanker position or in a column for the vertical flanker position (see Fig. 4). The central word was displayed either in the same color as in the meaning of the word (Stroop congruent) or in a different color (Stroop incongruent). The flanking words were presented either in the same color as the color of the central word (flanker congruent) or in a different color (flanker incongruent).

**Procedure** The procedure was similar to Experiment 1, except for the following modifications. First, in the mixed blocks, participants were instructed to indicate the color of the central color word while ignoring the meaning of the word and the flanking color words. Second, each pure block included 192 trials, and each mixed blocks included 144 trials. Third, in the 0%-catch condition, no catch trials were presented. The 25%-catch condition was similar to Experiment 1 with 25% of the trials being catch trials. Overall, the experiment consisted of one session of about 90 minutes.

**Data preparation** The data preparation was similar to Experiment 1, except for the following modifications. The trimming procedure dismissed 21.95% of the trials for the mixed blocks (44.36% when only the 25%-catch condition was considered), and 22.91% of the trials for the pure Stroop blocks (46.30% when only the 25%-catch condition was considered). The additional exclusions for RTs removed 11.86% of the trials for the mixed blocks, 13.07% of the trials for the pure Stroop blocks, and 9.40% of the trials for the pure flanker blocks.

**Data analysis** Analyses for the pure blocks are presented in Appendix B (i.e., Tables 13 and 14, as well as Fig. 10), and





**Fig. 5** Experiment 2. Modulation of the Stroop congruency effect by flanker congruency within a trial in both catch conditions (0% and 25%). Mean reaction times (RTs, **a**) and mean raw error rates (**b**) for

both Stroop and flanker congruency. Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008)

analyses for the catch trials of the mixed blocks are presented in Appendix C (see Table 17 and Fig. 13). As in Experiment 1, the focus was on performance on standard (no-catch) trials, and ANOVAs from both a NHST and Bayesian hypothesis testing approach were conducted. That is, a four-way

repeated-measures ANOVA were carried out with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-subject factors, and catch condition (0%, 25%) and flanker position (horizontal, vertical) as between-subjects factors. In a further step, a state-trace

**Table 7** Experiment 2: Inferential statistical values for the four-way analysis of variance (ANOVA), with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-

subject factors, and catch condition (0%, 25%) and flanker position (horizontal, vertical) as between-subjects factors, and Bayes factors (BF) from the model comparisons

Effect–Model	Reaction times					Arcsine transformed error rates				
	ANOVA		Bayesian analysis			ANOVA		Bayesian analysis		
	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Stroop congruency	446.19	<.001	.21	$9.93 \times 10^{64}$	$1.01 \times 10^{-65}$	90.67	<.001	.09	$2.11 \times 10^{20}$	$4.75 \times 10^{-21}$
Flanker congruency	42.10	<.001	.008	2.03	0.49	2.45	.121	.002	0.23	4.41
Catch	10.93	.001	.10	21.56	0.05	0.13	.717	.001	0.31	3.23
Flanker position	0.01	.932	<.001	0.29	3.44	1.50	.224	.01	0.51	1.97
Stroop congruency × Flanker congruency	17.66	<.001	.003	4.15	0.24	0.31	.578	<.001	0.17	5.81
Stroop congruency × Catch	45.25	<.001	.03	$1.51 \times 10^{12}$	$6.61 \times 10^{-13}$	6.01	.016	.006	6.64	0.15
Stroop congruency × Flanker position	0.21	.648	<.001	0.17	5.74	0.01	.925	<.001	0.16	6.21
Flanker congruency × Catch	2.85	.095	<.001	0.19	5.32	0.08	.778	<.001	0.17	5.82
Flanker congruency × Flanker position	0.86	.356	<.001	0.17	5.89	0.21	.646	<.001	0.16	6.14
Catch × Flanker position	3.25	.075	.03	1.17	0.86	5.64	.020	.05	2.92	0.34
Stroop congruency × Flanker congruency × Catch	8.25	.005	.001	1.53	0.65	1.70	.196	.001	0.40	2.51
Stroop congruency × Flanker congruency × Flanker position	1.16	.285	<.001	0.30	3.36	0.04	.842	<.001	0.21	4.77
Stroop congruency × Catch × Flanker position	7.21	.009	.004	37.35	0.03	0.87	.352	<.001	0.35	2.83
Flanker congruency × Catch × Flanker position	2.14	.147	<.001	0.23	4.28	0.88	.350	<.001	0.27	3.74
Four-way interaction	2.61	.110	<.001	0.47	2.12	2.11	.150	.001	0.62	1.61

Note. *d*f<sub>s</sub> = (1, 91) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

**Table 8** Experiment 2: Inferential statistical values for the follow-up two-way repeated-measures analysis of variance (ANOVA), with the factors Stroop congruency (incongruent, congruent) and flanker congruency

(incongruent, congruent), and Bayes factors (BF) from the model comparisons for each catch condition (0%, 25%) separately

Catch condition/Effect–Model	ANOVA				Bayesian analysis	
	<i>df</i>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
0%-catch condition						
Stroop congruency	1, 47	181.37	<.001	.13	2.20 x 10 <sup>30</sup>	4.54 x 10 <sup>-31</sup>
Flanker congruency	1, 47	48.21	<.001	.02	17.74	0.06
Two-way interaction	1, 47	1.53	.223	<.001	0.31	3.27
25%-catch condition						
Stroop congruency	1, 46	246.65	<.001	.28	1.03 x 10 <sup>41</sup>	9.74 x 10 <sup>-42</sup>
Flanker congruency	1, 46	8.63	.005	.004	0.26	3.89
Two-way interaction	1, 46	16.68	<.001	.007	10.57	0.09

*Note.* Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

analysis was performed similar to Experiment 1, except that age group (young vs. older) was replaced by catch condition (0% vs. 25%).

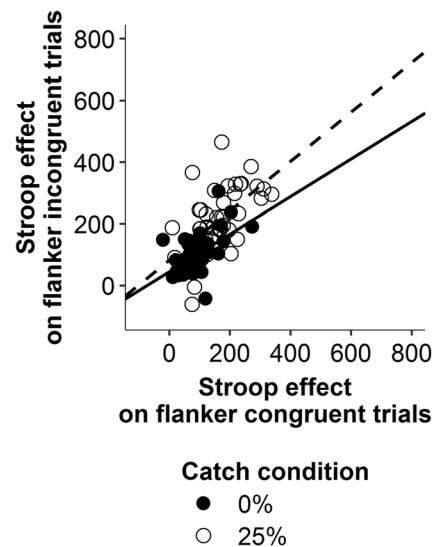
**Results**

The main objective of the study was to determine the impact of catch trials on the interaction between Stroop and flanker congruency. First, the results of the ANOVAs from the NHST and Bayesian hypothesis testing approach are reported. Then, the results of the state-trace analysis are presented.

**Analysis of variance** For the ANOVAs, the most relevant results are the Stroop congruency effect, depending on whether the trials are flanker congruent or incongruent in both catch conditions (0% and 25%). The descriptive results are depicted in Fig. 5. The results of the four-way ANOVA are presented in Table 7.

Of most interest, the ANOVA for RTs showed a significant interaction between Stroop congruency, flanker congruency, and catch condition. The Bayesian analysis suggested, however, only weak evidence in favor of the model including the three-way interaction. This interaction was disentangled by conducting follow-up two-way repeated-measures ANOVAs, with the factors Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) for each catch condition separately. As shown in Table 8, the interaction between Stroop congruency and flanker congruency was not significant in the 0%-catch condition. In line with the NHST analysis, the Bayesian analysis suggested positive evidence against the interaction model. Thus, the magnitude of the Stroop congruency effect was relatively similar across both flanker incongruent and congruent trials (89 and 99 ms, respectively). In both trial types, it was significant,  $t(47) = 11.96, p < .001, d = 1.73, BF_{10} = 8.87 \times$

$10^{12}, BF_{01} = 1.13 \times 10^{-13}$ , and,  $t(47) = 11.68, p < .001, d = 1.69, BF_{10} = 3.91 \times 10^{12}, BF_{01} = 2.55 \times 10^{-13}$ , respectively. In the 25%-catch condition, however, the two-way ANOVA showed a significant interaction between Stroop congruency and flanker congruency, and the Bayesian analysis suggested positive evidence in favor of the interaction model (see Table 8). Thus, the Stroop congruency effect was smaller when the stimuli were flanker incongruent (156 ms),  $t(46) = 14.30, p < .001, d = 2.09, BF_{10} = 3.41 \times 10^{15}, BF_{01} = 2.93 \times 10^{-16}$ , compared with when they were flanker congruent (206 ms),  $t(46) = 13.83, p < .001, d = 2.02, BF_{10} = 1.02 \times 10^{15}, BF_{01} = 9.76 \times 10^{-16}$ . Together, this shows an underadditive interaction between Stroop and flanker congruency in the 25%-catch



**Fig. 6** Experiment 2: State-trace plot. Solid line results from the regression analysis using the 0%-catch condition only. Dotted line results from the regression analysis using the 25%-catch condition only. Reaction times are given in milliseconds

condition, but no interaction in the 0%-catch condition (see Fig. 5).

For the error rates, the ANOVA from the NHST revealed no significant interaction including Stroop congruency and flanker congruency. In line with the NHST approach, the Bayesian hypothesis testing approach suggested small to positive evidence against the models including such interactions. Therefore, no speed–accuracy trade-offs compromised the effects of interest.

**State-trace analysis** The state-trace plot is depicted in Fig. 6. Goodness-of-fit statistics and model comparison results are presented in Table 5 (lower part). Estimates of the fixed parameters are summarized in Table 6 (lower part). As shown in Table 5, model fit was better for the model without the interaction term than the model including the interaction term. Moreover, the  $\Delta\chi^2$  was not significant, and the Bayesian analysis suggested positive evidence for the model without the interaction term. Thus, the data displayed in the state-trace plot were better explained with one single line (see Fig. 6).

In sum, the state-trace results speak against a difference in the Stroop congruency effect between both catch conditions in the relation between flanker congruent and incongruent trials. Therefore, no process was added in the 25%-catch condition compared with the 0%-catch condition. This suggests that the interaction between Stroop and flanker congruency became significant in the 25%-catch condition only because the Stroop congruency effect was larger in that condition than in the 0%-catch condition.

## Discussion

At first, the results of Experiment 2 suggested that the interaction between Stroop and flanker congruency only occurred when catch trials were occasionally presented (i.e., in the 25%-catch condition). However, the state-trace analysis indicated that the interaction was found in that condition only because the Stroop congruency effect was larger than in the condition without catch trials (i.e., the 0%-catch condition). These findings are in line with the view that a large Stroop congruency effect is a necessary condition for observing a significant interaction between Stroop and flanker congruency (Hommel, 1997).

## Experiment 3

The goal of Experiment 3 was to increase the size of the Stroop congruency effect without presenting catch trials. Previous research has shown that the Stroop congruency effect was larger when responses were given vocally (e.g., Sharma & McKenna, 1998; Shichel & Tzelgov, 2018). Experiment 3 thus had the same design as the 0%-catch condition of the previous experiment, but participants were asked to respond vocally. Again, if a large

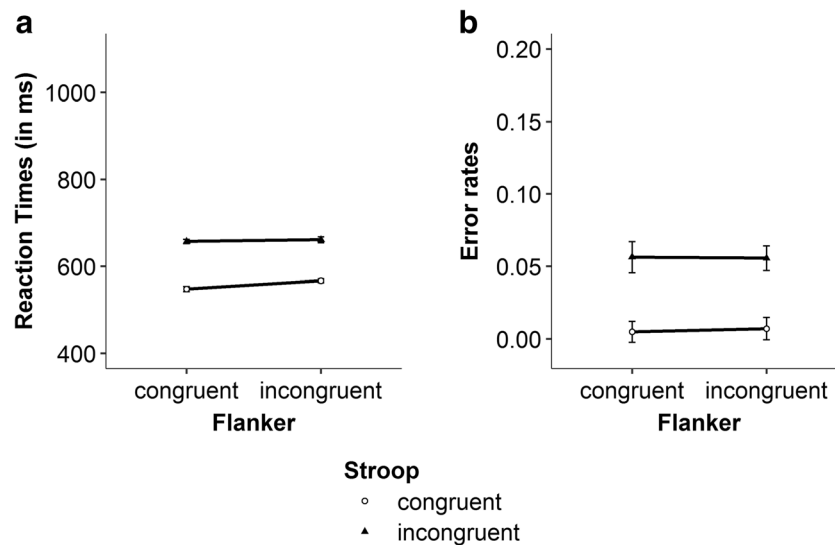
Stroop congruency effect is the necessary condition for observing an interaction (Hommel, 1997), the interaction between Stroop and flanker congruency would be expected to be significant in Experiment 3.

## Method

**Participants** In total, 28 new young participants were tested. Two participants were removed because they already participated in a previous study using a similar design, and three participants were removed because the computer malfunctioned. The final sample consisted of 23 participants (21 women, mean age = 21.7 years,  $SD = 2.2$ ). The study was carried out according to the guidelines of the ethics committee of the Catholic University of Eichstätt-Ingolstadt. Informed consent was acquired from all participants. At study completion, participants received 8 € per hour (about US\$9) or course credits.

**Material and procedure** Material and procedure were similar to the 0%-catch condition with vertical flankers of Experiment 2, except that participants were asked to respond vocally. To this end, a microphone was used, and the experiment was programmed using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The stimulus remained on-screen until the participant responded vocally into the microphone, which stopped the timer and eliminated the stimulus from the screen. RT in milliseconds was measured from the stimulus onset until the participant responded. Then, the screen blanked for 500 ms. During this blank screen, the experimenter typed whether or not the response was correct so that the computer could evaluate errors. In case of an error, the word “Fehler” (i.e., “error” in English) was displayed for 500 ms. Finally, an intertrial interval of 500 ms occurred. In addition, the pure Stroop and flanker blocks were divided into miniblocks of 96 trials each so that participants had more breaks. In the mixed and pure flanker blocks, the distance between the flankers and the central stimulus was 0.1 cm. All stimuli were presented in uppercase in 28-point Arial Bold font.

**Data preparation** The same data preparation was used as in Experiment 1, except that trials in which the microphone was malfunctioning and the experimenter did not code the accuracy of the response were additionally excluded (i.e., 1.01% of trials for the mixed blocks, 1.51% for the pure Stroop blocks, and 1.54% for the pure flanker blocks). Overall, the trimming procedure dismissed 1.70% of the trials for the mixed blocks, 2.03% of the trials for the pure Stroop blocks, and 2.05% of the trials for the pure flanker blocks. The additional exclusions for RTs removed 8.19% of the trials for the mixed blocks, 6.82% of the trials for the pure Stroop blocks, and 4.69% of the trials for the pure flanker blocks.



**Fig. 7** Experiment 3. Modulation of the Stroop congruency effect by flanker congruency within a trial. Mean reaction times (RTs, **a**) and mean raw error rates (**b**) for both Stroop and flanker congruency. Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008)

**Data analysis** Analyses on pure blocks are reported in Appendix B (see Tables 14 and 15 as well as Fig. 11). Performance from the mixed blocks were analyzed using both a NHST and Bayesian hypothesis testing approach. That is, a two-way repeated-measures ANOVA was conducted with the factors Stroop congruency (incongruent, congruent), and flanker congruency (incongruent, congruent).

**Results**

In Experiment 3, the focus was on the modulatory effect of the flanker congruency on the Stroop congruency effect. The most relevant results are thus the RTs from the Stroop congruency effect, depending on whether the trials are flanker congruent or incongruent. The descriptive results are depicted in Fig. 7. The results of the two-way ANOVA are shown in Table 9.

Of most interest, the ANOVA on the RTs showed a significant two-way interaction between Stroop congruency and flanker

congruency. Bayesian analysis suggested small evidence in favor of the interaction model. The interaction was disentangled by conducting follow-up *t* tests. The Stroop congruency effect was smaller when the stimuli were flanker incongruent (94 ms),  $t(22) = 17.03, p < .001, d = 3.55, BF_{10} = 1.65 \times 10^{11}, BF_{01} = 6.04 \times 10^{-12}$ , compared with when they were flanker congruent (109 ms),  $t(22) = 20.63, p < .001, d = 4.30, BF_{10} = 7.26 \times 10^{12}, BF_{01} = 1.38 \times 10^{-13}$ . For the error rates, the ANOVA from the NHST approach showed no significant interaction, and Bayesian analysis suggested positive evidence against the interaction model (see Table 9). Together, this shows an underadditive interaction between Stroop and flanker congruency for the RTs.

**Discussion**

The results of Experiment 3 showed an underadditive interaction in RTs between Stroop and flanker congruency when participants were asked to respond vocally and when no catch trials were presented. These findings further emphasize the

**Table 9** Experiment 3: Inferential statistical values for the two-way repeated-measures analysis of variance (ANOVA), with the factors Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent), and Bayes factors (BF) from the model comparisons

Effect-Model	Reaction times					Arcsine transformed error rates				
	ANOVA		$\eta_g^2$	Bayesian analysis		ANOVA		$\eta_g^2$	Bayesian analysis	
<i>F</i>	<i>p</i>	BF <sub>10</sub>		BF <sub>01</sub>	<i>F</i>	<i>p</i>	BF <sub>10</sub>		BF <sub>01</sub>	
Stroop congruency	398.39	<.001	.34	1.45 x 10 <sup>31</sup>	6.88 x 10 <sup>-32</sup>	63.44	<.001	.48	1.07 x 10 <sup>15</sup>	9.32 x 10 <sup>-16</sup>
Flanker congruency	7.65	.011	.006	0.30	3.32	0.61	.443	.003	0.24	4.14
Two-way interaction	16.28	<.001	.003	1.15	0.87	0.69	.414	.002	0.32	3.13

Note. *dfs* = (1, 22) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

magnitude of the Stroop congruency effect as the key component in observing an interaction between congruency conditions of conflict tasks in RTs (Hommel, 1997).

## General discussion

The purpose of the present study was to determine whether a large congruency effect is a necessary condition for observing an interaction in RTs between the congruency conditions of two conflict tasks combined within the same trial. To this end, the Stroop and flanker tasks were integrated within the same trial by asking participants to respond the color of the central letter or word while ignoring the meaning of the word and the color of the flanking letters or words. Critically, the size of the Stroop congruency effect was increased by three different manipulations. First, older adults were tested (Experiment 1). Second, the proportion of catch trials (0% vs. 25%)—that is, trials in which participants were asked to respond to the meaning of the word instead of responding to the color of the color word—was manipulated (Experiment 2). Third, participants were asked to respond vocally (Experiment 3). Data were analyzed using different analyses: ANOVA from the NHST approach, Bayesian ANOVA from the Bayesian hypothesis testing approach, and a state-trace analysis. The results of the present study are summarized in Fig. 8, which illustrates the magnitude of the interaction and of the Stroop congruency effect for each participant of each experimental condition in each experiment, as well as both effects averaged across experimental conditions in each experiment.

The results showed an interaction in RTs between Stroop and flanker congruency in conditions with a large congruency effect, such as when young and older adults performed a single-letter Stroop task with 25% of catch trials (Experiment 1), when young adults performed a standard Stroop task with 25% of catch trials (Experiment 2, 25%-catch condition), and when young adults responded vocally to the standard Stroop task (without any catch trials, Experiment 3). In all these conditions, the Stroop congruency effect was larger than 100 ms. There was only one condition in which no significant interaction between the congruency conditions was observed, namely when young adults performed the standard Stroop task without any catch trials (Experiment 2, 0%-catch condition). In this condition, the congruency effect was slightly smaller than 100 ms (i.e., 94 ms). Moreover, as shown in Fig. 8, the size of the interaction was found to be affected by the magnitude of the Stroop congruency effect. This was also observed in Experiment 1 in which older adults were tested in addition to young adults. In this experiment, the reduction of the Stroop congruency effect in flanker incongruent trials was still substantial,

even after differences in general processing speed between young and older adults was taken into consideration by applying a log-transformation to all RTs.

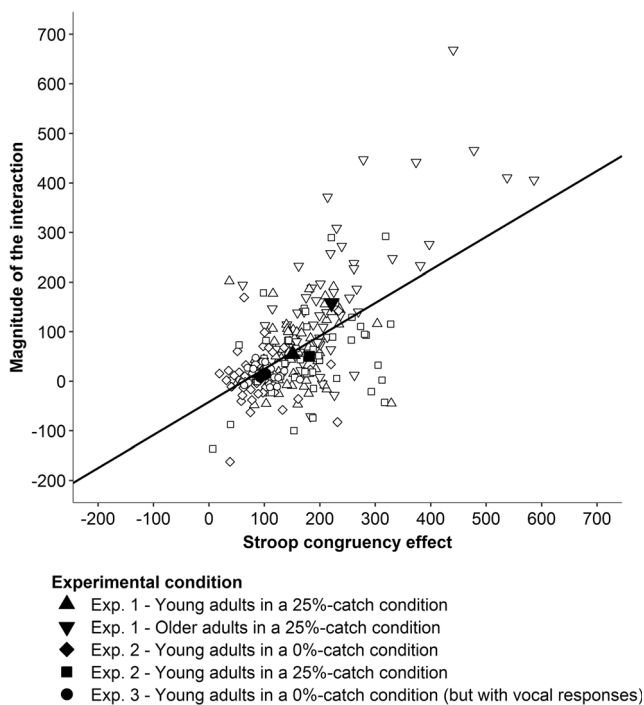
At first glance, it might be surprising that as soon as the Stroop congruency effect was slightly under 100 ms (i.e., in the 0%-catch condition of Experiment 2), the interaction was no longer significant. In contrast, it was sufficient that the Stroop congruency effect was slightly above the limit of 100 ms to observe a significant interaction between Stroop and flanker congruency (e.g., 102 ms in Experiment 3). However, participants were overall faster at responding in Experiment 3 than in Experiment 2 (608 ms vs. 755 ms for the 0%-catch condition, respectively). Thus, when the Stroop congruency effect is expressed as a proportional score (i.e., as the difference in RTs between Stroop incongruent and congruent trials, divided by the RTs of Stroop congruent trials), the differences between Experiment 2 (0%-catch) and Experiment 3 become more evident (0.12 and 0.17, respectively).<sup>2</sup>

This suggests that it is not the magnitude of the Stroop congruency effect per se which is critical for the interaction between Stroop and flanker congruency. It is rather the magnitude in relationship with the overall response speed (Frühholz et al., 2011; Hommel, 1997; but see Rey-Mermet & Gade, 2016). This possibility was investigated by computing delta plots for each experiment (see Appendix D). Delta plots illustrate the changes in the congruency effect as a function of response speed (Pratte, Rouder, Morey, & Feng, 2010). In case of the Stroop task, delta plots show an increase of the congruency effect as responses slow down. As shown in Appendix D, this pattern of changes was similar for flanker incongruent trials and flanker congruent trials across the different conditions and experiments. Critically, as soon as the mean RTs were slower than ca. 800 ms, the size of the interaction—that is, the difference in the Stroop congruency effect between flanker incongruent and flanker congruent trials—also increased. This occurred irrespective of the conditions (young vs. older, 0%-catch vs. 25%-catch) and the experiments. These results are in line with the findings of the state-trace analysis showing that the Stroop congruency effect was simply larger for the older adults than for the young adults in Experiment 1, and for the 25%-catch condition than for the 0%-catch condition in Experiment 2. Thus, although there was a difference between age groups and between both catch conditions, no additional process is necessary to explain the interaction.

Now, the question is: Why is a Stroop congruency effect of 100 ms a limit to observe the interaction between Stroop and flanker tasks? First, the reason of that limit might be some biological constraints, such as the speed of execution when pressing

<sup>2</sup> For the sake of completeness, the proportional Stroop congruency effect for the other cases is also reported: that is, 0.20 for the young adults in Experiment 1, 0.21 for the older adults in Experiment 1, as well as 0.22 for the 25%-catch condition in Experiment 2.





**Fig. 8** Modulation of the interaction between Stroop and flanker congruency (i.e. the Stroop congruency effect in flanker congruent trials minus the Stroop congruency effect in flanker incongruent trials) by the magnitude of the Stroop congruency effect. Reaction times are given in milliseconds. In this scatterplot, each empty point represents a participant of each experimental condition (i.e., young or older adults either in the 0%- or 25% catch condition for Experiment 1, Experiment 2, or Experiment 3). Each solid point represents the mean for each experimental condition and each experiment. The solid line results from the linear mixed model in which the magnitude of the interaction was predicted by the fixed Stroop congruency effect with experimental condition and participant as random intercept effects. The slope relating the Stroop congruency effect to the magnitude of the interaction was significant (slope = 0.67, standard error = .07,  $t(226.14) = 10.07, p < .001$ )

the response key or when articulating the response. Another possibility may be related to the processes underlying the Stroop congruency effect. In previous research, the Stroop congruency effect has been assumed to result from at least two processes, such as the interference induced by incongruent trials and the priming or facilitation induced by congruent trials (e.g., Glaser & Glaser, 1982; Tzelgov, Henik, & Berger, 1992), or the processes induced by a task conflict and those induced by the response conflict (e.g., Steinhauser & Hübner, 2009). A limit of 100 ms in the Stroop congruency effect may suggest that when the Stroop congruency effect is smaller than that limit, one of the two processes or conflicts is not present or not sufficiently processed, thus resulting in the absence of the interaction between the congruency conditions. In particular, in a previous experiment (Rey-Mermet & Gade, 2016, Experiment 1c), the necessity of interference in Stroop incongruent trials was put forward because the interaction between Stroop and flanker congruency was observed when the Stroop congruency effect was measured as the

difference between incongruent trials and neutral trials (i.e., rows of colored Xs), but not when it was measured as the difference between congruent and neutral trials. Thus, it is possible that experiencing interference in the Stroop incongruent trials is necessary to observe the interaction between Stroop and flanker congruency, in particular to preactivate the correct response as proposed in Rey-Mermet et al. (2019) account.

In the present study, we increased the Stroop congruency effect in three different ways, that is: (1) by testing older adults, (2) by manipulating the proportion of catch trials in which participants were asked to respond to the word meaning instead of responding to the print color, and (3) by asking participants to respond vocally. One may wonder whether the cognitive processes triggered by each of these manipulations were not important for the interaction to occur and whether these processes did not differ across the different manipulations. For example, as catch trials require participants to switch between two decisions (i.e., responding to the word meaning vs. responding to the print color), the cognitive processes underlying Stroop performance may be affected by this task switching (see, e.g., Kalanthroff & Henik, 2014). Whether these cognitive processes are affected in the same way by vocal responding is so far an open question. To provide a well-founded response, further research is necessary in which Stroop performance in the flanker–Stroop combination would be decomposed by using ex-Gaussian distribution analysis (e.g., Steinhauser & Hübner, 2009) or diffusion modeling (e.g., White, Servant, & Logan, 2018), for example.

To summarize, the findings of the present study show that when conflict tasks—such as the Stroop and flanker tasks—are combined within the same trial, finding an interaction in RTs between the congruency conditions of both tasks requires a large congruency effect, which, in turn, requires slow responses. This emphasizes that finding interactions of that sort can be used to validate or invalidate theoretical explanations only when the precondition—that is, a large congruency effect—is fulfilled.

**Acknowledgements** This work was supported by a grant from the Swiss National Science Foundation (Grant P300P1\_164598) to A. Rey-Mermet. Some results of this article were presented at the 60th Conference of Experimental Psychologists (TeaP), Marburg, Germany, March 2018, and at the 3rd International Meeting of the Psychonomic Society, Amsterdam, The Netherlands, May 2018. The author thanks Danielle Pessach, Julia Spaaij, Paola Keller, Stefanie Ochsenkühn, and Melina Härter for their help in data collection, as well as Miriam Gade and Marco Steinhauser for helpful discussions and comments on a previous version.

The data and materials for all experiments in this work are available at the Open Science Framework (<https://osf.io/nq76s>).

**Author note** Alodie Rey-Mermet, Department of Psychology, General Psychology, Catholic University of Eichstätt-Ingolstadt, Germany, and University of Zurich, Department of Psychology, Cognitive Psychology Unit, Switzerland.

## Appendix A: Sample characteristics and background measures

**Table 10** Experiment 1: Description of the sample characteristics and background measures

Measure	Young adults (subsample <sup>a</sup> )	Older adults	Comparison
Sample size	24	59	
Demographics			
Age (years)	23.92 (2.83)	70.41 (2.74)	$t(41.54) = -68.52, p < .001, d = 16.81, BF_{10} = 5.56 \times 10^{69}, BF_{01} = 1.80 \times 10^{-70}$
Gender (female/male)	18/6	24/35	$\chi^2(1, N = 83) = 6.73, p = .010, d = 0.59, BF_{10} = 19.88, BF_{01} = 0.05$
CERAD-Plus tests			
Boston Naming Test (accuracy)	0.95 (0.11)	0.97 (0.06)	$t(29.33) = -0.62, p = .538, d = 0.19, BF_{10} = 0.32, BF_{01} = 3.11$
Figure recall (accuracy)	0.95 (0.10)	0.87 (0.16)	$t(66.71) = 2.56, p = .013, d = 0.51, BF_{10} = 1.67, BF_{01} = 0.60$
Word list recall (accuracy)	0.90 (0.05)	0.76 (0.11)	$t(74.60) = 2.48, p = .015, d = 0.47, BF_{10} = 1.25, BF_{01} = 0.80$
Word list recognition (discriminability) <sup>b</sup>	1 (0)	0.99 (0.03)	$t(58.00) = 4.20, p < .001, d = 0.65, BF_{10} = 4.89, BF_{01} = 0.20$
Verbal fluency (number of words)	24.58 (7.27)	23.61 (4.87)	$t(31.74) = 0.60, p = .551, d = 0.17, BF_{10} = 0.31, BF_{01} = 3.24$
S-words fluency (number of words)	17.38 (4.71)	14.75 (4.65)	$t(42.26) = 2.31, p = .022, d = 0.56, BF_{10} = 2.42, BF_{01} = 0.41$
TMT score (in sec) <sup>c</sup>	1.84 (0.58)	2.14 (0.73)	$t(52.83) = -1.95, p = .057, d = 0.43, BF_{10} = 0.95, BF_{01} = 1.06$
Health			
Physical index (standardized score) <sup>d</sup>	56.61 (5.73)	52.54 (5.69)	$t(42.46) = 2.94, p = .005, d = 0.71, BF_{10} = 9.27, BF_{01} = 0.11$
Mental index (standardized score) <sup>d</sup>	45.59 (7.91)	57.00 (4.37)	$t(28.90) = -6.66, p < .001, d = 2.03, BF_{10} = 4.38 \times 10^9, BF_{01} = 2.28 \times 10^{-10}$

*Note.* Standard deviations (*SD*) are given in parentheses. Because of unequal sample size, Welch *t* tests were performed. Effect sizes are expressed with Cohen’s *d*. CEARD-Plus = Consortium to Establish a Registry for Alzheimer’s Disease-Plus.  $BF_{10}$  = Bayes factor in favor the alternative hypothesis (i.e., an age-related difference).  $BF_{01}$  = Bayes factor in favor the null hypothesis (i.e., in favor of the absence of an age-related difference)

<sup>a</sup> Only a subsample of young adults participated in the previous study (Rey-Mermet et al., 2018) and thus had background measures. Overall, 67 young adults were tested in Experiment 1 ( $M_{age} = 23.92, SD_{age} = 3.56, 56$  females, 11 males)

<sup>b</sup> The discriminability score was computed as the difference  $\{1 - [(10 - hits) + (10 - correct\ rejections)] / 20\}$

<sup>c</sup> Trail Making Test (TMT) score was calculated as the time required to connect alternatively numbers and letters (TMT-B) divided by the time required to connect numbers only (TMT-A)

<sup>d</sup> Higher scores indicate better health status

## Appendix B: Performance on the pure Stroop and flanker blocks

**Table 11** Experiment 1: Inferential statistical values for the two-way analysis of variance (ANOVA), with congruency (incongruent, congruent) as a within-subject factor and age group (young, older) as a between-subjects factor, and Bayes factors (BF) from the model comparisons (upper part: pure Stroop blocks separately for each trial type [standard vs. catch trials]; lower part: pure flanker blocks)

Block type	Trial type	Dependent measure	Effect-Model	ANOVA			Bayesian analysis	
				<i>F</i>	<i>p</i>	$\eta_g^2$	$BF_{10}$	$BF_{01}$
Stroop	Standard	Reaction times	Congruency	737.11	<.001	.41	$2.92 \times 10^{41}$	$3.43 \times 10^{-42}$
			Age group	157.90	<.001	.53	$7.17 \times 10^{19}$	$1.39 \times 10^{-20}$
			Interaction	57.06	<.001	.05	$4.78 \times 10^8$	$2.09 \times 10^{-9}$
	Log-transformed reaction times		Congruency	1,017.25	<.001	.42	$5.70 \times 10^{54}$	$1.76 \times 10^{-55}$
			Age group	187.26	<.001	.58	$4.31 \times 10^{22}$	$2.32 \times 10^{-23}$
			Interaction	17.18	<.001	.01	265.71	$3.76 \times 10^{-3}$
	Arcsine transformed error rates		Congruency	179.93	<.001	.33	$1.02 \times 10^{24}$	$9.79 \times 10^{-25}$
			Age group	38.33	<.001	.17	$6.45 \times 10^4$	$1.55 \times 10^{-5}$
			Interaction	1.70	.195	.005	0.41	2.45

**Table 11** (continued)

Block type	Trial type	Dependent measure	Effect–Model	ANOVA			Bayesian analysis	
				<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Flanker	Catch	Reaction times	Congruency	462.05	<.001	.24	1.25 × 10 <sup>40</sup>	7.98 × 10 <sup>-41</sup>
			Age group	124.11	<.001	.48	2.17 × 10 <sup>17</sup>	4.61 × 10 <sup>-18</sup>
			Interaction	0.83	.363	<.001	0.27	3.65
		Log-transformed reaction times	Congruency	560.97	<.001	.26	7.02 × 10 <sup>42</sup>	1.43 × 10 <sup>-43</sup>
			Age group	137.08	<.001	.50	4.97 × 10 <sup>18</sup>	2.01 × 10 <sup>-19</sup>
			Interaction	13.04	<.001	.008	50.02	0.02
	Arcsine transformed error rates	Congruency	254.85	<.001	.44	2.04 × 10 <sup>34</sup>	4.89 × 10 <sup>-35</sup>	
		Age group	9.28	.003	.04	1.97	0.51	
		Interaction	0.64	.426	.002	0.27	3.76	
	Standard	Reaction times	Congruency	225.33	<.001	.03	4.49 × 10 <sup>24</sup>	2.23 × 10 <sup>-25</sup>
			Age group	136.76	<.001	.52	1.74 × 10 <sup>18</sup>	5.74 × 10 <sup>-19</sup>
			Interaction	10.44	.002	.001	15.67	0.06
Log-transformed reaction times		Congruency	324.50	<.001	.04	7.42 × 10 <sup>32</sup>	1.35 × 10 <sup>-33</sup>	
		Age group	182.04	<.001	.59	3.61 × 10 <sup>22</sup>	2.77 × 10 <sup>-23</sup>	
		Interaction	1.17	.282	<.001	0.31	3.24	
Arcsine transformed error rates		Congruency	7.28	.008	.01	2.64	0.38	
		Age group	34.25	<.001	.18	2.72 × 10 <sup>5</sup>	3.68 × 10 <sup>-6</sup>	
		Interaction	4.08	.045	.006	1.13	0.89	

Note. *dfs* = (1, 125) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

**Table 12** Experiment 1: Inferential statistical values for the *t* test comparing incongruent and congruent trials, and Bayes factors (BF) from the model comparisons (upper part: pure Stroop blocks separately for each trial type [standard vs. catch trials]; lower part: pure flanker blocks)

Block type	Trial type	Dependent measures	Age group	<i>t</i> test				Bayesian <i>t</i> test	
				<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>	BF <sub>10</sub>	BF <sub>01</sub>
Stroop	Standard	Reaction times	Young	67	20.10	<.001	2.44	6.80 × 10 <sup>26</sup>	1.47 × 10 <sup>-27</sup>
			Older	58	18.96	<.001	2.47	2.20 × 10 <sup>23</sup>	4.54 × 10 <sup>-24</sup>
		Log-transformed reaction times	Young	67	23.76	<.001	2.88	1.07 × 10 <sup>31</sup>	9.36 × 10 <sup>-32</sup>
			Older	58	21.54	<.001	2.80	1.25 × 10 <sup>26</sup>	7.98 × 10 <sup>-27</sup>
		Arcsine transformed error rates	Young	67	11.04	<.001	1.34	6.22 × 10 <sup>13</sup>	1.61 × 10 <sup>-14</sup>
			Older	58	8.07	<.001	1.05	2.07 × 10 <sup>8</sup>	4.83 × 10 <sup>-9</sup>
	Catch	Reaction times	Young	67	18.13	<.001	2.20	2.18 × 10 <sup>24</sup>	4.58 × 10 <sup>-25</sup>
			Older	58	13.16	<.001	1.71	1.13 × 10 <sup>16</sup>	8.82 × 10 <sup>-17</sup>
		Log-transformed reaction times	Young	67	21.31	<.001	2.58	1.91 × 10 <sup>28</sup>	5.24 × 10 <sup>-29</sup>
			Older	58	12.87	<.001	1.68	4.46 × 10 <sup>15</sup>	2.24 × 10 <sup>-16</sup>
		Arcsine transformed error rates	Young	67	14.71	<.001	1.78	3.74 × 10 <sup>19</sup>	2.68 × 10 <sup>-20</sup>
			Older	58	8.93	<.001	1.16	4.82 × 10 <sup>9</sup>	2.07 × 10 <sup>-10</sup>
Flanker	Standard	Reaction times	Young	67	11.04	<.001	1.34	6.16 × 10 <sup>13</sup>	1.62 × 10 <sup>-14</sup>
			Older	58	10.36	<.001	1.35	8.54 × 10 <sup>11</sup>	1.17 × 10 <sup>-12</sup>
		Log-transformed reaction times	Young	67	13.19	<.001	1.60	1.81 × 10 <sup>17</sup>	5.51 × 10 <sup>-18</sup>
	Older		58	12.27	<.001	1.60	6.14 × 10 <sup>14</sup>	1.63 × 10 <sup>-15</sup>	
	Arcsine transformed error rates	Young	67	0.55	.581	0.07	0.15	6.48	
		Older	58	2.91	.005	0.38	6.32	0.16	

Note. Effect sizes are expressed as Cohen’s *d*. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

**Table 13** Experiment 2—Standard trials: Inferential statistical values for the three-way analysis of variance (ANOVA), with congruency (incongruent, congruent) as a within-subject factor and catch condition (0%, 25%) and flanker position (horizontal, vertical) as between-subjects factors, and Bayes factors (BF) from the model comparisons (upper part: pure Stroop blocks; lower part: pure flanker blocks)

Block type/Effect or Model	Reaction times					Arcsine transformed error rates				
	ANOVA			Bayesian analysis		ANOVA			Bayesian analysis	
	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
<b>Pure Stroop blocks</b>										
Congruency	381.83	<.001	.28	$1.13 \times 10^{23}$	$8.86 \times 10^{-24}$	47.87	<.001	.11	$1.25 \times 10^6$	$7.99 \times 10^{-7}$
Catch	29.47	<.001	.23	$2.03 \times 10^4$	$4.94 \times 10^{-5}$	1.90	.172	.02	0.51	1.96
Flanker position	0.17	.685	.002	0.25	4.03	0.89	.348	.008	0.32	3.15
Congruency × Catch	54.26	<.001	.05	$7.55 \times 10^6$	$1.32 \times 10^{-7}$	16.39	<.001	.04	173.08	0.01
Congruency × Flanker position	0.02	.877	<.001	0.21	4.67	0.35	.557	<.001	0.25	4.02
Catch × Flanker position	3.64	.060	.04	1.17	0.85	2.89	.093	.02	0.93	1.07
Three-way interaction	14.05	<.001	.01	67.99	0.01	2.66	.107	.007	0.86	1.16
<b>Pure flanker blocks</b>										
Congruency	149.17	<.001	.04	$1.22 \times 10^{18}$	$8.18 \times 10^{-19}$	2.49	.118	.003	0.49	2.05
Catch	0.30	.585	.003	0.52	1.93	0.01	.934	<.001	0.40	2.50
Flanker position	0.00	.997	<.001	0.47	2.14	0.53	.467	.005	0.47	2.11
Congruency × Catch	0.26	.610	<.001	0.24	4.22	0.16	.686	<.001	0.23	4.34
Congruency × Flanker position	0.23	.631	<.001	0.24	4.19	0.61	.437	<.001	0.28	3.55
Catch × Flanker position	1.87	.175	.02	0.94	1.06	4.56	.035	.04	1.99	0.50
Three-way interaction	0.07	.791	<.001	0.31	3.23	0.05	.820	<.001	0.30	3.30

*Note.* *d*f<sub>s</sub> = (1, 91) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect). As the three-way interaction between Stroop congruency, catch condition, and flanker position was significant for the RTs of the pure Stroop blocks and the Bayesian analysis suggested strong evidence for the model including this interaction, follow-up two-way ANOVAs were conducted, with Stroop congruency (incongruent, congruent) as a within-subject factor and flanker position (horizontal, vertical) as a between-subjects factor for each catch condition (0%, 25%) separately. The analyses showed a significant interaction between Stroop congruency and flanker position for each catch type, 0% catch:  $F(1, 46) = 12.40, p = .001, \eta_g^2 = .02, BF_{10} = 18.83, BF_{01} = 0.05$ ; and 25% catch:  $F(1, 45) = 5.08, p = .029, \eta_g^2 = .01, BF_{10} = 1.98, BF_{01} = 0.50$ . These interactions were disentangled by carrying out follow-up *t* tests. These revealed a significant Stroop congruency effect for each catch condition in each flanker position, ensuring the presence of the effect in each case, 0% catch with horizontal flanker position: 123 ms,  $t(23) = 8.34, p < .001, d = 1.70, BF_{10} = 6.49 \times 10^5, BF_{01} = 1.54 \times 10^{-6}$ ; 0% catch with vertical flanker position: 67 ms,  $t(23) = 11.15, p < .001, d = 2.28, BF_{10} = 1.02 \times 10^8, BF_{01} = 9.76 \times 10^{-9}$ ; 25% catch with horizontal flanker position: 180 ms,  $t(23) = 10.40, p < .001, d = 2.12, BF_{10} = 2.88 \times 10^7, BF_{01} = 3.47 \times 10^{-8}$ ; 25% catch with vertical flanker position: 241 ms,  $t(22) = 11.48, p < .001, d = 2.39, BF_{10} = 1.03 \times 10^8, BF_{01} = 9.70 \times 10^{-9}$ .

**Table 14** Experiment 2—Catch trials in the pure Stroop blocks: Inferential statistical values for the two-way analysis of variance (ANOVA), with congruency (incongruent, congruent) as a within-subject factor, and flanker position (horizontal, vertical) as a between-subjects factor, and Bayes factors (BF) from the model comparisons

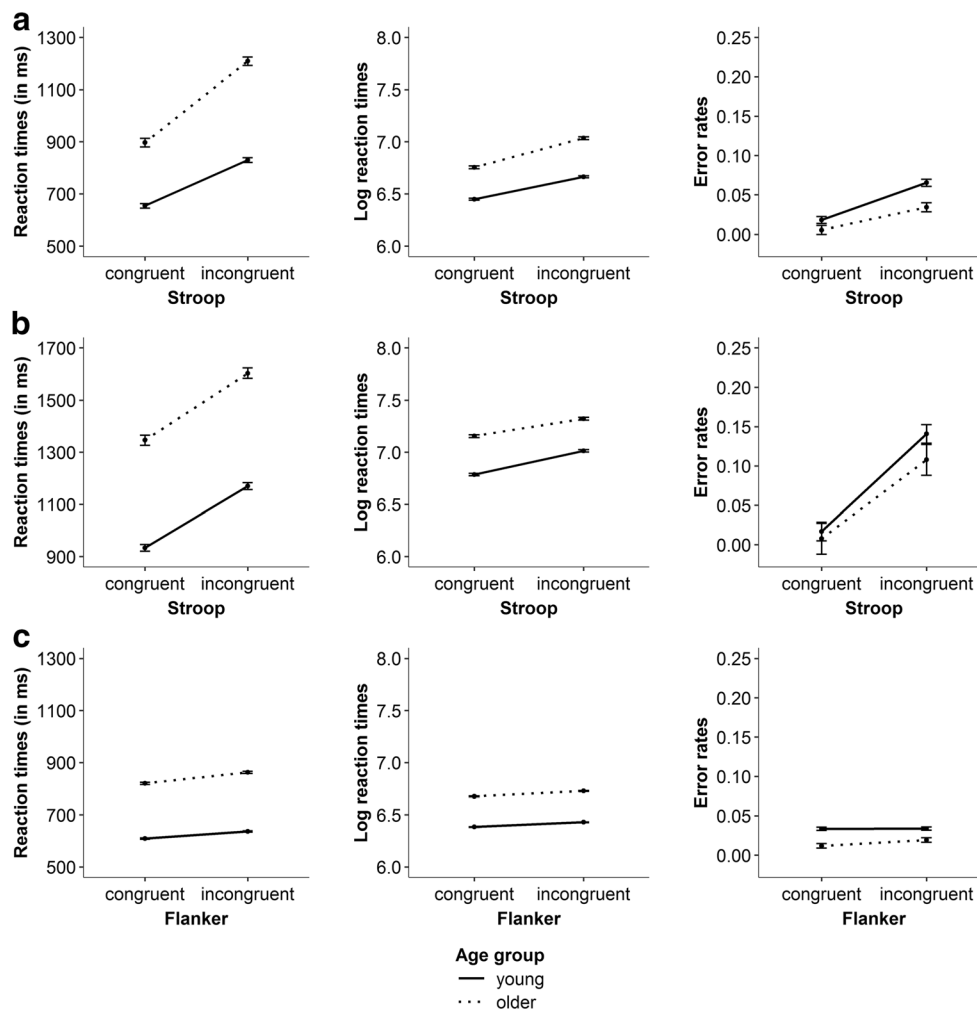
Effect-Model	Reaction times					Arcsine transformed error rates				
	ANOVA			Bayesian analysis		ANOVA			Bayesian analysis	
	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Stroop congruency	207.57	<.001	.27	$2.38 \times 10^{15}$	$4.20 \times 10^{-16}$	244.09	<.001	.63	$2.96 \times 10^{20}$	$3.38 \times 10^{-21}$
Flanker position	0.10	.755	.002	0.31	3.25	4.28	.044	.06	0.52	1.93
Interaction	0.09	.767	<.001	0.29	3.39	6.43	.015	.04	3.84	0.26

*Note.* *d*f<sub>s</sub> = (1, 45) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect). As the ANOVA showed a significant interaction between Stroop congruency and flanker position for the error rates and Bayesian analysis suggested positive evidence in favor of the interaction model, follow-up *t* tests comparing performance on Stroop incongruent and congruent trials were computed for each flanker position separately. In both flanker positions, error rates were higher for incongruent trials than for congruent trials, horizontal flanker position:  $t(23) = 12.27, p < .001, d = 2.50, BF_{10} = 6.09 \times 10^8, BF_{01} = 1.64 \times 10^{-9}$ ; and vertical flanker position:  $t(22) = 9.79, p < .001, d = 2.04, BF_{10} = 6.42 \times 10^6, BF_{01} = 1.56 \times 10^{-7}$ . However, this difference was larger for the horizontal flanker position (0.18) than for the vertical flanker position (0.11)

**Table 15** Experiment 3: Inferential statistical values for the *t* test comparing performance in congruent and incongruent trials, and Bayes factors (upper part: pure Stroop blocks; lower part: pure flanker blocks)

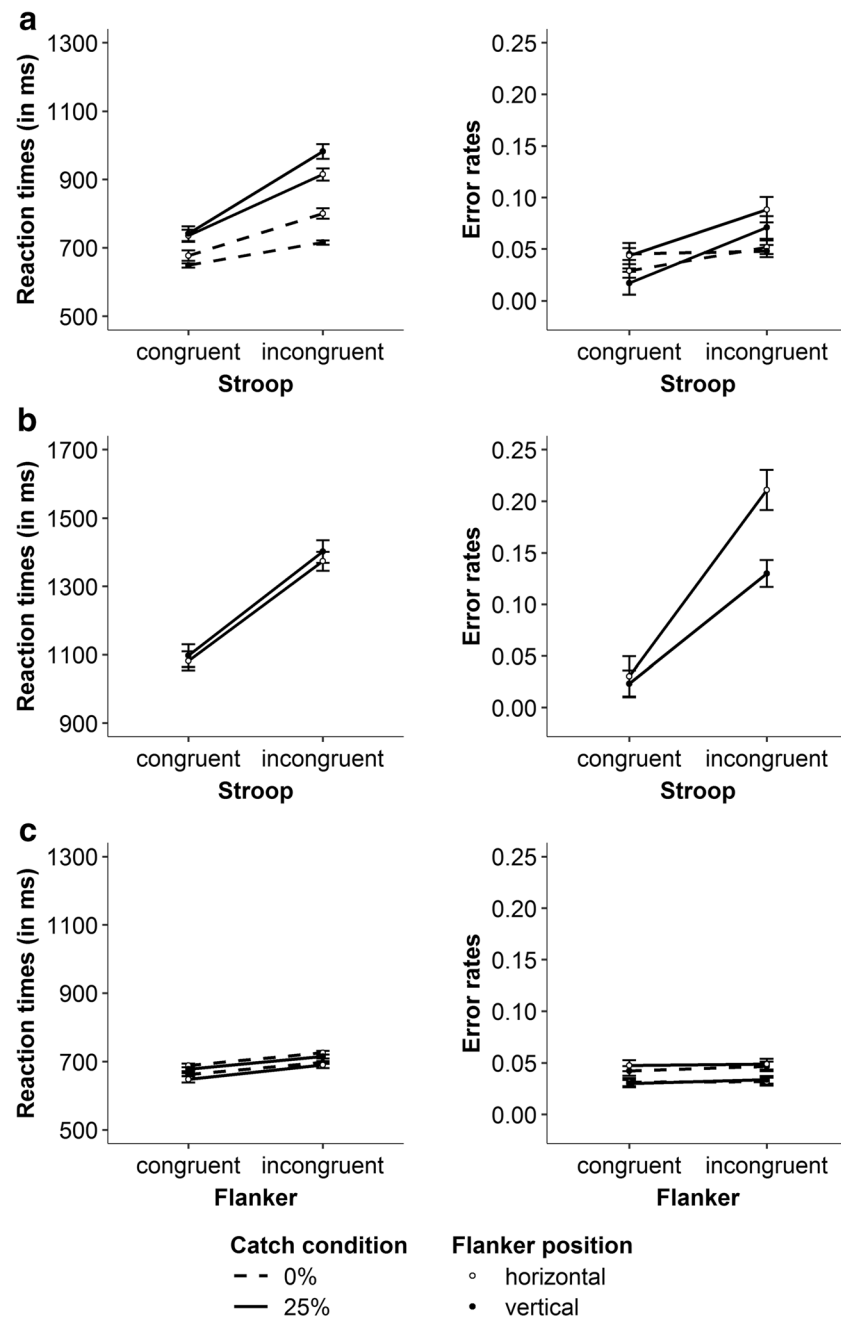
Block type	Reaction times					Arcsine transformed error rates				
	<i>t</i> test			Bayesian <i>t</i> test		<i>t</i> test			Bayesian <i>t</i> test	
	<i>t</i>	<i>p</i>	<i>d</i>	BF <sub>10</sub>	BF <sub>01</sub>	<i>t</i>	<i>p</i>	<i>d</i>	BF <sub>10</sub>	BF <sub>01</sub>
Pure Stroop blocks	17.95	<.001	3.74	$4.62 \times 10^{11}$	$2.16 \times 10^{-12}$	6.31	<.001	1.31	$8.01 \times 10^3$	$1.25 \times 10^{-4}$
Pure flanker blocks	5.32	<.001	1.11	962.58	$1.04 \times 10^{-3}$	0.62	.545	0.13	0.26	3.85

Note. *df* = 22 for the *t* statistic. Effect sizes are expressed as Cohen’s *d*. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)



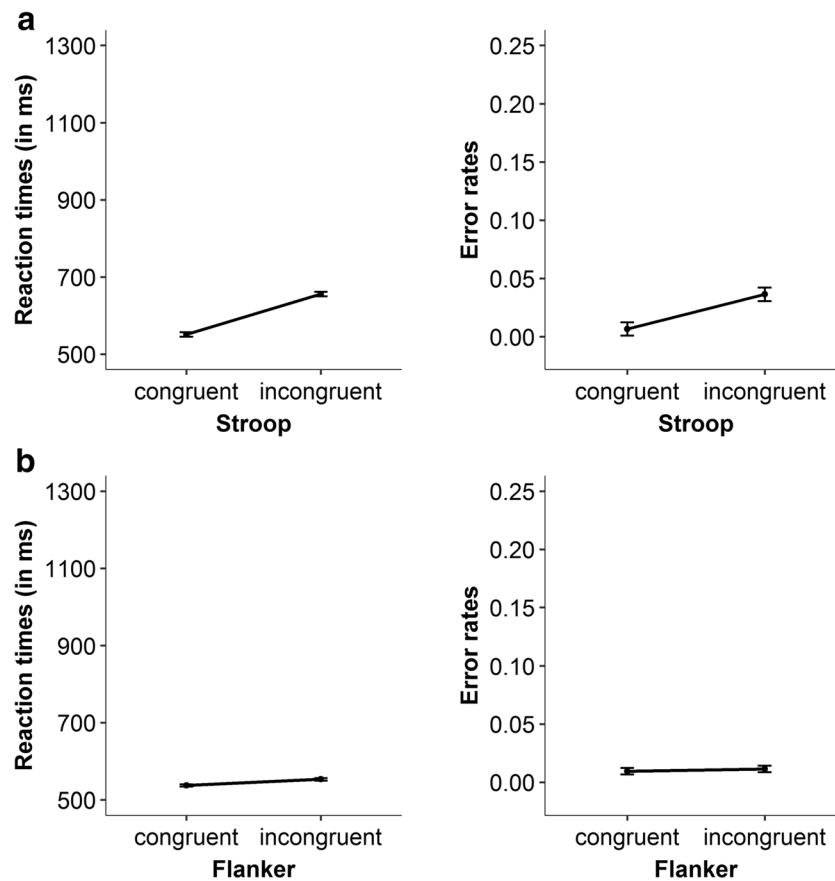
**Fig. 9** Experiment 1: Performance on the pure blocks. Mean reaction times (left part), mean log-transformed reaction times (middle part), and mean raw error rates (right part). Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008). **a** Stroop congruency: standard (no-catch) trials. **b** Stroop congruency: catch trials. **c** Flanker congruency





**Fig. 10** Experiment 2: Performance on the pure blocks. Mean reaction times (left part) and mean raw error rates (right part). Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008).

**a** Stroop congruency: standard (no-catch) trials. **b** Stroop congruency: catch trials. **c** Flanker congruency



**Fig. 11** Experiment 3: Performance on the pure blocks. Mean reaction times (left part) and mean raw error rates (right part). Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008). **a** Stroop congruency. **b** Flanker congruency

### Appendix C: Performance on catch trials in the mixed blocks

In this appendix, the focus was on performance on catch trials in the mixed blocks. For reaction times (RTs), data trimming removed errors, RTs faster than three standard deviations (*SD*) from the mean and RTs slower than three standard deviations from the mean for each congruency condition, session, and participant for Experiment 1 and for each congruency

condition and participant for Experiment 2. This removed 8.01% and 11.44% of the raw dataset for Experiment 1 and 2, respectively. An arcsine square root transformation was applied to the error rates for statistical analysis. Descriptive results for Experiments 1 and 2 are shown in Figures 12 and 13, respectively, and statistical results are presented in Tables 16 and 17.

**Table 16** Experiment 1: Inferential statistical values for the three-way analysis of variance (ANOVA) with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-

subject factors, and age group (young, older) as a between-subjects factor, and Bayes factors (BF) from the model comparisons

Dependent measure	Effect-Model	ANOVA			Bayesian analysis	
		<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Reaction times	Stroop congruency	451.46	<.001	.16	$1.99 \times 10^{43}$	$5.03 \times 10^{-44}$
	Flanker congruency	749.23	<.001	.13	$1.30 \times 10^{30}$	$7.69 \times 10^{-31}$
	Age group	136.14	<.001	.49	$3.45 \times 10^{18}$	$2.90 \times 10^{-19}$
	Stroop congruency × Flanker congruency	0.22	.637	<.001	0.15	6.49
	Stroop congruency × Age group	0.08	.774	<.001	0.14	7.17

**Table 16** (continued)

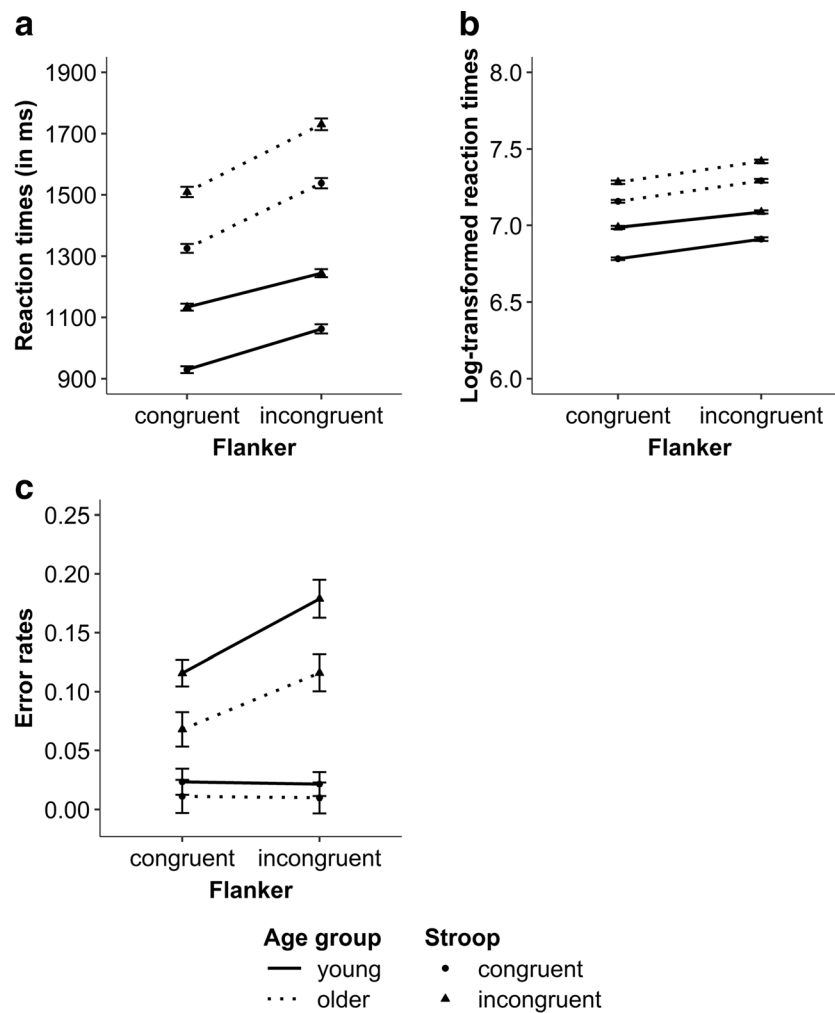
Dependent measure	Effect–Model	ANOVA			Bayesian analysis	
		<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Log-transformed reaction times	Flanker congruency × Age group	58.64	<.001	.01	147.93	6.76 × 10 <sup>-3</sup>
	Three-way interaction	1.39	.241	<.001	0.33	3.04
	Stroop congruency	511.40	<.001	.19	9.41 × 10 <sup>52</sup>	1.06 × 10 <sup>-53</sup>
	Flanker congruency	826.88	<.001	.13	8.32 × 10 <sup>26</sup>	1.20 × 10 <sup>-27</sup>
	Age group	160.45	<.001	.54	8.65 × 10 <sup>20</sup>	1.16 × 10 <sup>-21</sup>
	Stroop congruency × Flanker congruency	2.65	.106	<.001	0.30	3.38
	Stroop congruency × Age group	20.77	<.001	.010	118.78	8.42 × 10 <sup>-3</sup>
Arcsine transformed error rates	Flanker congruency × Age group	6.10	.015	.001	0.23	4.39
	Three-way interaction	3.10	.081	<.001	0.53	1.89
	Stroop congruency	255.62	<.001	.39	2.88 × 10 <sup>60</sup>	3.47 × 10 <sup>-61</sup>
	Flanker congruency	41.80	<.001	.03	8.44	0.12
	Age group	20.79	<.001	.08	413.28	2.42 × 10 <sup>-3</sup>
	Stroop congruency × Flanker congruency	42.41	<.001	.04	3.55 × 10 <sup>4</sup>	2.82 × 10 <sup>-5</sup>
	Stroop congruency × Age group	4.86	.029	.01	5.92	0.17
	Flanker congruency × Age group	0.01	.929	<.001	0.14	7.18
3-way interaction	2.46	.119	.002	0.40	2.49	

Note. *dfs* = (1, 125) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect)

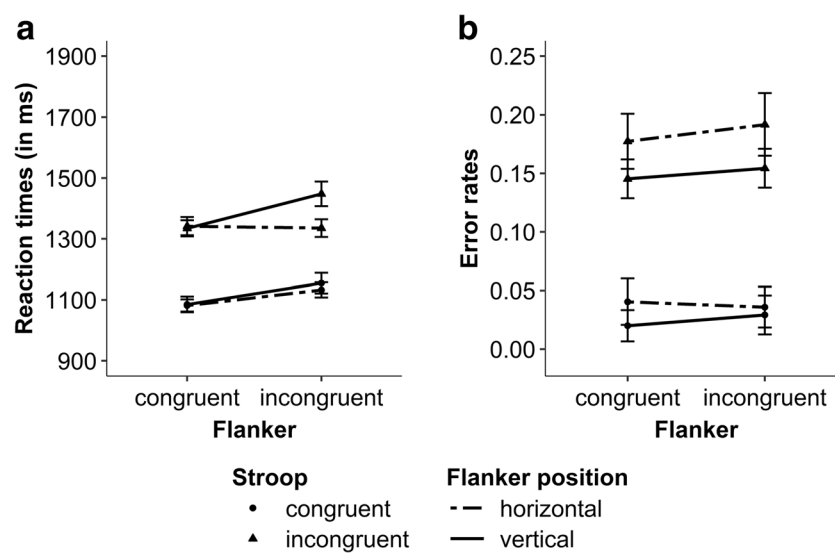
**Table 17** Experiment 2: Inferential statistical values for the three-way analysis of variance (ANOVA), with Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) as within-subject factors, and flanker position (horizontal, vertical) as a between-subjects factor, and Bayes factors (BF) from the model comparisons

Effect–Model	Reaction times					Arcsine transformed error rates				
	ANOVA			Bayesian analysis		ANOVA			Bayesian analysis	
	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>	<i>F</i>	<i>p</i>	$\eta_g^2$	BF <sub>10</sub>	BF <sub>01</sub>
Stroop congruency	198.80	<.001	.18	1.07 × 10 <sup>31</sup>	9.33 × 10 <sup>-32</sup>	279.40	<.001	.47	1.02 × 10 <sup>38</sup>	9.77 × 10 <sup>-39</sup>
Flanker congruency	23.40	<.001	.01	1.41	0.71	0.25	.622	<.001	0.16	6.24
Flanker position	0.18	.674	.004	0.41	2.44	1.20	.278	.02	0.35	2.87
Stroop congruency × Flanker congruency	0.08	.780	<.001	0.22	4.49	0.16	.692	<.001	0.23	4.39
Stroop congruency × Flanker position	1.26	.267	.001	0.46	2.18	0.01	.924	<.001	0.22	4.59
Flanker congruency × Flanker position	8.67	.005	.004	0.49	2.05	0.52	.476	.001	0.22	4.49
Three-way interaction	4.49	.040	.002	1.03	0.97	2.26	.140	.004	0.55	1.83

Note. *dfs* = (1, 45) for the *F* statistic. Effect sizes are expressed as generalized  $\eta^2$  values. BF<sub>10</sub> = Bayes factor in favor of the alternative hypothesis (i.e., in favor of the effect). BF<sub>01</sub> = Bayes factor in favor of the null hypothesis (i.e., in favor of the absence of the effect). As the ANOVA showed a significant three-way interaction for RTs and the Bayesian analysis suggested weak evidence for the model including this interaction, follow-up two-way repeated-measures ANOVAs were conducted, with the factors Stroop congruency (incongruent, congruent) and flanker congruency (incongruent, congruent) for each flanker position separately. Contrary to the omnibus ANOVA, these ANOVAs showed no significant interaction between Stroop congruency and flanker congruency, and the Bayesian analysis suggested positive evidence against the model including the interaction, horizontal flanker position: *F*(1, 23) = 2.95, *p* = .099,  $\eta_g^2$  = .003, BF<sub>10</sub> = 0.22, BF<sub>01</sub> = 4.53; and vertical flanker position: *F*(1, 22) = 1.65, *p* = .213,  $\eta_g^2$  = .002, BF<sub>10</sub> = 0.22, BF<sub>01</sub> = 4.45



**Fig. 12** Experiment 1: Performance on the catch trials in the mixed blocks. a Mean reaction times. b Mean log-transformed reaction times. c Mean raw error rates. Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008)

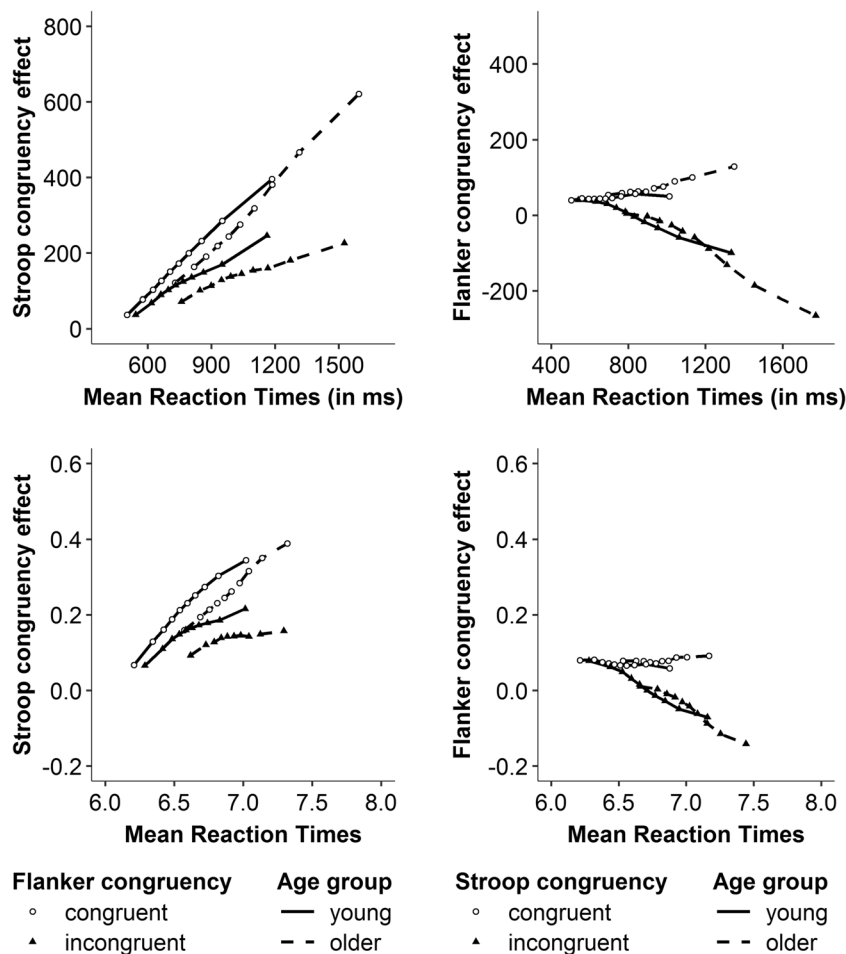


**Fig. 13** Experiment 2: Performance on the catch trials in the mixed blocks. a Mean reaction times. b Mean raw error rates. Error bars represent within-subject confidence intervals (see Cousineau, 2005; Morey, 2008)

### Appendix D: Delta plots

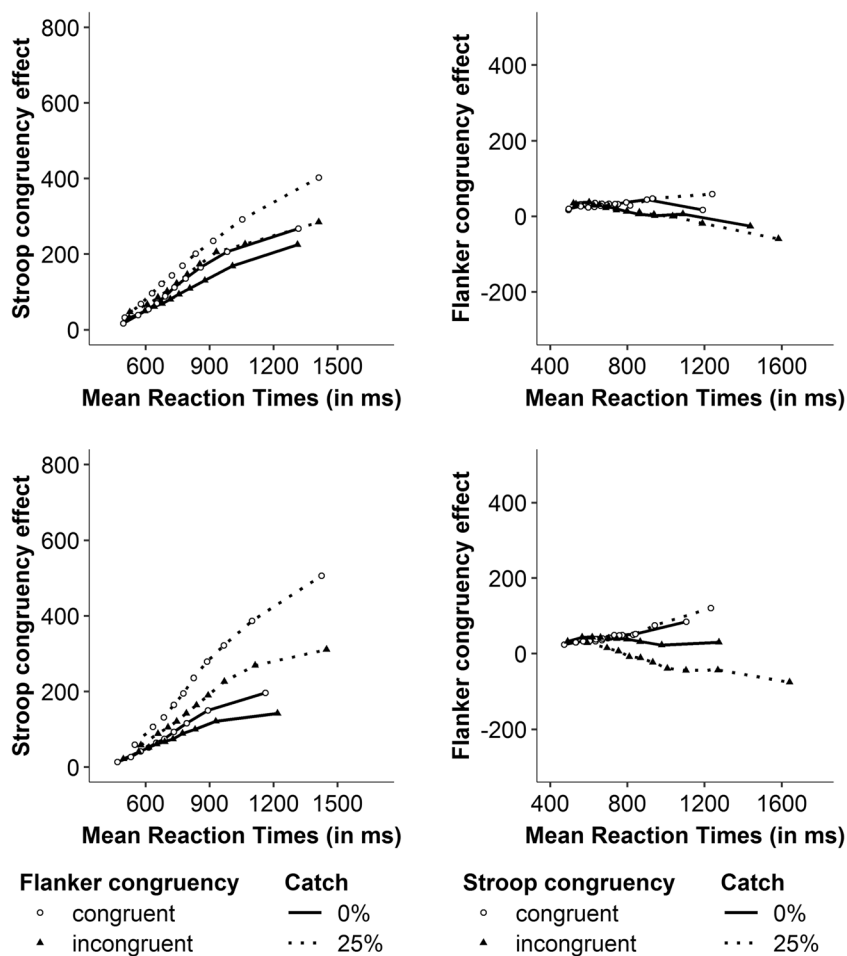
Delta plots illustrate how the magnitude of the congruency effect changes as a function of response speed. The *y*-axis of a delta plot shows the magnitude of the congruency effect at each percentile, whereas its *x*-axis shows the RT performance averaged across incongruent and congruent trials (see Pratte et al., 2010, for an example). A first delta plot was computed

for the Stroop congruency effect depending on whether the flanker trials were incongruent or congruent (left part in each figure of Appendix D). For the sake of completeness, a second delta plot was computed for the flanker congruency effect depending on whether the Stroop trials were incongruent or congruent (right part in each figure of Appendix D). Delta plots for Experiments 1, 2, and 3 are shown in Figs. 14, 15, and 16, respectively.

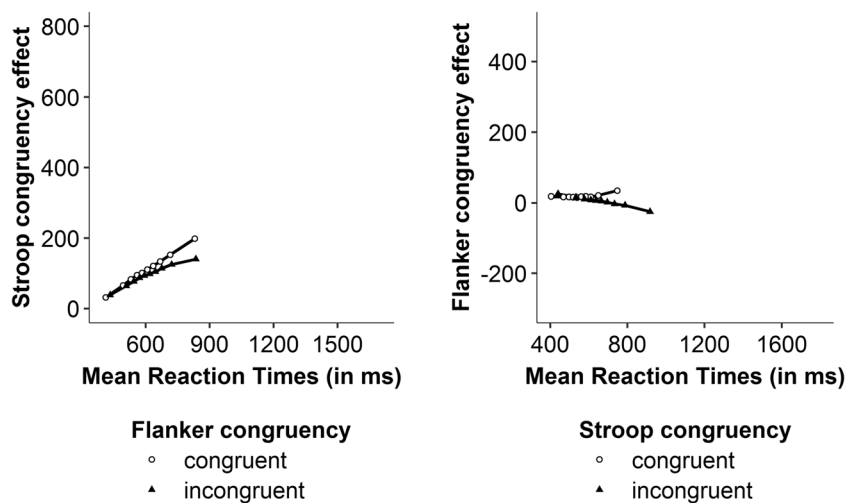


**Fig. 14** Experiment 1. Delta plots for the Stroop congruency effect (left part) and the flanker congruency effect (right part). Upper part: Reaction times given in milliseconds (ms). Lower part: Log-transformed reaction times





**Fig. 15** Experiment 2. Delta plots for the Stroop congruency effect (left part) and the flanker congruency effect (right part). Upper part: Horizontal flanker position. Lower part: Vertical flanker position. Reaction times are given in milliseconds (ms)



**Fig. 16** Experiment 3. Delta plots for the Stroop congruency effect (left part) and the flanker congruency effect (right part). Reaction times are given in milliseconds (ms)

## References

- Akçay, Ç., & Hazeltine, E. (2011). Domain-specific conflict adaptation without feature repetitions. *Psychonomic Bulletin & Review*, *18*, 505–511. doi:<https://doi.org/10.3758/s13423-011-0084-y>
- Besner, D., Stolz, J. A., & Boutilier, C. (1997). The Stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*, 221–225. doi: <https://doi.org/10.3758/BF03209396>
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652. doi:<https://doi.org/10.1037/0033-295X.108.3.624>
- Boy, F., Husain, M., & Sumner, P. (2010). Unconscious inhibition separates two forms of cognitive control. *Proceedings of the National Academy of Sciences*, *107*, 11134–11139. doi:<https://doi.org/10.1073/pnas.1001925107>
- Buchanan, E., Scofield, J., & Valentine, K. D. (2019). MOTE: Effect Size and Confidence Interval Calculator (Version 1.0.1) [R package].
- Bugg, J. M., DeLosh, E. L., Davalos, D. B., & Davis, H. P. (2007). Age differences in Stroop interference: Contributions of general slowing and task-specific deficits. *Aging, Neuropsychology, and Cognition*, *14*, 155–167. doi:<https://doi.org/10.1080/138255891007065>
- Bullinger, M. (1998). SF-36. Fragebogen zum Gesundheitszustand SF-36: Health Questionnaire (I. Kirchberger, Ed.). Göttingen, Germany: Hogrefe.
- Cerella, J., & Hale, S. (1994). The rise and fall in information-processing rates over the life span. *Acta Psychologica*, *86*, 109–197. doi:[https://doi.org/10.1016/0001-6918\(94\)90002-7](https://doi.org/10.1016/0001-6918(94)90002-7)
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.
- Egner, T. (2008). Multiple conflict-driven control mechanisms in the human brain. *Trends in Cognitive Sciences*, *12*, 374–380. doi: <https://doi.org/10.1016/j.tics.2008.07.001>
- Egner, T., Delano, M., & Hirsch, J. (2007). Separate conflict-specific cognitive control mechanisms in the human brain. *NeuroImage*, *35*, 940–948. doi:<https://doi.org/10.1016/j.neuroimage.2006.11.061>
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143–149. doi:<https://doi.org/10.3758/BF03203267>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. doi:<https://doi.org/10.3758/BF03193146>
- Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*, 189–198.
- Frühholz, S., Godde, B., Finke, M., & Herrmann, M. (2011). Spatio-temporal brain dynamics in a combined stimulus–stimulus and stimulus–response conflict task. *NeuroImage*, *54*, 622–634. doi: <https://doi.org/10.1016/j.neuroimage.2010.07.071>
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 875–894. doi:<https://doi.org/10.1037/0096-1523.8.6.875>
- Hasher, L., Zacks, R., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 653–675). Cambridge, MA: MIT Press.
- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.
- Hautzinger, M., Keller, F., & Kühner, C. (2006). Beck Depression Inventar II (BDI 2) [Beck Depression Inventory II]. Frankfurt, Germany: Harcourt Test Service.
- Hommel, B. (1997). Interactions between stimulus–stimulus congruence and stimulus–response compatibility. *Psychological Research*, *59*, 248–260. doi:<https://doi.org/10.1007/BF00439302>
- Kalanthroff, E., & Henik, A. (2014). Preparation time modulates proactive control and enhances task conflict in task switching. *Psychological Research*, *78*, 276–288. doi:<https://doi.org/10.1007/s00426-013-0495-7>
- Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*, 1017–1048. doi:<https://doi.org/10.1037/xge0000184>
- Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of Stroop- and Simon-like stimuli. *Psychological Research*, *56*, 130–135. doi:<https://doi.org/10.1007/BF00419699>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203. doi: <https://doi.org/10.1037/0033-2909.109.2.163>
- Mayr, S., Erdfelder, E., Buchner, A., & Faul, F. (2007). A short tutorial of G\*Power. *Tutorials in Quantitative Methods for Psychology*, *3*, 51–59.
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*, 8–14. doi: <https://doi.org/10.1177/0963721411429458>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*, 61–64.
- Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.12-2) [R package]. Retrieved from <https://CRAN.R-project.org/package=BayesFactor>
- Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*, 475–487.
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, *8*, 434–447. doi:<https://doi.org/10.1037/1082-989X.8.4.434>
- Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between Stroop and Simon effects using delta plots. *Attention, Perception, & Psychophysics*, *72*, 2013–2025. doi:<https://doi.org/10.3758/APP.72.7.2013>
- Prince, M., Brown, S., & Heathcote, A. (2012). The design and analysis of state-trace experiments. *Psychological Methods*, *17*, 78–99. doi: <https://doi.org/10.1037/a0025809>
- R Core Team. (2019). R: A language and environment for statistical computing (Version 3.5.2) [Computer software]. Retrieved from <https://www.R-project.org/>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. doi:<https://doi.org/10.2307/271063>
- Rey-Mermet, A., & Gade, M. (2016). Contextual within-trial adaptation of cognitive control: Evidence from the combination of conflict tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1505–1532. doi:<https://doi.org/10.1037/xhp0000229>
- Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*, 1695–1716. doi:<https://doi.org/10.3758/s13423-017-1384-7>

- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 501–526. doi:<https://doi.org/10.1037/xlm0000450>
- Rey-Mermet, A., Gade, M., Souza, A. S., von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *48*(8), 1335–1372. doi: <https://doi.org/10.1037/xge0000593>
- Rey-Mermet, A., Gade, M., & Steinhauser, M. (2019). Sequential conflict resolution under multiple concurrent conflicts: An ERP study. *NeuroImage*, *188*, 411–418. doi: <https://doi.org/10.1016/j.neuroimage.2018.12.031>
- Rohatgi, A. (2018). WebPlotDigitizer [Computer software] (Version 4.1). Retrieved from <https://automeris.io/WebPlotDigitizer/index.html>
- Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*, 304–321. doi:<https://doi.org/10.1037/met0000057>
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776. doi:<https://doi.org/10.1037/0012-1649.27.5.763>
- Satzger, W., Hampel, H., Padberg, F., Bürger, K., Nolde, T., Ingrassia, G., & Engel, R. R. (2001). Zur praktischen Anwendung der CERAD-Testbatterie als neuropsychologisches Demenzscreening [The practical application of the CERAD test battery as neuropsychological dementia screening]. *Der Nervenarzt*, *72*, 196–203. doi:<https://doi.org/10.1007/s001150050739>
- Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*, 1033–1040. doi:<https://doi.org/10.3758/BF03201181>
- Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, *189*, 93–102. <https://doi.org/10.1016/j.actpsy.2017.10.007>
- Simon, J. R., & Berbaum, K. (1990). Effect of conflicting cues on information processing: The ‘Stroop effect’ vs. the ‘Simon effect.’ *Acta Psychologica*, *73*, 159–170. doi:[https://doi.org/10.1016/0001-6918\(90\)90077-S](https://doi.org/10.1016/0001-6918(90)90077-S)
- Simon, J. R., & Small, A. M. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, *53*, 433–435. doi:<https://doi.org/10.1037/h0028034>
- Singmann, H., Bolker, B., & Westfall, J. (2018). afex: Analysis of factorial experiments (Version 0.19-1) [R package]. Retrieved from <https://CRAN.R-project.org/package=afex>
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer’s type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 461–479. doi:<https://doi.org/10.1037/0096-1523.22.2.461>
- Stahl, C., Voss, A., Schmitz, F., Nuszbaum, M., Tüscher, O., Lieb, K., & Klauer, K. C. (2014). Behavioral components of impulsivity. *Journal of Experimental Psychology: General*, *143*, 850–886. doi: <https://doi.org/10.1037/a0033981>
- Steinhauser, M., & Hübner, R. (2009). Distinguishing response conflict and task conflict in the Stroop task: Evidence from ex-Gaussian distribution analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 1398–1412. doi:<https://doi.org/10.1037/a0016467>
- Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods*, *38*, 280–286. doi:<https://doi.org/10.3758/BF03192779>
- Stoffels, E. J., & van der Molen, M. W. (1988). Effects of visual and auditory noise on visual choice reaction time in a continuous-flow paradigm. *Perception & Psychophysics*, *44*, 7–14. doi:<https://doi.org/10.3758/BF03207468>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:<https://doi.org/10.1037/h0054651>
- Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition*, *20*, 727–735. doi:<https://doi.org/10.3758/BF03202722>
- Verhaeghen, P. (2014). Age-related differences in the speed of executive control. In *The elements of cognitive aging: Meta-analyses of age-related differences in processing speed and their consequences* (pp. 169–210). Oxford, England: Oxford University Press.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*, 779–804. doi: <https://doi.org/10.3758/BF03194105>
- Weissman, D. H. (2019). Interacting congruency effects in the hybrid Stroop–Simon task prevent conclusions regarding the domain specificity or generality of the congruency sequence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi: <https://doi.org/10.1037/xlm0000769>
- Wendt, M., Kluwe, R. H., & Peters, A. (2006). Sequential modulations of interference evoked by processing task-irrelevant stimulus features. *Journal of Experimental Psychology: Human Perception & Performance*, *32*, 644–667. doi:<https://doi.org/10.1037/0096-1523.32.3.644>
- White, C. N., Servant, M., & Logan, G. D. (2018). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin & Review*, *25*, 286–301. doi:<https://doi.org/10.3758/s13423-017-1271-2>
- Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*, 37–49.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.