https://dx.doi.org/10.1016/j.bandl.2021.104952

Can inhibition deficit hypothesis account for age-related differences in semantic fluency? Converging evidence from Stroop color and word test and an ERP flanker task

Manson Cheuk-Man Fong[a,*], Tammy Sheung-Ting Law[a], Matthew King-Hang Ma[b], Nga Yan Hui[a], William Shiyuan Wang[a,b,*]

[a] *Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong*
[b] *Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong*

Abstract

The inhibition deficit hypothesis (IDH) proposed that individual differences in inhibitory control is an underlying reason for age-related language decline. This study examined whether the hypothesis holds within the domain of lexico-semantic retrieval. Sixty-six older adults aged 60-79 were tested in a semantic fluency task comprising 16 categories; each response was classified as automatic or controlled. Also, Stroop color and word test and an ERP flanker task were employed to yield both behavioral and neural measures of inhibitory control. Mixed-effects modelling revealed that the number of controlled (but not automatic) responses was negatively associated with age. This interaction could be partially accounted for by the behavioral Stroop inhibition score and two neural measures from the ERP flanker task (P2 and Pc amplitudes). These results not only provide converging evidence supporting the IDH, but also demonstrate the involvement of specific inhibitory control components, including attentional control and performance monitoring.

1. Introduction

1.1. *Controlled retrieval and inhibition deficit hypothesis*

According to the controlled semantic cognition (CSC) theory, the semantic system is partitioned into two components, a representational component for storing lexico-semantic information and a control component to enable the strategic retrieval of information (Lambon Ralph et al., 2017). Recent research has suggested that although older adults are often similar to or even better than young adults in accuracy for tasks that require primarily automatic activation of lexico-semantic information, they perform more poorly when there is a greater demand for controlled processing (Baciu et al., 2016). For example, in receptive tests of semantic knowledge (e.g., lexical decision), older adults were reported to attain similar or even higher accuracy (Baciu et al., 2016). However, when asked to judge, among a pair of words (e.g., “dove” and “pepper”), the one shared the same color with a probe word “salt”, older adults were significantly less accurate and disproportionately slower than younger adults in the presence of the semantic competitor “pepper” (Hoffman, 2018).

The *inhibition deficit hypothesis* (IDH; Hasher & Zacks, 1988; Hasher et al., 1991) represents a plausible and deeper account for the differential effect of ageing in semantic retrieval highlighted above. According to the IDH, inhibitory control is instrumental to many of our cognitive functions, because it can prevent attentional resources to be misallocated to task-irrelevant information. As a result, it is hypothesized that differences in inhibitory control can account for

[*] Corresponding authors
Email addresses: cmmfong@polyu.edu.hk (Manson Cheuk-Man Fong), wsywang@polyu.edu.hk (William Shiyuan Wang)

the age-related differences in various cognitive domains, including language. The IDH has found support in language comprehension, episodic memory tasks (Hasher & Zacks, 1988; Hasher et al., 1991), and fragmented picture identification task (Lindfield et al., 1994). In these studies, the task performance of younger and older adults were compared, and the poorer performance in older adults was attributed to their declining inhibition function. However, these studies did not necessarily include an inhibition task to formally establish the association. It is only in some later studies that the predictions of the IDH were rigorously tested with partial correlation analysis or modelling. For example, inhibition was shown to directly account for the age-related performance deficit in both a verbal list learning task and an attention task (Persad et al., 2002), meaning that the predictive power of age on task performance was significantly weakened once inhibition was taken into account.

The question can be asked whether the IDH can account for age-related differences in language production tasks such as verbal fluency—a routine neuropsychological screening instrument that evaluates the access to and controlled retrieval of lexico-semantic knowledge. There are two common forms of verbal fluency tasks, where the task is to produce as many items belonging to a semantic category (i.e., semantic fluency) or beginning with a certain letter (i.e., phonemic fluency). Various authors have remarked that inhibitory control is required in verbal fluency. Specifically, inhibitory control is required for implementing a strategy for controlled semantic retrieval, detecting semantic conflicts between retrieved concepts, controlling for undesired concepts, and ensuring the quality of the retrieved information (e.g., Rosen & Engle, 1997; Sauzéon et al., 2011; Unsworth, Spillers & Brewer, 2010). Inhibitory control is also required for retrieval-related search and post-retrieval control processes for organizing words into clusters (Rosen & Engle, 1997). In alignment with the IDH, it is plausible that the decline in inhibitory control functions directly contributed to the decline in verbal fluency performance. Consistent with this view, semantic fluency, which involves greater executive control than phonemic fluency, has consistently been observed in older adults (Gordon et al., 2018; Troyer et al., 1997; Vonk et al., 2019).

However, the evidence that the IDH may extend to the verbal fluency task has remained scarce. First, although some studies reported significant associations between verbal fluency and inhibition measures derived from Stroop and flanker tasks, age was not partialled out in reporting the correlations as this was not their research focus (McDowd et al., 2011; Unsworth et al., 2010). Thus, no inference can be drawn on whether the age-related differences in verbal fluency is directly due to the individual variations in inhibitory control. The study by Shao et al. (2014) presented a rare exception in that age was factored into the regression analysis, but no significant association between stop-signal RT and verbal fluency measures was found. The authors suggested that the reason is that the stop-signal task measured only motoric inhibition, yet verbal fluency might rely more on the ability to suppress the activation of highly potent lexical competitors.

Second, inhibitory control is not a unitary concept; instead, it is supported by several component processes. According to the conflict monitoring theory, the dorsolateral prefrontal cortex (DLPFC) is hypothesized to be functionally responsible for regulating the amount of attentional resources devoted in stimulus evaluation, whereas the dorsal anterior cingulate cortex (dACC) plays a key role to detect conflicts between competing stimulus features and/or response options (Botvinick et al., 2001; Egner, 2008). The VLPFC and the anterior insula have also been linked to response inhibition and selection (Aron et al., 2014). These advances in the conflict monitoring theory and related research have suggested a complex, but exciting picture on the component processes. Due to this non-unitary nature of inhibitory control, it is pertinent to employ experimental techniques that allow the component processes of inhibitory control to be readily measured. One means for this purpose is the event-related potential (ERP), which allows the component processes to be temporally segregated. In the following, we briefly review the relevant ERP literature.

1.2. ERP correlates of inhibitory control

In the ERP literature, the N2 component—a frontocentral component that peaks about 200–350 ms post-stimulus—is by far the most well-known index of inhibition (Folstein & Van Petten, 2008), being associated with various related functions such as conflict detection (Botvinick et al., 2001) and error detection (Yeung et al., 2004). In young adults, the N2 amplitude is often larger (i.e., more negative) in the presence of response conflict (Gajewski & Falkenstein, 2013). For example, in flanker tasks, the N2 amplitude is larger for incompatible than compatible trials (Kopp et al., 1996). In older adults, N2 amplitude was found to be significantly less negative (Donohue et al., 2016).

Besides the N2 component, other components have also been reported to play a role in conflict processing, including P2 and P3b components (Korsch et al., 2016; Zhou et al., 2019). The P2 component peaks around 200 ms, and it has been linked with selective attention and analysis of visual features (Hillyard & Münte, 1984; Luck & Hillyard, 1994). For example, in oddball detection tasks, P2 amplitude was found to be larger for targets than non-targets regardless of whether an overt response was required (Potts, 2004). Also, Kopp et al. (2007) showed that the P2 amplitude elicited by a visual non-target was modulated by its perceptual distinctiveness to the target. These findings suggest that P2 amplitude is modulated by attentional control.

The P3b component, a centroparietal component that peaks about 400 ms post-stimulus, is modulated by task difficulty, and reflects the amount of attentional resources allocated (Polich, 2007). Consistent with this general view, the P3b amplitude is smaller for incompatible than compatible condition in flanker paradigms (Zhou et al., 2019), and it may reflect the resources allocated during response selection (Verleger et al., 2005). In normal ageing, the P3b amplitude decreases in the flanker task (Wild-Wall et al., 2008), indicating that response selection function may decline with age.

Apart from stimulus-locked components, some response-locked components are informative about conflict processing (Falkenstein et al., 2000; Ullsperger, Danielmeier & Jocham, 2014). For example, the error-related negativity (ERN; also known as Ne) is a negative-going component elicited about 100 ms post-error (Gehring et al., 1993), and it is followed by a positive deflection named the error positivity (Pe) that peaks about 200-400 ms post-error. Functionally, the ERN and Pe have been proposed to reflect error monitoring processes, with the ERN being associated with early error detection and the Pe with conscious error processing (Baldwin et al., 2015). In correct trials, two analogous components have been observed (Maier, Di Pellegrino & Steinhauser, 2012)—correct-related negativity (CRN) and correct positivity (Pc). While their precise functional significance have remained understudied, they likely reflect some aspects of performance monitoring (Ullsperger et al., 2014). Typically, the ERN is larger (more negative) for incompatible than compatible trials, and larger than the CRN. However, older adults would show smaller ERN but larger CRN than younger adults, indicating reduced awareness to own errors and increased uncertainty for correct answers (Schreiber et al., 2011).

1.3. The present study—the role of inhibitory control in lexico-semantic retrieval

The present study aimed to provide both behavioral and neural evidence for testing whether the IDH holds within the domain of lexico-semantic retrieval. Towards this end, a semantic fluency task was employed[1]. Sixteen categories were included such that representative lexico-semantic retrieval performance could be obtained; these categories have previously been tested in younger adults (Fong et al., 2020).

[1] Phonemic fluency was not run because of the non-alphabetic nature of Chinese orthographies.

Two conflict tasks were employed to provide both behavioral and neural measures of inhibitory control. The behavioral, Stroop color and word test (Golden & Freshwater, 1978) was used to evaluate the overall inhibitory control performance of the participants. However, behavioral scores of inhibition are composite measures resulting from the component processes of inhibitory control. For this reason, an ERP flanker task was also used to study the possible roles of these component processes in the age-related differences in semantic fluency. In contrast to the original Eriksen flanker task that used letters to construct the stimulus array, arrows were used to minimize the involvement of verbal processing, so that the domain-general inhibitory control processes would be evoked. Our task design was inspired by the taxonomy of conflicts proposed by (Kornblum et al., 1990), which classifies a conflict either as stimulus–response (S–R) or stimulus–stimulus (S–S) conflicts. S–R conflicts are induced when the target stimulus primes a response that is incompatible with the response required, like in the Simon task, while S–S conflicts occur when the task-irrelevant information is incompatible with the target, as in the flanker task. Previous works combined Simon and flanker tasks (Frühholz et al., 2011; Korsch et al., 2016), and demonstrated that the S–R and S–S conflicts are resolved by different brain mechanisms. Like these studies, the present study also manipulated two factors (Fig. 1): Response (response-compatible, RC, vs. response-incompatible, RI) and Flanker (flanker-compatible, FC, vs. flanker-incompatible, FI), so that our inhibitory control measures are not conflict-specific. Because the conflict in RI conditions arose between the target arrow (up or down) and the response required (left or right), while that in the FI conditions was due to the stimulus-level conflict between the target arrow and the flanker, the RI and FI conditions could be readily understood as a type of S–R and S–S conflict, respectively.

Following a similar logic to the study by Persad et al. (2002), we focused on the question of whether the age-related differences in semantic fluency could be accounted for by the behavioral and ERP measures of inhibitory control. For much of the literature, semantic fluency has been scored in terms of the number of concepts produced, without taking into consideration the time required to retrieve each concept. In light of the CSC theory, concepts that are rapidly retrieved likely evoke less semantic control, which in turn suggests less involvement of inhibitory control. Consistent with this view, a recent study also showed that semantic fluency relies on both automatic and controlled retrieval (Gordon et al., 2018). To incorporate retrieval mode in our analyses, each correct response was classified according to inter-response time (i.e., the response latency relative to the preceding response). For each category, two main performance measures were thus obtained per participant, namely, number of concepts retrieved with automatic or with controlled processes (N_A or N_C).

Generalized linear mixed-effects modelling (GLMM; Baayen et al., 2008; Faraway, 2016) was used for hypothesis testing. Specifically, we first constructed a baseline Demographics-model, in which the interaction between age and retrieval mode (Mode: automatic/controlled) on semantic fluency were included as predictors. We then constructed other models and tested whether including one or more behavioral/ERP predictors of inhibitory control and their respective interactions with Mode in these models would significantly weaken the effect of age and its interaction with mode. If so, the age-related differences in semantic fluency could be explained by the individual differences in inhibitory control. While we examined the three stimulus-locked ERP components reviewed above, we focused only on one response-locked component (Pc) because there was an insufficient number of error trials for examining ERN, and that the CRN was not discernible in the correct trials (see section 3.3.2).

Beyond testing the IDH, we also sought to shed light on the mechanism behind the potential relationship between inhibitory control and age-related differences in semantic fluency. The functional associations of the ERP components with semantic fluency can be manifested in many ways. For example, in accordance with the conflict monitoring theory, individual N2 compatibility effect may predict the semantic fluency performance because the conflict detection mechanism would be implicated in detecting the semantic conflicts between the retrieved concepts and the semantic category in question. Here, however, the present paper was especially interested to determine whether some inhibitory control processes are differentially recruited depending on the automaticity of semantic retrieval. For example, if the attentional control process is differentially required for lexico-semantic selection, then older adults with good attentional control would be especially capable of retrieving concepts in a controlled manner. If so, a significant association between individual P2 amplitude and controlled semantic retrieval performance may be observed. The GLMMs obtained were used to explore the exact patterns of associations.

2. Methods

2.1. Participants

Participants were 66 cognitively normal older adults (28 male) aged 60-79 ($M = 67.4$, $SD = 4.7$), with 12.9 years of education ($SD = 4.9$). The majority of them were right-handed except that four were ambidextrous according to the Edinburgh inventory (Oldfield, 1971). All participants were native Cantonese speakers, had no known neurological disorders and normal/correct-to-normal vision. All procedures were approved by the Ethical Review Committee, Hong Kong Polytechnic University. Written signed informed consent was obtained from all participants, who were paid HKD 200 for their participation.

2.2. Experimental design and data acquisition

Participants attended two sessions each, in which behavioral and EEG tasks were administered. The two sessions were separated by an average of 3.7 days ($SD = 1.7$). The critical tasks in the present report—two behavioral tasks (Stroop color and word test and semantic fluency) and the ERP flanker task—were administered in the first and second sessions, respectively.

The Stroop color and word test was adapted from a standard test in English (Golden & Freshwater, 1978) into Chinese, and comprised three sub-tasks: word naming, color naming, and Stroop color–word naming. The details of the adaptation could be found in our previous works (Fong et al., 2020; Hui et al., 2020).

In the semantic fluency task, sixteen semantic categories (e.g., mammals and countries) covering a broad range of common concepts encountered in daily life were tested as separate questions in order to yield a representative measure of semantic retrieval. All the categories and procedures were identical to those in Fong et al. (2020). See Table S1, for the full list of categories.

In the flanker task, each stimulus array comprised a centrally presented target arrow (< or >) flanked by arrows on each side, e.g., >><>> (Fig. 1). A two-by-two design was used, with Response (RC/RI) and Flanker (FC/FI) being orthogonally manipulated to yield four conditions. Participants were instructed to ignore the flankers, and respond to the target arrow at the central fixation point only; they should press the left button if the target symbol was < or $\wedge$, and the right button if the target symbol was > or $\vee$. In the RC conditions, the target symbol matched the response required (e.g., <<<<< and <<><<), but for the RI condition, the target symbol (e.g., $\wedge\wedge\wedge\wedge\wedge$ and $\wedge\wedge\vee\wedge\wedge$) and the response required (left and right, respectively) did not match. A balanced design was adopted, meaning that stimuli for the four conditions were presented with equal probability. Participants were given 2000 ms to respond as quickly and accurately as possible; there were 384 trials in total. During the task, electroencephalograms were acquired at 2048 Hz with a 32-channel ActiveTwo EEG system, using Ag/AgCl electrodes. Two electrodes were placed over the two mastoids for offline re-referencing. Horizontal/vertical electroocculograms (HEOG/VEOG) were recorded using four electrodes placed near the two outer canthi, and above/below left eye.

2.3. Data analysis

2.3.1. Stroop color and word test

A single performance measure was obtained: StroopInhibition. It was defined as the number of Stroop color-words named within 45 s, divided by the average number of words and colors separately named in 45 s. This formula ensures that the processing speed of individual participants was factored out (Belleville et al., 2006).

2.3.2. Semantic fluency

The data transcription procedure largely followed our previous work (Fong et al., 2020). In brief, individual responses produced by each participants were first transcribed faithfully word-by-word. The complete list of responses from all participants, with all duplicates removed, was then compiled per category. Intrusions were identified by three independent judges. After obtaining the list of correct responses, correct responses that were synonymous from one another were grouped under the same concept if at least two of the three judges considered the responses to be synonymous. For example, in Cantonese, “monkey” can be referred to in colloquial (/maa5 lau1/) or literary form (/hau4 zi2/).

For each category, the total semantic fluency was calculated as the number of correct responses produced (repeated occurrence(s) of a concept were considered as perseveration errors and were not counted towards this score). In addition, in order to test our main hypothesis that different components of inhibitory control are differentially involved depending on the retrieval mode (automatic/controlled), the onset latency of each response was recorded manually up to a precision of 0.1 s, such that each response can be classified as being automatic or controlled according to the time required for producing the response. To take into account the different speaking habits of different individuals, we defined the threshold appropriate for each individual according to the probability distribution of inter-response times t, pooled across categories (note: for the first response in each question, t was taken as its onset latency; for the remaining responses, t was measured simply as the onset latency with respect to the onset of the previous response). An exponential distribution ($\lambda e^{-\lambda t}$) was fitted to describe the distribution of the inter-response times, which specifies that the probability for a response to be produced within time t is proportional to t. A response was considered as automatic if the time required to produce it was less than the half-life of the fitted distribution, i.e., $-log(0.5)/\lambda$; this threshold took into account the different spread of individual distributions. For each category, the number of automatic and controlled responses (N_A/N_C) were thus obtained.

2.3.3. Flanker task

In analyzing the behavioral data, incorrect and outlying responses faster than 100 ms, slower than 2000 ms, or with a reaction time (RT) beyond 3 SDs from the individual mean, were excluded in the RT analysis. Repeated-measures ANOVAs were run to test the effects of Response and Flanker on RT and error rate.

The procedure for EEG preprocessing closely followed our previous study (Fong et al., 2018). Raw EEG data were filtered between 0.1–30 Hz and downsampled to 512 Hz using EEGLAB (Delorme & Makeig, 2004). Stimulus-locked epochs from 200 ms pre-stimulus to 1500 ms post-stimulus were obtained. The data were then re-referenced to the average mastoid. Baseline correction was performed based on the 200 ms pre-stimulus. Epochs with voltage fluctuation of over 120 μV in any of the scalp electrodes or the two reference electrodes, eye movements (threshold: 120 μV), or blinks (threshold: 200 μV) between −200 ms and 400 ms were excluded. Eye movements and blinks were removed based on independent component analysis (ICA), computed using the `runica` function. The stimulus-locked ERP for each condition was obtained by averaging the artefact-free epochs. Mean amplitude for P2, N2, and P3b, were computed at the following time-windows (electrodes): 130-230 ms (Fz and Cz), 240-360 ms (Fz and Cz), and 400-700 ms (Cz and Pz). To obtain the response-locked ERP, the stimulus-locked epochs were re-time-locked to the RT; a baseline of −200 ms to −100 ms pre-response was employed. The correct positivity (Pc) was computed at 80–280 ms (Fz and Cz) based on visual inspection.

However, the mean amplitudes in each condition thus obtained were highly correlated, making them inappropriate for inclusion as predictors in the GLMM analyses. In order to derive a set of uncorrelated indices that capture the individual variations in inhibitory control, principal component analysis (PCA) was applied to the condition amplitudes. Specifically, for each ERP component, we computed the mean amplitude across the two electrodes selected per condition; PCA was then applied to the resultant measures across the four conditions to yield four principal component scores (PC1-4).

2.3.4. Evaluation of the IDH with generalized linear mixed effects modelling

GLMM was adopted due to its strength in simultaneously incorporating the variability across participants and categories as crossed random effects (Baayen et al., 2008). To examine our two main research issues, a series of six GLMMs were constructed. In Demographics-model, we started with two random-effect terms (participant and category) and six fixed-effect terms: Age, Gender, Education, and their interactions with retrieval mode (Mode: automatic/controlled). A stepwise backward selection procedure based on the Bayesian Information Criterion (BIC) was applied to determine a parsimonious model for comparison with the remaining, more complex models. Because a small BIC indicates a more parsimonious model, at each step, the term that leads to the largest increase in BIC when excluded would be excluded if the increase was larger than 2. If all the increases in BIC were smaller than 2, all the terms were kept to yield the final model.

In the remaining GLMMs, additional predictors and their interactions with Mode were added to the final Demographics-model. These predictors included: StroopInhibition and the PC1-4 for each of the ERP components (e.g., P2), each yielding a corresponding model (Stroop-model, P2-model, etc.). For each model, a likelihood ratio test was conducted to test whether the model fit was significantly improved. If so, stepwise backward selection was applied to obtain a final model.

These models were then used to investigate our main research issues. To examine the extent to which age-related differences in semantic fluency could be explained by IDH, we used permutation tests to compare the contributions of the two age-dependent terms (i.e., Age and its interaction with Mode) in the latter five models against the Demographics-model. For each permutation test, 10000 iterations were run to obtain the null distribution of the change in $\beta_{Mode \times Age}$ and β_{Age} while the permuted predictors were added to the Demographics-model. The p values (one-tailed) associated with the actual observed changes under the null distribution were then determined to a precision of .0001.

To test whether the cognitive processes underlying the four ERP components are differentially involved during automatic and controlled semantic retrieval, Wald Z tests with Tukey's adjustment for multiple comparisons were conducted. Follow-up trend analysis was conducted for each significant interaction between a categorical variable (e.g., Mode) and a continuous predictor (e.g., Age), to test whether the *simple slope* was significant at each level of the categorical variable.

All models were estimated based on maximum likelihood using the function `glmer` in the package `lme4` in R software environment (version 3.6.1). All the models used the logarithm as the link function, and assumed a Poisson error structure suitable for analyzing counts (Gu & Gong, 2016). The backward selection procedure was implemented using the `drop1` function. The function `emtrends` of the package `emmeans` (Lenth et al., 2019) was used to conduct the follow-up trend analyses. For ease of interpretation of the fitted models, the default dummy coding in R was changed to effects coding.

3. Results

3.1. Stroop color and word test

On average, participants scored 87.0 ± 13.8, 59.7 ± 12.2, and 27.5 ± 9.1, in the word, color, and colored word conditions, amounting to an inhibition score of $\text{StroopInhibition} = 0.37 \pm 0.10$. This score was significantly correlated with age, $r = -.34$, $df = 64$, $p = .006$, and with education, $r = .35$, $df = 64$, $p = .004$.

3.2. Semantic fluency

The total semantic fluency, i.e., the individual mean number of correctly produced concepts across the 16 categories, was $N = 11.09 \pm 2.37$. The number of automatic and controlled responses was $N_A = 3.03 \pm 1.18$ and $N_C = 8.06 \pm 2.54$, respectively (see Table S2, for the mean and SD for each category). The number of intrusion and perseveration errors committed by the participants per category were 0.61 ± 0.53 and 0.83 ± 0.59, respectively. Total semantic fluency was significantly correlated with age, $r = -.37$, $df = 64$, $p < .002$, and education, $r = .49$, $df = 64$, $p < .001$.

3.3. Flanker task

3.3.1. Behavioral analysis

In the RT analysis, 0.008% of outlying responses were rejected; this analysis revealed a significant main effect of Response, $F(1, 65) = 116.04$, $p < .001$, $\eta_p^2 = .64$, and of Flanker, $F(1, 65) = 305.49$, $p < .001$, $\eta_p^2 = .82$ (Fig. 2). Participants were faster for RC and FC conditions. Response × Flanker was also significant, $F(1, 65) = 6.14$, $p < .05$, $\eta_p^2 = .09$. This interaction effect was due to the significantly larger simple main effect of Flanker in the RC condition (92.16 ms) than in the RI condition (76.71 ms), as well as the significantly larger simple main effect of Response in the FC condition (75.50 ms) than in the FI condition (60.04 ms).

For error rate, the main effect was significant for Response, $F(1, 65) = 59.98$, $p < .001$, $\eta_p^2 = .48$, and for Flanker, $F(1, 65) = 8.51$, $p < .01$, $\eta_p^2 = .12$ (Fig. 2). Participants were more accurate for the RC and FC conditions. Response × Flanker interaction was non-significant, $F(1, 65) = 0.73$, $p > .39$, $\eta_p^2 = .01$.

3.3.2. ERP analysis

Fig. 3 shows the stimulus- and response-locked grand-averaged ERP waveforms recorded at the three midline electrodes (Fz, Cz, and Pz); the mean number of accepted epochs, out of a maximum of 96 epochs, was: RC-FC: 85.4; RC-FI: 82.8; RI-FC: 80.4; and RI-FI: 77.6. It is worth noting that the P3b amplitude was especially sensitive to the two experimental factors, Response and Flanker. As for the response-locked waveform, it comprised primarily Pc, the positive-going component. For comparison, the grand-averaged waveform for the trials in which the participants committed an error was also shown, clearly showing that an ERN and a subsequent Pe. However, because most participants (38 of 66) committed no more than 20 errors (corresponding to an error rate of about 5%), the ERN could not be reliably measured on an individual level. Also, the CRN was not discernible for all correct conditions[2]. Consequently, only the correct positivity (Pc) was measured.

Given that the main focus of the present study was on the functional associations of inhibitory control components with semantic fluency in older adults but not on the neural correlates of specific inhibitory control processes *per se*, the effects of Response and Flanker were not reported here in detail. Readers are referred to Table S3 for the statistical results based on repeated-measures ANOVA.

3.3.3. Principal component analysis

To obtain a set of linearly independent predictors for each ERP component, PCA was first applied to yield a set of four principal component scores (PC1-PC4), which were ordered by the percentage of variance explained. While the functional meaning of the resultant principal component scores could be deduced from the transformation matrices, we illustrated the meaning by directly correlating each score with four weighted sum, calculated for each individual participant: (1) the average amplitude [(RC-FC + RC-FI + RI-FC + RI-FI)/4]; (2) the Response effect, i.e., the amplitude difference between RI and RC conditions [(RI-FC + RI-FI) − (RC-FC + RC-FI)], (3) the Flanker effect, i.e., the amplitude difference between FI and FC conditions [(RC-FI + RI-FI) − (RC-FC + RI-FC)], and (4) an Interaction Effect, i.e., the difference in Flanker effect between RC and RI conditions [(RC-FI − RC-FC) − (RI-FI − RI-FC)]. A given principal component could then be interpreted in terms of the weighted sum that exhibited the strongest association with it (see Table S4). For example, for the P2 component, PC1 reflected the individual average P2 amplitude due to the high correlation between the two ($r = 1.00$). Similarly, PC2 represented the individual interaction effect ($r = .98$), PC3 the Flanker effect ($r = -.97$), and PC4 the Response effect ($r = -.95$).

3.4. Generalized linear mixed-effects modelling of semantic fluency

3.4.1. Demographics-model

The Demographics-model was summarized in Table 1 (first column). N was larger for controlled than automatic, $z = 46.66$, $p < .001$, and was positively associated with Education, $z = 4.35$, $p < .001$. Importantly, Mode × Age was significant, $z = 5.46$, $p < .001$. Follow-up trend analysis revealed that Age had a significant negative association with N_C, $\beta = -0.021$, $SE = 0.005$, $CI = (-0.032, -0.010)$, $p < .001$, but not with N_A, $\beta = 0.004$, $SE = 0.006$, $CI = (-0.009, 0.017)$, $p > .42$. This interaction was depicted in Fig. 4A. For example, the β of -0.021 for controlled responses indicated that N_C would reduce by a factor of $\exp(-0.021) = 0.979$ per year.

3.4.2. Stroop-model

Comparing the Stroop-model against the Demographics-model using a likelihood ratio test, we found that the two added terms (StroopInhibition and Mode × StroopInhibition) improved the model fit significantly, $\chi^2 = 31.68$, $df = 2$, $p < .001$. No term was excluded by the backward selection procedure (Table 1, second column). Wald Z tests revealed a significant positive association between StroopInhibition and N, $\beta = 0.53$, $SE = 0.22$, $z = 2.40$, $p = .02$. Mode×StroopInhibition was also significant, $\beta = -0.66$, $SE = 0.12$, $z = -5.65$, $p < .001$. Follow-up trend analysis

[2] The CRN component could be revealed by applying Surface Laplacian to the ERP data (Allain et al., 2004). However, due to the relatively small number of electrodes and the lack of information about the head size of the participants, the SL transformation was not adopted in the present study, to guard against the possibility that doing so would introduce artefactual variances in the resultant measures.

revealed that the interaction was explained by the higher association of StroopInhibition with N_C, $\beta = 1.19$, $SE = 0.23$, $CI = (0.75, 1.63)$, $p < .001$, than with N_A, $\beta = -0.12$, $SE = 0.27$, $CI = (-0.66, 0.41)$, $p > .66$.

3.4.3. ERP-models

The likelihood ratio tests between the four ERP-models and the baseline Demographics-model were all significant, $ps < .001$, indicating that the ERP-models were significantly better. Stepwise backward selection was thus applied to each ERP-model, leading to the final models reported in Table 1 (last four columns).

To examine IDH, we used permutation tests to compare the coefficients for the Age and Mode × Age in the models obtained against those in Demographics-model. The results showed that the change in $\beta_{Mode \times Age}$ (Δ) was significant in three models: (1) P2-model, $\Delta = -0.0042$, $SD = 0.0020$, $p = .03$, (2) Pc-model, $\Delta = -0.0031$, $SD = 0.0010$, $p = .02$, and (3) Stroop-model, $\Delta = -0.0037$, $SD = 0.0010$, $p = .02$ (p values were adjusted by the FDR procedure). These indicated that the added predictors in these three models could account for some of the Mode × Age interaction present in the baseline model. The change in $\beta_{Mode \times Age}$ for the two remaining models were non-significant, P3b-model, $p > .79$, N2-model, $p = .08$. We visualized the Mode × Age interaction in (Fig. 4B), where N_C/N_A was plotted against Age for each model; the slope of this plot depended only on $-\beta_{Mode \times Age}$. Here, the Demographics-model had a steep slope, exceeded only by P3b-model. The slope in the remaining models was less steep, and decreased in the following order: N2-model, Pc-model, Stroop-model, and P2-model.

Also, although there was only a trend for a negative association of N with Age in the baseline model, β_{Age} was also significantly reduced in magnitude in Pc-model, $\Delta = 0.0031$, $SD = 0.0009$, $p = .007$, and in Stroop-model, $\Delta = 0.0037$, $SD = 0.0006$, $p = .007$. The change in the remaining model was non-significant, P2-model, $p > .89$, N2-model, $p > .89$, P3b-model, $p > .53$. Taken together, these results provided converging evidence that the differential effect of ageing on semantic retrieval was partially mediated by the individual differences in inhibitory control.

Next, to explore the nature of the association of each ERP component with semantic fluency, Wald Z tests were conducted to examine the significance of each fixed-effects term in the final ERP-models. For each significant interaction, follow-up trend analysis was conducted. However, due to the large number of interactions tested, the detailed statistics of the trend analyses were only reported for each significant association identified. The complete list of significant associations are shown in Table 2. For P2 amplitude, all four PCs significantly interacted with Mode: PC1, $p < .001$, PC2, $p = .015$, PC3, $p < .001$, PC4, $p = .004$, but only PC1 was negatively associated with N, $p = .019$ (Table 1). Follow-up trend analysis revealed a negative association between N_A and PC2, a positive association between N_A and PC3, and negative associations of N_C with both PC1 and PC3 (Table 2). For example, for the association between N_C and PC1, that $\beta = -0.101$ indicated that N_C would decrease by a factor of $\exp(-0.101) = 0.904$ if PC1 increased by 1 µV.

Similarly, for N2 amplitude, only PC2 was kept in the final model. It showed a significant interaction with Mode, $p < .001$, which was due to a negative association with N_A. For P3b amplitude, all four PCs significantly interacted with Mode: PC1, $p < .001$, PC2, $p = .004$, PC3, $p < .001$, PC4, $p = .004$. Trend analysis revealed only a negative association of PC3 with N_A. Finally, for Pc amplitude, PC2 had a significant positive association with N, $p = .005$, while only PC1 significantly interacted with N, $p < .001$. Follow-up trend analysis revealed a positive association of PC1 with N_C.

4. Discussion

4.1. Implications on the inhibition deficit hypothesis

The main aim of the present study was to test whether the IDH extends to the domain of lexico-semantic retrieval using the semantic fluency task. Individual variability in inhibitory control was measured using a behavioral Stroop color and word test and an ERP flanker task within a group of older adults aged 60-79. First, with the baseline Demographics-model, we revealed a significant Mode × Age interaction on the number of correct concepts, with age being negatively associated with number of controlled responses (N_C) but not automatic responses (N_A). This indicated that the age-related decline was more prominent for controlled retrieval. We then showed that the extent of this interaction, as captured by the magnitude of $\beta_{Mode \times Age}$, was significantly reduced in the Stroop-model. Also, although the effect of Age was only marginally significant in the Demographics-model, the effect was significantly weakened in Stroop-model. These results directly demonstrate that the variability in inhibition function contributes to both the interaction and the overall trend of semantic fluency decline.

Considering the non-unitary nature of inhibitory control, we sought to obtain converging neural evidence with the ERP flanker task, which allowed different component processes to be measured. The results showed that, in all four ERP-models, the added ERP predictors significantly improved the model fit over the Demographics-model. However, the magnitude of $\beta_{Mode \times Age}$ was diminished only in two of the models (P2-model and Pc-model), while that of β_{Age} was diminished only in Pc-model. Not only did these results provided converging evidence to show that the differential effect of ageing during lexico-semantic retrieval is partially accounted for by the individual differences in inhibitory control, these results also shed light on the possible mechanism for the association. Specifically, the differential effects of aging on retrieval were mediated by attentional control and performance monitoring, as reflected by P2 and Pc amplitudes.

In most studies on the IDH, the association between inhibitory control and the age-related differences in question have seldom been statistically tested. Instead, the decline of performance in older adults was simply attributed to their declining inhibition function. Adding to this uncertainty is that there have been some inconsistent findings, both in comparing the inhibitory functions of younger and older adults (Brache, Scialfa & Hudson, 2010) and the relative performance of the two groups in the task in question (McArthur et al., 2015). One possible reason for these inconsistencies is that the group differences are specific to the inhibition task in question (Rey-Mermet & Gade, 2018). Alternatively, inhibitory functions may continue to grow during early adulthood, such that as a group, older adults are not necessarily worse than younger adults as the decline begins. In our view, the predictions of IDH should best be tested using either partial correlation analysis or modelling to formally establish the direct association between inhibitory control and performance. To our knowledge, only Persad et al. (2002) has employed a similar rationale. They found that the age-related differences in the verbal list learning task and an attention task were significantly accounted for by inhibition scores in a sample of older adults. The present findings added to their finding by showing that, regardless of whether the conflict task involves suppressing a verbal dimension (Stroop) or not (flanker), the link between inhibitory control and the differential effect of ageing can be found.

4.2. Detailed recruitment of inhibitory mechanisms in semantic fluency

The four ERP-models also shed light on whether inhibitory control processes are differentially recruited during automatic and controlled semantic retrieval. For the P2 component, five P2–fluency associations were found: PC1 (Average) was negatively associated with both N_C and N, PC2 (Interaction Effect) was negatively associated with N_A, while PC3 (Flanker) was positively associated with N_A but negatively with N_C. The PC1 findings suggested that a smaller average P2 amplitude was associated with better overall semantic retrieval, especially with controlled retrieval. The other findings were more subtle in that they involve individual main and interaction effects. The PC2 finding suggested that the participants tended to have a smaller N_A if the individual interaction effects [(RC-FI − RC-FC) − (RI-FI − RI-FC)] was larger. Similarly, the PC3 finding suggested that the participants with more negative individual Flanker effect tended to produce a larger N_A but a smaller N_C. In the verbal fluency literature, attentional control is hypothesized to be important for implementing a consistent strategy in semantic retrieval (Rosen & Engle, 1997). Our findings were consistent with this view.

For the N2 component, PC2 (Flanker) was found to be negatively associated with N_A. Because PC2 was positively correlated to the Flanker effect, this result meant that participants with a larger (more negative) flanker effect tended to produced more automatic responses. This result is consistent with the general view that individuals with better conflict detection function have better automatic lexico-semantic retrieval.

For the P3b component, PC3 (Response) was found to be negatively associated with N_A. Because a smaller PC3 reflects a larger suppression in the P3b amplitude in the RI conditions (hence better response selection function), this suggests that a participant with better response selection function would tend to have a larger N_A. We speculated that response selection is relevant in semantic fluency for choosing the appropriate response from the list of activated concepts. Hence, there is likely some functional overlap between manual and verbal response selection.

For the response-locked Pc component, PC1 (Average) and PC2 (Flanker) were positively associated with N_C and N, respectively. Given that PC1 positively reflected the average Pc amplitude, the PC1 finding supports the hypothesis that performance monitoring is critical in determining the effectiveness of controlled semantic retrieval. As for the PC2 finding, it should be noted that PC2 was most positively associated with individual Flanker effect. A large PC2 thus indicated a smaller reduction in Pc amplitude for incompatible conditions (both RI and FI). Hence, a participant who had smaller Pc amplitude suppression tended to have better overall semantic fluency performance. Taken together, the two results suggested that performance monitoring contributes to semantic fluency performance by allowing adjustments to be made in semantic retrieval. In the literature, higher verbal fluency performance was associated with a less negative CRN amplitude in a sample of older individuals with normal cognition or mild cognitive impairment (Thurm et al., 2013). Our PC1 finding corroborated with this previous result, in that a larger Pc, much like a less negative CRN, could indicate better controlled retrieval performance.

4.3. Significance and limitations

The significance of our findings is two-fold. First, focusing on the age-related differences within an age range of 60-79, the present study provided both behavioral and neural evidence that the variability in inhibitory control within a group of older adults directly contributes to the age-related differences in semantic fluency. Our results complemented those from cross-sectional studies that compared younger and older adults, which often had not formally tested the associations statistically. Second, our ERP findings help clarify the mechanisms behind the association between inhibitory control and semantic fluency. Specifically, we showed that the cascade of domain-general inhibitory control processes involved in a non-verbal conflict task—attentional control, conflict detection, response inhibition, and performance monitoring—are all differentially involved during automatic and controlled lexico-semantic retrieval; however, only the variability in attentional control and performance monitoring contribute to the age-related differences in semantic fluency. Thus, our results also help clarify the possible mechanism behind the link between cognitive and semantic control.

Beyond theoretical interest, the relationship between inhibitory and semantic control that this study demonstrated reiterates the importance of inhibitory functions in older adults. Recent studies have suggested that bilinguals possess better inhibitory control functions than monolinguals (e.g., Bialystok, Craik & Ryan, 2006; Hui et al., 2020), because bilingualism may act as a cognitive reserve and protect individuals against cognitive decline. As such, foreign language learning has been proposed as an effective cognitive training paradigm for improving the inhibitory control functions of older adults (Antoniou, Gunasekera & Wong, 2013). In light of the present findings, it is plausible that older language learners may also see improvement in semantic control, which is essential for everyday communications. These practical ramifications should be explored in future studies.

The present study was limited in two aspects. First, we have not compared inhibitory functions against other cognitive functions such as working memory capacity, which could have played an even more important role than inhibitory control in accounting for age-related differences in semantic fluency. Future works should evaluate the relative importance of inhibitory control and other potential cognitive correlates of semantic fluency. Second, the present study revealed some novel associations between semantic fluency and ERP components in an exploratory manner. Because the model selection procedure of GLMM inherently involved many model comparisons, the specific associations identified may not be sufficiently well-protected against false discoveries. It was for this reason that we limited our analyses to ERP amplitudes but not latencies. Nonetheless, GLMM is known to keep a good balance between false positives and the sensitivity for detection (e.g., Gilmore et al., 2017). Hence, even from the present exploratory findings, it should be clear that the ERP components examined were informative about individual semantic fluency performances. Future studies could extend the present finding by clarifying the link between inhibitory control and other expressive language tasks which also involve active inhibition and selection of concepts.

4.4. Conclusions

The present work showed that ageing affects primarily controlled but not automatic lexico-semantic retrieval. Such differential effect of ageing is partially mediated by the individual variations in inhibitory control in general, but especially attentional control and performance monitoring, as indexed by the amplitudes of P2 and Pc. Our findings suggest that the IDH can extend to accounting for the individual differences in semantic fluency. Our analyses also revealed that all four ERP components examined are engaged during lexico-semantic retrieval.

Supplementary information

Table S1. **The sixteen categories in the semantic fluency task.**

Table S2. **The main performance measures for each category.**

Table S3. **Summary of repeated-measures ANOVAs conducted on the mean ERP amplitudes.**

Table S4. **Principal component analysis on different ERP components.**

Acknowledgements

This work was supported by HKRGC-GRF 15601718 awarded to W.W. and a post-doctoral fellowship awarded to M.F. by Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University.

References

Allain, S., Carbonnell, L., Falkenstein, M., Burle, B., & Vidal, F. (2004). The modulation of the ne-like wave on correct responses foreshadows errors. *Neuroscience Letters*, *372*(1-2), 161–166. https://doi.org/10.1016/j.neulet.2004.09.036.

Antoniou, M., Gunasekera, G. M., & Wong, P. C. (2013). Foreign language training as cognitive therapy for age-related cognitive decline: a hypothesis for future research. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2689–2698. https://doi.org/10.1016/j.neubiorev.2013.09.004.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: one decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Baciu, M., Boudiaf, N., Cousin, E., Perrone-Bertolotti, M., Pichat, C., Fournet, N., Chainay, H., Lamalle, L., & Krainik, A. (2016). Functional MRI evidence for the decline of word retrieval and generation during normal aging. *Age*, *38*(1), 3. https://doi.org/10.1007/s11357-015-9857-y.

Baldwin, S. A., Larson, M. J., & Clayson, P. E. (2015). The dependability of electrophysiological measurements of performance monitoring in a clinical sample: A generalizability and decision analysis of the ERN and Pe. *Psychophysiology*, *52*(6), 790–800. https://doi.org/10.1111/psyp.12401.

Belleville, S., Rouleau, N., & Van der Linden, M. (2006). Use of the Hayling task to measure inhibition of prepotent responses in normal aging and Alzheimer's disease. *Brain and Cognition*, *62*(2), 113–119. https://doi.org/10.1016/j.bandc.2006.04.006.

Bialystok, E., Craik, F. I., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(6), 1341–1354. https://doi.org/10.1037/0278-7393.32.6.1341.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624.

Brache, K., Scialfa, C., & Hudson, C. (2010). Aging and vigilance: Who has the inhibition deficit? *Experimental aging research*, *36*(2), 140–152. https://doi.org/10.1080/03610731003613425.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009.

Donohue, S. E., Appelbaum, L. G., McKay, C. C., & Woldorff, M. G. (2016). The neural dynamics of stimulus and response conflict processing as a function of response complexity and task demands. *Neuropsychologia*, *84*, 14–28. https://doi.org/10.1016/j.neuropsychologia.2016.01.035.

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

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology*, *51*(2-3), 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9.

Faraway, J. J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x.

Fong, M. C.-M., Hui, N. Y., Fung, E. S. W., Chu, P. C. K., & Wang, W. S.-Y. (2018). Conflict monitoring in multisensory flanker tasks: Effects of cross-modal distractors on the N2 component. *Neuroscience Letters*, *670*, 31–35. https://doi.org/10.1016/j.neulet.2018.01.037.

Fong, M. C.-M., Hui, N. Y., Fung, E. S. W., Ma, M. K. H., Law, T. S.-T., Wang, X., & Wang, W. S.-Y. (2020). Which cognitive functions subserve clustering and switching in category fluency? Generalizations from an extended set of semantic categories using linear mixed-effects modelling. *Quarterly Journal of Experimental Psychology*, *73*(12), 2132–2147. https://10.1177/1747021820957135.

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*(1), 622–634. https://doi.org/10.1016/j.neuroimage.2010.07.071.

Gajewski, P. D., & Falkenstein, M. (2013). Effects of task complexity on ERP components in Go/Nogo tasks. *International Journal of Psychophysiology*, *87*(3), 273–278. https://doi.org/10.1016/j.ijpsycho.2012.08.007.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x.

Gilmore, R. O., Diaz, M. T., Wyble, B. A., & Yarkoni, T. (2017). Progress toward openness, transparency, and reproducibility in cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1396*(1), 5–18. https://doi.org/10.1111/nyas.13325.

Golden, C. J., & Freshwater, S. M. (1978). *Stroop Color and Word Test*. Stoelting.

Gordon, J. K., Young, M., & Garcia, C. (2018). Why do older adults have difficulty with semantic fluency? *Aging, Neuropsychology, and Cognition*, *25*(6), 803–828. https://doi.org/10.1080/13825585.2017.1374328.

Gu, Y., & Gong, P. (2016). The dynamics of memory retrieval in hierarchical networks. *Journal of Computational Neuroscience*, *40*(3), 247–268. https://doi.org/10.1007/s10827-016-0595-7.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 163–169. http://doi.org/10.1037/0278-7393.17.1.163.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* 22 (pp. 193–225). Academic Press.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, *36*(2), 185–198. https://doi.org/10.3758/bf03202679.

Hoffman, P. (2018). An individual differences approach to semantic cognition: Divergent effects of age on representation, retrieval and selection. *Scientific Reports*, *8*(1), 8145. https://doi.org/10.1038/s41598-018-26569-0.

Hui, N. Y., Yuan, M., Fong, M. C.-M., & Wang, W. S. (2020). L2 proficiency predicts inhibitory ability in l1-dominant speakers. *International Journal of Bilingualism*, *1*(1), 1–15. https://doi.org/10.1177/1367006920914399.

Kopp, B., Rist, F., & Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*, 282–294. https://doi.org/10.1111/j.1469-8986.1996.tb00425.x.

Kopp, B., Tabeling, S., Moschner, C., & Wessel, K. (2007). Temporal dynamics of selective attention and conflict resolution during cross-dimensional go-nogo decisions. *BMC Neuroscience*, *8*(1), 68. https://doi.org/10.1186/1471-2202-8-68.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility—a model and taxonomy. *Psychological Review*, *97*(2), 253–270. https://doi.org/10.1037/0033-295X.97.2.253.

Korsch, M., Frühholz, S., & Herrmann, M. (2016). Conflict-specific aging effects mainly manifest in early information processing stages—an ERP study with different conflict types. *Frontiers in Aging Neuroscience*, *8*, 53. https://doi.org/10.3389/fnagi.2016.00053.

Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150.

Lenth, R., Singmann, H., Love, J. et al. (2019). Emmeans: Estimated marginal means, aka least-squares means. R package version 1.4. 3.01.

Lindfield, K. C., Wingfield, A., & Bowles, N. L. (1994). Identification of fragmented pictures under ascending versus fixed presentation in young and elderly adults: Evidence for the inhibition-deficit hypothesis. *Aging, Neuropsychology, and Cognition*, *1*(4), 282–291. https://doi.org/10.1080/13825589408256582.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*(3), 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x.

Maier, M. E., Di Pellegrino, G., & Steinhauser, M. (2012). Enhanced error-related negativity on flanker errors: Error expectancy or error significance? *Psychophysiology*, *49*(7), 899–908. https://doi.org/10.1111/j.1469-8986.2012.01373.x.

McArthur, A. D., Sears, C. R., Scialfa, C. T., & Sulsky, L. M. (2015). Aging and the inhibition of competing hypotheses during visual word identification: evidence from the progressive demasking task. *Aging, Neuropsychology, and Cognition*, *22*(2), 220–243. https://doi.org/10.1080/13825585.2014.911240.

McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. *Neuropsychology*, *25*(2), 210–225. https://doi.org/10.1037/a0021531.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

Persad, C. C., Abeles, N., Zacks, R. T., & Denburg, N. L. (2002). Inhibitory changes after age 60 and their relationship to measures of attention and memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(3), P223–P232. https://doi.org/10.1093/geronb/57.3.P223.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019.

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13. https://doi.org/10.1016/j.bandc.2004.03.006.

Rey-Mermet, A., & Gade, M. (2018). Inhibition in aging: What is preserved? What declines? A meta-analysis. *Psychonomic Bulletin & Review*, *25*(5), 1695–1716. https://doi.org/10.3758/s13423-017-1384-7.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211–227. https://doi.org/10.1037/0096-3445.126.3.211.

Sauzéon, H., Raboutet, C., Rodrigues, J., Langevin, S., Schelstraete, M.-A., Feyereisen, P., Hupet, M., & N'Kaoua, B. (2011). Verbal knowledge as a compensation determinant of adult age differences in verbal fluency tasks over time. *Journal of Adult Development*, *18*(3), 144–154. https://doi.org/10.1007/s10804-010-9107-6.

Schreiber, M., Pietschmann, M., Kathmann, N., & Endrass, T. (2011). ERP correlates of performance monitoring in elderly. *Brain and Cognition*, *76*(1), 131–139. https://doi.org/10.1016/j.bandc.2011.02.003.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772.

Thurm, F., Antonenko, D., Schlee, W., Kolassa, S., Elbert, T., & Kolassa, I.-T. (2013). Effects of aging and mild cognitive impairment on electrophysiological correlates of performance monitoring. *Journal of Alzheimer's Disease*, *35*(3), 575–587. https://doi.org/10.3233/jad-121348.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138–146. https://doi.org/10.1037/0894-4105.11.1.138.

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of performance monitoring and adaptive behavior. *Physiological reviews*, *94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *The Quarterly Journal of Experimental Psychology*, *64*(3), 447–466. https://doi.org/10.1080/17470218.2010.505292.

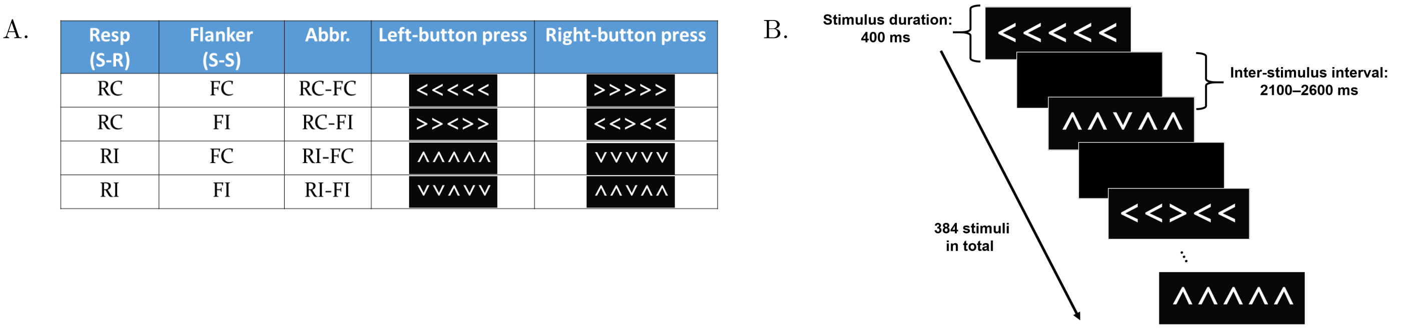
Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165–181. https://doi.org/10.1027/0269-8803.19.3.165.

Vonk, J. M., Rizvi, B., Lao, P. J., Budge, M., Manly, J. J., Mayeux, R., & Brickman, A. M. (2019). Letter and category fluency performance correlates with distinct patterns of cortical thickness in older adults. *Cerebral Cortex*, *29*(6), 2694–2700. https://doi.org/10.1093/cercor/bhy138.

Wild-Wall, N., Falkenstein, M., & Hohnsbein, J. (2008). Flanker interference in young and older participants as reflected in event-related potentials. *Brain Research*, *1211*, 72–84. https://doi.org/10.1016/j.brainres.2008.03.025.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931.

Zhou, S., Xiong, S., Cheng, W., & Wang, Y. (2019). Flanker paradigm contains conflict and distraction factors with distinct neural mechanisms: an ERP analysis in a 2-1 mapping task. *Cognitive Neurodynamics*, *13*(4), 341–356. https://doi.org/10.1007/s11571-019-09529-w.



A.

Resp (S-R)	Flanker (S-S)	Abbr.	Left-button press	Right-button press
RC	FC	RC-FC	<<<<<	>>>>>
RC	FI	RC-FI	>><>>	<<><<
RI	FC	RI-FC	∧∧∧∧∧	∨∨∨∨∨
RI	FI	RI-FI	∨∨∧∨∨	∧∧∨∧∧

B.
Stimulus duration: 400 ms
<<<<<
Inter-stimulus interval: 2100–2600 ms
∧∧∨∧∧
<<><<
384 stimuli in total
∧∧∧∧∧

Fig. 1. A. The two-by-two design of the flanker task involving Response (RC/RI) and Flanker (FC/FI). B. Experimental paradigm used in the present study; stimuli in the four conditions were randomly presented, at an inter-stimulus interval (ISI) of 2100-2600 ms.

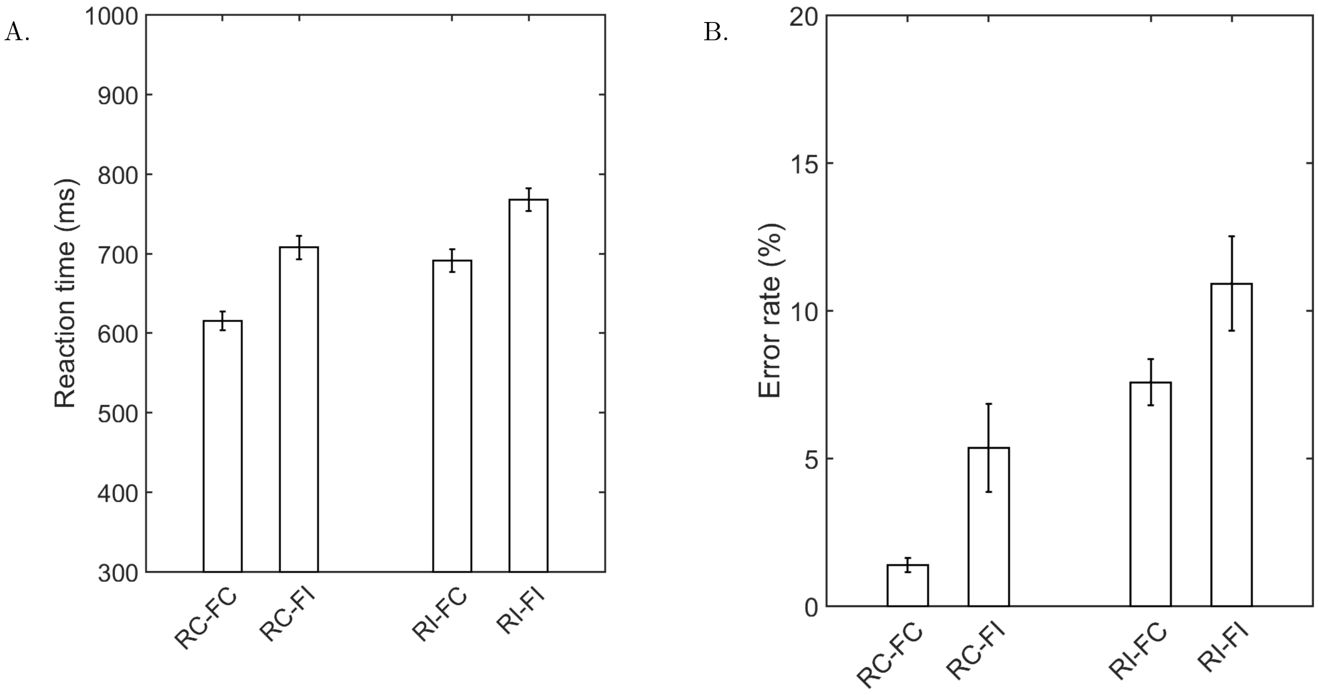


Fig. 2. Behavioral performance across the four conditions in the ERP flanker task. A. RT (ms); B. Error rate (%).

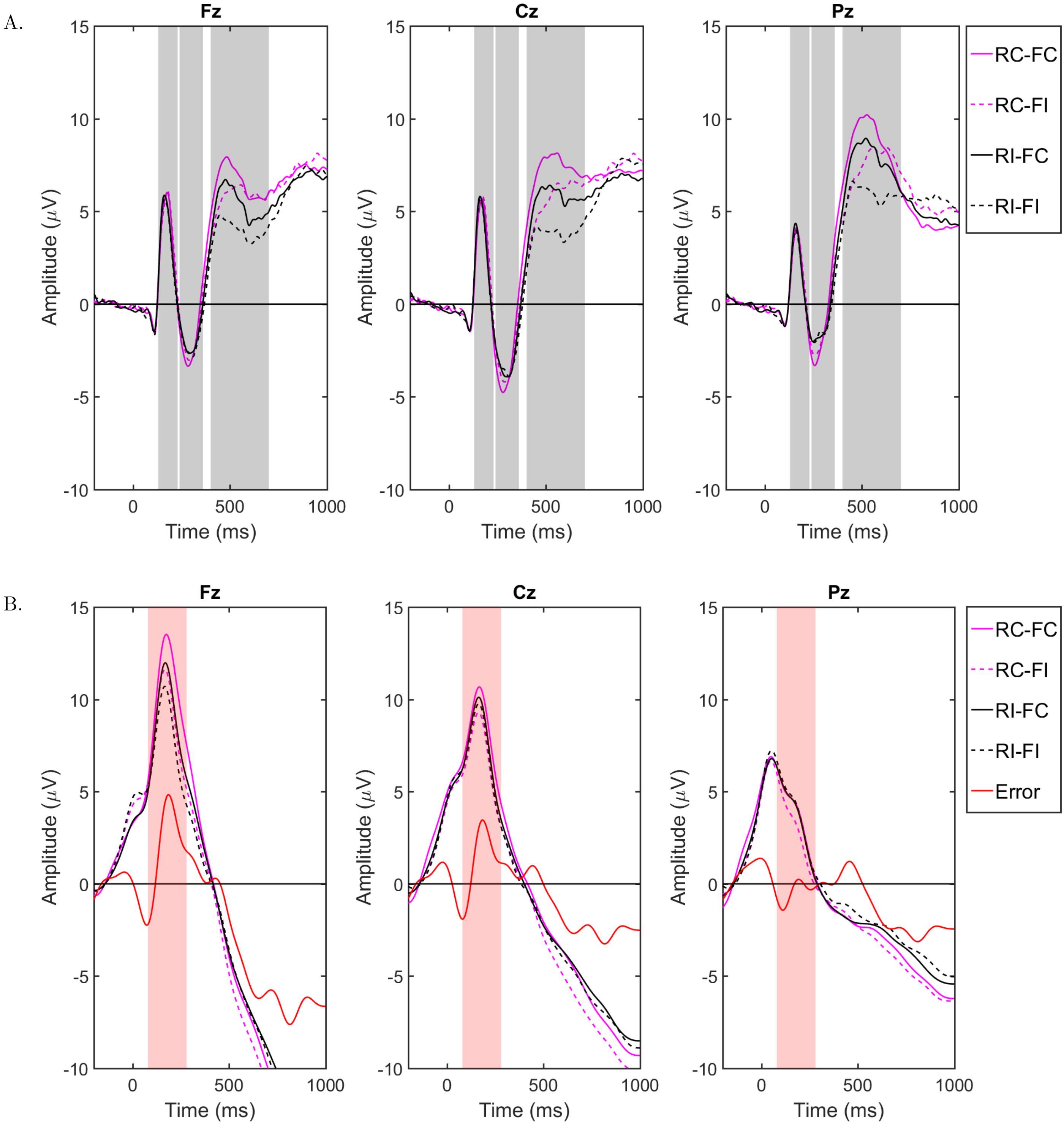


Fig. 3. A. Time-locked grand-averaged ERP at Fz (left), Cz (middle), and Pz (right). B. Response-locked grand-averaged ERP at the same three electrodes. Three windows were highlighted in grey for time-locked ERP: P2 (130–230 ms), N2 (240–360 ms), and P3b (400–700 ms), while one window was highlighted in pink for Pc (80–280 ms). The response-locked ERP for the error trials was also shown for comparison with the correct trials.

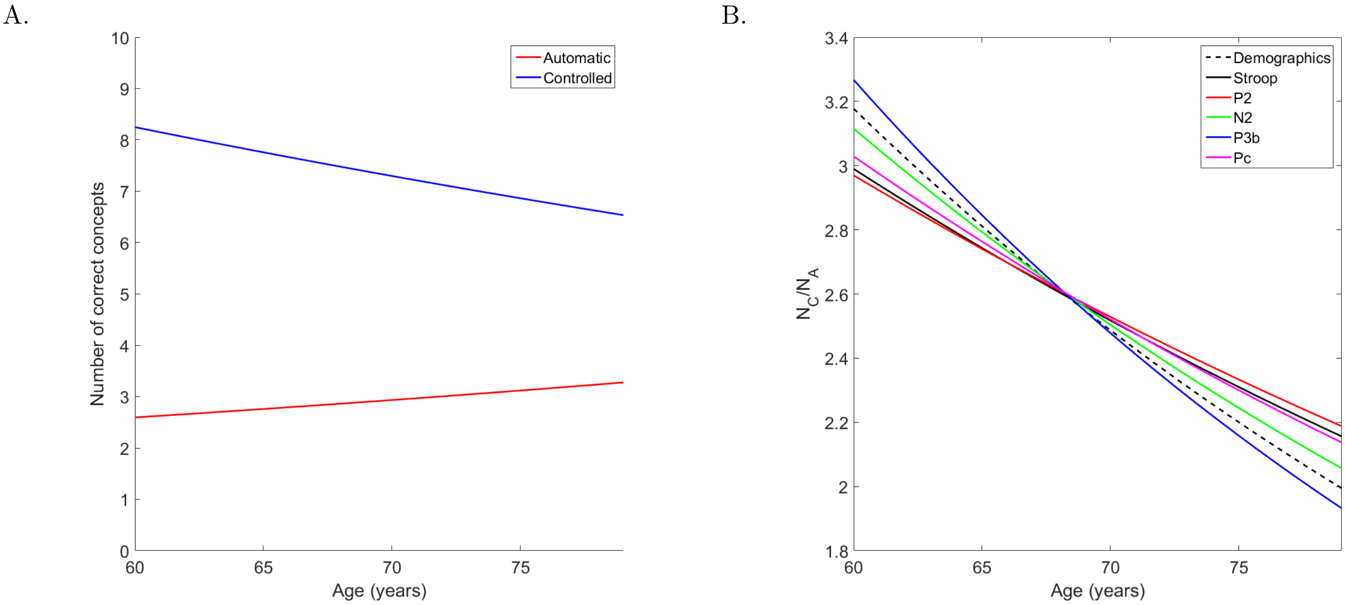


Fig. 4. A. Interaction between Mode (automatic/controlled) and Age. B. Changes in the Mode × Age interaction across the six GLMM models.

Table 1. The six GLMM models.

	Demographics	**Stroop**	**P2**	**N2**	**P3b**	**Pc**
Statistics						
AIC	10156.42	10128.74	10092.06	10136.25	10130.75	10115.35
BIC	10195.90	10179.51	10176.68	10187.02	10215.37	10171.76
logLik	-5071.21	-5055.37	-5031.03	-5059.12	-5050.38	-5047.67
deviance	3433.92	3412.34	3370.94	3409.30	3392.91	3399.46
df.resid	2075	2073	2067	2073	2067	2072
Random effects						
Participant†	0.159	0.147	0.139	0.160	0.159	0.144
Category†	0.290	0.290	0.289	0.290	0.290	0.290
Fixed effects (β)						
(Intercept)ˆ	1.532***	1.532***	1.530***	1.531***	1.529***	1.531***
Ageˆ	-0.008	-0.006	-0.010	-0.009	-0.008	-0.005
Educationˆ	0.019***	0.015***	0.017***	0.019***	0.019***	0.021***
Modeˆ	-0.487***	-0.484***	-0.485***	-0.487***	-0.490***	-0.490***
Age:Modeˆ	0.012***	0.009***	0.008**	0.011***	0.014***	0.009***
StroopInhibitionˆ	-	0.047*	-	-	-	-
StroopInhibition:Modeˆ	-	-0.054***	-	-	-	-
PC1ˆ	-	-	-0.047**	#	-0.016	0.008
PC1:Modeˆ	-	-	0.053***	#	-0.038***	-0.064***
PC2ˆ	-	-	-0.038	-0.020	-0.014	0.058**
PC2:Modeˆ	-	-	-0.026*	-0.053***	-0.031**	#
PC3ˆ	-	-	0.005	#	-0.029	#
PC3:Modeˆ	-	-	0.057***	#	-0.038***	#
PC4ˆ	-	-	0.005	#	-0.012	#
PC4:Modeˆ	-	-	0.031**	#	-0.031**	#

Note. †, SD; ˆ, a fixed-effect coefficient β; #, term was removed by stepwise backward selection; ***, $p < .001$, **, $p < .01$, *, $p < .05$.

Table 2. Post hoc trend analysis of the significant interactions between the PCs and Mode in the four ERP-models.

Component	PC	Interpretation	r	Automatic			Controlled		
				β	SE	p	β	SE	p
P2	PC1	Average	1.00	0.006	0.025	.81	-0.101	0.02	$< .001^{***}$
	PC2	Interaction Effect	0.98	-0.064	0.026	$.03^{*}$	-0.012	0.021	1.00
	PC3	Flanker	-0.97	0.062	0.027	$.04^{*}$	-0.053	0.022	$.03^{*}$
	PC4	Response	-0.95	0.036	0.026	.33	-0.026	0.021	.44
N2	PC2	Flanker	0.94	-0.073	0.028	$.02^{*}$	0.033	0.023	.31
P3b	PC1	Average	1.00	-0.054	0.028	.11	0.021	0.024	.74
	PC2	Flanker	0.95	-0.045	0.027	.20	0.017	0.023	.91
	PC3	Response	0.89	-0.067	0.026	$.02^{*}$	0.009	0.023	1.00
	PC4	Interaction Effect	-0.72	-0.043	0.027	.23	0.019	0.023	.80
Pc	PC1	Average	1.00	-0.056	0.026	.07	0.072	0.021	$.001^{**}$

Note. r, the correlation between the PC and weighted sum that supported our interpretation of the PC.