Original Research Article

# Assessment of Automatic and Controlled Retrieval Using Verbal Fluency Tasks

Assessment
2023, Vol. 30(7) 2198–2211

DOI: 10.1177/10731911221117512
journals.sagepub.com/home/asm
SAGE

**Martin Marko[1,2], Drahomír Michalko[1], Jozef Dragašek[3], Zuzana Vančová[3], Dominika Jarčušková[3], and Igor Riečanský[1,4,5]**

## Abstract

Category and letter verbal fluency assessment is widely used in basic and clinical research. Yet, the nature of the processes measured by such means remains a matter of debate. To delineate automatic (free-associative) versus controlled (dissociative) retrieval processes involved in verbal fluency tasks, we carried out a psychometric study combining a novel lexical-semantic retrieval paradigm and structural equation modeling. We show that category fluency primarily engages a free-associative retrieval, whereas letter fluency exerts executive suppression of habitual semantic associates. Importantly, the models demonstrated that this dissociation is parametric rather than absolute, exhibiting a degree of unity as well as diversity among the retrieval measures. These findings and further exploratory analyses validate that category and letter fluency tasks reflect partially distinct forms of memory search and retrieval control, warranting different application in basic research and clinical assessment. Finally, we conclude that the novel associative-dissociative paradigm provides straightforward and useful behavioral measures for the assessment and differentiation of automatic versus controlled retrieval ability.

## Keywords

behavioral assessment, clinical assessment, semantic fluency, phonemic fluency, semantic memory retrieval, cognitive control, inhibition, structural equation modeling

The ability to retrieve context-relevant knowledge from memory is fundamental for everyday adaptive behavior, thus attracting much attention in basic and applied psychological sciences. Lexical-semantic retrieval is commonly assessed by means of verbal fluency tasks requiring individuals to continuously generate verbal responses according to specified rules in a limited time (Zemla et al., 2020). Typically, the testing requires exploration and retrieval of words either belonging to a semantic category (e.g., *animals*) or starting with a letter (e.g., *F*) for up to 1 minute (Abwender et al., 2001), thus referred to as *category fluency* (CF; also known as *semantic fluency*) and letter fluency (LF; also known as *phonemic fluency*), and *letter fluency* (LF), respectively. Importantly, except for probing human mnestic, language, and executive functions in healthy individuals (Diamond, 2013; Shao et al., 2014; Whiteside et al., 2016), verbal fluency tasks have been extensively applied to evaluate the severity of cognitive impairment in a range of neuropsychiatric disorders, including attention-deficit/hyperactivity disorder, Alzheimer's disease, schizophrenia, depression, and autism (Baldo et al., 2006; Henry & Crawford, 2004; Monsch et al., 1994; Schmidt et al., 2019; Tröster et al., 1998; Zhao et al., 2013).

Intense research and debates concern the cognitive and neural processes supporting CF versus LF, their specificity, and diagnostic sensitivity. On one hand, both forms of fluency share several core processes, such as self-monitoring, planning, maintaining and updating cues, post-retrieval selection, and mental flexibility (Rende et al., 2002; Troyer et al., 1998; Unsworth et al., 2013), and are related to individual processing speed as well as verbal abilities (Henry & Phillips, 2006; Whiteside et al., 2016). Consistent with this view, evidence from neuroimaging and neuropsychiatric patients indicate that CF and LF tasks engage overlapping

[1]Centre of Experimental Medicine, Slovak Academy of Sciences, Bratislava, Slovakia
[2]Comenius University in Bratislava, Slovakia
[3]Pavol Jozef Šafárik University, Košice, Slovakia
[4]University of Vienna, Austria
[5]Slovak Medical University in Bratislava, Slovakia

**Corresponding Author:**
Igor Riečanský, Department of Behavioural Neuroscience, Institute of Normal and Pathological Physiology, Centre of Experimental Medicine, Slovak Academy of Sciences, Sienkiewiczova 1, 813 71 Bratislava, Slovakia.
Email: igor.riecansky@savba.sk

temporal-parietal and frontal brain regions supporting semantic representation and retrieval control, respectively (Biesbroek et al., 2016; Katzev et al., 2013; Schmidt et al., 2019). Yet, several fine-grained analyses suggest that CF versus LF tasks may engage partially distinct cognitive processes and resources. For instance, CF seems to exert greater demands on the integrity of semantic representation (i.e., long-term stores) underpinned by the temporal lobe (Grogan et al., 2009; Tröster et al., 1998; Troyer et al., 1998) and the hippocampus (Henry & Crawford, 2004; Schmidt et al., 2019), whereas LF was proposed to predominantly engage controlled access to memory representations and mental lexicon, supported by the frontal lobe (Baldo et al., 2010; Gourovitch et al., 2000). Furthermore, the two tasks may encourage distinct search strategies—while the category retrieval urges exploiting a well-organized semantic knowledge, the letter-cued retrieval requires individuals to follow phonological/orthographic criteria. Consequently, given that category search mimics the natural way we search and retrieve knowledge (Iudicello et al., 2012; Shao et al., 2014), it more closely reflects automatic associative retrieval processes. By contrast, as letter search has to comply with artificial phonological constrains, this task entails the need to suppress the natural mode of retrieving concepts by meaning and semantic associations (Heim et al., 2009; Henry & Crawford, 2004; Perret, 1974), and thus can be considered more executively demanding. Therefore, despite some similarities between the CF and the LF tasks, the current evidence suggests that they differ in the relative involvement of automatic versus controlled retrieval processes. Nevertheless, direct psychometric evaluation of such dissociation is lacking. Moreover, the assumption that LF demands more cognitive control than CF, particularly the capacity to suppress retrieving words by meaning (i.e., prepotent semantic associations), also awaits an explicit verification. These open questions are addressed in the present study.

Recent years have brought attempts to differentiate automatic versus controlled (executive) processes in memory retrieval, striving to provide more informative measures for research and applied settings. With this regard, we have developed and introduced a novel generative retrieval paradigm, the Associative Chain Test (ACT) (Marko, Michalko, et al., 2019; Marko & Riečanský, 2021b). In this paradigm, individuals continuously generate sequences of related (free-*associative*) and unrelated (*dissociative*) words first separately, in two blocks with a *fixed* rule, and then deliver related and unrelated word alternately, in a third block with a *switching* rule. This structure enables addressing two important conceptual contrasts. The first stems from the evidence that free-associative production engages an automatic mode of retrieval with little demands on cognitive control (Collins & Loftus, 1975; Gray et al., 2019; Marron et al., 2018), whereas generating unrelated words exerts more time and demands on controlled inhibition, i.e., the *inhibition cost* (IC) (Allen et al., 2008; Collette et al., 2001; Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a, 2021b). Following that, the dissociative condition (and IC, see Figure 1B) operationalizes the efficiency of controlled memory retrieval, particularly the ability to inhibit automatic but semantically inappropriate responses. The second contrast stems from the evidence that flexible alternating between the retrieval rules further increases retrieval latency (particularly for dissociative trials), which has been attributed to demands on controlled switching, i.e., the *switching cost* (SC) (see Figure 1B; Marko, Cimrová, & Riečanský, 2019; Marko & Riečanský, 2021a; Mayr & Kliegl, 2000; Troyer et al., 1997). Since switching is essential to produce words while rules and semantic sets are rapidly changing, the SC operationally defines the cognitive (in)flexibility during lexical-semantic memory retrieval.

Notably, in line with the proposed distinction between automatic (associative) versus controlled (dissociative, switching) processes in ACT, recent studies have shown that free-*associative* retrieval engages the default mode brain network (assumed to support spontaneous thought), whereas *dissociative* retrieval and switching likely recruit the executive control brain network (Allen et al., 2008; Hyafil et al., 2009; Marko & Riečanský, 2021a; Marron et al., 2020). Taken together, ACT is a novel behavioral research method, which may provide informative measures for exploring the variety of processes involved in verbal fluency, leading to a more proper understanding and assessment of the core functions governing lexical-semantic memory retrieval.

## The Present Study

The present study aimed to evaluate the hypothesized parametric differences in the contribution of automatic versus controlled processes in (category vs. letter) verbal fluency by using the (associative vs. dissociative) measures from the novel retrieval paradigm (ACT). Before addressing that, we tested three related assumptions: (a) retrieving words by *letter* is more demanding (i.e., takes more time) than by *category*, while using standard testing stimuli; (b) the *dissociative* retrieval takes more time than the *associative* retrieval (due to demands on controlled inhibition); and that (c) alternating the rules further increases the latency of *dissociative* retrieval (due to demands on controlled switching).

Subsequently, we investigated the main hypotheses. Our first hypothesis stemmed from previous theoretical views that the retrieval measures may share a set of cognitive and neural resources (Henry & Crawford, 2004; Rende et al., 2002), yet differentially engage automatic versus controlled/executive processes (Heim et al., 2009; Martin et al., 1994; Perret, 1974). In particular, we hypothesized that

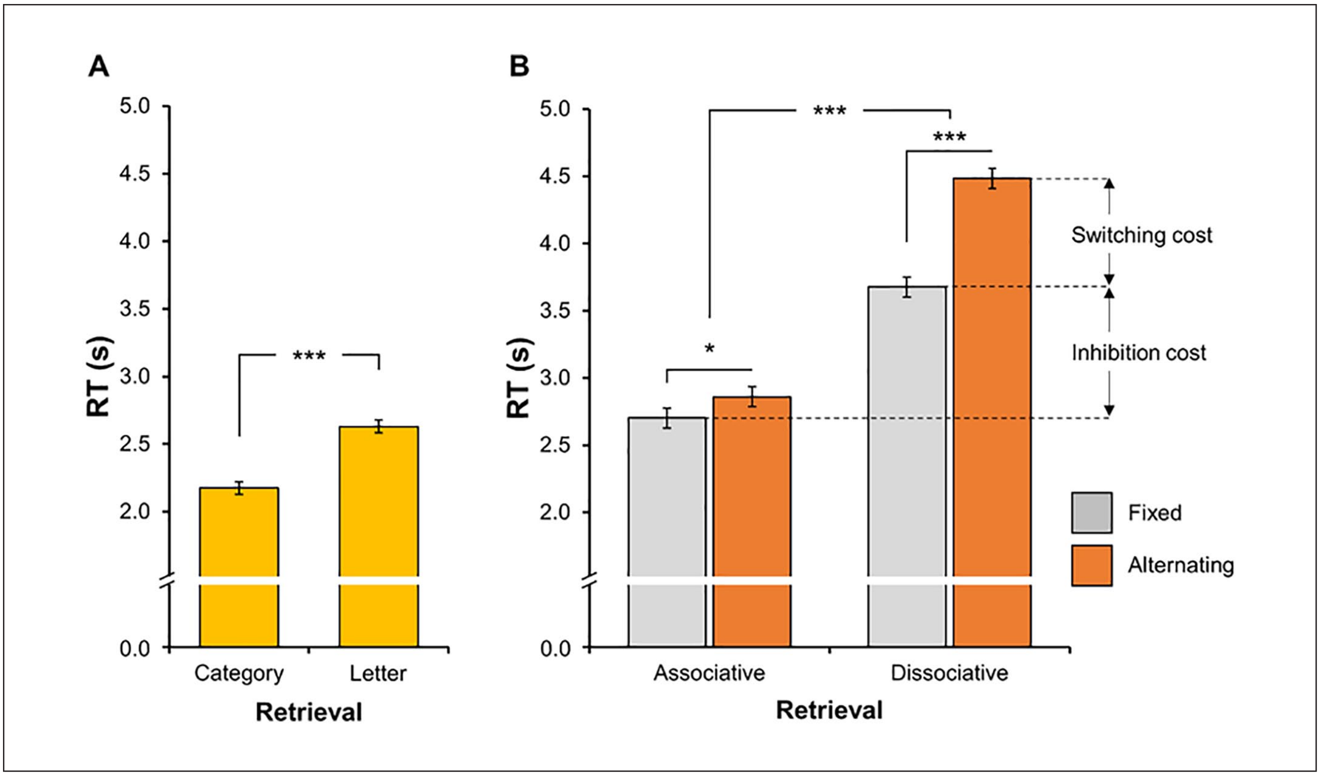


**Figure 1.** Retrieval performance in Verbal Fluency Tasks (A) and the Associative Chain Test (B)
*Notes. Inhibition cost* represents the difference between the dissociative and the associative retrieval in fixed chains (gray bars). *Switching cost* represents the difference between the alternating (orange bars) and the fixed dissociative retrieval. Error bars represent $\pm SE$.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$ (adjusted using Holm correction where appropriate).

retrieval in the *category* condition would be more strongly coupled with the *associative* (i.e., automatic) than the *dissociative* (i.e., controlled) mode of retrieval in a confirmatory correlated factor model (CFM). Complementarily, we hypothesized that retrieval in the *letter* condition would be more strongly related to the *dissociative* than the *associative* retrieval in the same model. To test this complex hypothesis, latent scores were estimated for each retrieval measure (i.e., *category/letter* latency and *associative/dissociative* latency) using structural equation modeling (SEM) (Friedman & Miyake, 2017), and their mutual associations were assessed. Here, we expected the latent constructs to be substantially correlated to each other (i.e., showing a degree of "unity") but, more importantly, we expected the parametric dissociation among these correlations (i.e., a degree of "diversity"). This expected pattern was further verified by comparing the differences among pair-wise correlations between latent retrieval scores estimated from the model. The next step was to address our second hypothesis that, in addition to the processes and resources shared by all tasks (e.g., access to memory representations, planning, or self-monitoring), the *dissociative* and *letter* retrieval exert additional demands on semantic control (e.g., inhibition and/or response selection; Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a; Perret, 1974). This hypothesis was tested using a bifactor model (Dunn & McCray, 2020), where we expected that, after partialling out the common variance shared across all retrieval measures (i.e., the general factor), the specific factors for *dissociative* and *letter* retrieval (but not for the *associative* and *category* retrieval) would retain significant factor variances and loadings, and show higher explanatory power ($R^2$). Here, significant variance, reliable loadings, and the remaining explanatory power of a specific factor were considered as evidence for a retrieval-specific process included in the respective task (e.g., inhibitory processes in dissociative and letter retrieval tasks) not captured fully by the general factor. Essentially, confirming the hypotheses above would imply that CF and LF tasks exert relatively distinct demands on automatic memory retrieval (memory storage and vocabulary) versus the retrieval control (executive regulation), bearing important consequences for their practical application and assessment of the respective memory functions.

Finally, a confirmatory factor analysis (CFA) including the IC and the SC derived from ACT was conducted to explore relationship between these two control processes (Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a, 2021b). These latent scores were further subjected to an exploratory correlation analysis, along with the latent verbal fluency scores, to explore whether LF has a particular link to inhibition (i.e., suppressing the natural mode of retrieval by the meaning and associations). Notably, all the considerations above were discussed with respect to the (convergent and divergent) validity of the respective retrieval tasks/measures, their specific utility for basic and applied research, as well as their diagnostic value in clinical settings.

## Methods

### Participants

Sample size was determined by an a priori power analysis in *G*Power* (Faul et al., 2007) and *semPower* (R package, see Jobst et al., 2021) at the level of $\alpha = .05$ and $1 - \beta = .80$. For SEM, the required sample size of $N \geq 176$ was estimated for detecting significant misfit (root mean square error of approximation [RMSEA] > .05) between hypothesized and saturated models (a detailed description of the sample size calculation for all procedures is provided in the Supplemental materials file). Following that, a group of 189 healthy young adults, primarily undergraduate students (mean age ± $SD = 22.75 \pm 1.89$ years; 51 males, 109 nonsmokers), took part and completed the study including two separate sessions. Four more individuals were recruited to participate but were excluded as they did not finish the whole procedure. Upon arrival for the first session, participants completed a short anamnestic questionnaire in a computerized form. Here, the study participants reported no history of neurological condition, psychiatric illness, or recent use of medication. Thereafter, trait anxiety and schizotypal personality traits were assessed using State-Trait Anxiety Inventory (Spielberger et al., 1983) and Schizotypal Personality Questionnaire (Raine, 1991), respectively. The mean trait anxiety score ($M = 40.58$, $SD = 10.71$) was not statistically different from the normative estimate ($M = 40.90$, $SD = 8.51$; $p > .695$; Heretik et al., 2009), whereas the mean schizotypy score ($M = 13.64$, $SD = 10.01$) was slightly below the previous estimate ($M = 15.74$, $SD = 7.45$; $p = .015$; Chylová et al., 2017). Research was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and was approved by the institutional review board. All procedures and methods were carried out in accordance with the relevant guidelines and regulations. All participants gave written informed consent and received a financial reward for their participation.

### Procedure

This study was a part of a larger project investigating declarative and procedural memory. Here we report the data and procedures relevant for the a priori hypotheses concerning lexical-semantic retrieval. Lexical-semantic retrieval was assessed in two separate sessions, each lasting up to 50 minutes, carried out in two consecutive days. At the beginning of the sessions, participants reported their sleep quality, sleep quantity (i.e., estimated sleep duration in hours), and perceived arousal. There were no statistically significant differences in these self-reported measures between the two sessions ($p > .132$, $d < 0.11$; see the supplemental materials for more details). The sessions were carried out by two trained experimenters unaware of the main research intent and followed the same standard procedure including a computerized version of (a) letter fluency tasks, (b) category fluency tasks, and (c) the ACT, in this fixed order. The tasks were administered in small groups ($N \leq 15$ individuals), intermitted by short breaks in-between them. In all retrieval tasks and conditions, participants typed the responses using computer keyboard. Response time (RT) for each trial (i.e., generated word) was calculated as the latency between the trial onset and first key press (words with at least two characters were allowed to enter as responses).

### Verbal Fluency Tasks

Verbal fluency was assessed in two forms, letter (phonemic) and category (semantic), both including a short practice trial and two test trials per session (i.e., there was a total of four test trials for each form). In the LF, participants were instructed to generate as many words as possible beginning with a specific letter ("*K*" and "*D*" in Session 1, "*L*" and "*T*" in Session 2). In the CF, participants were asked to generate exemplars belonging to a specific category ("*Animals*" and "*Sports*" in Session 1, "*Plants*" and "*Occupations*" in Session 2). Trials in both verbal fluency tasks lasted 50 seconds each. The sequences delivered by the participants were screened by two trained independent raters for inappropriate and repeated responses (5.6%), which were excluded before statistical processing. Notably, a sequence had to contain at least five appropriate and unique words/exemplars to be included in the analyses (see the Supplemental materials for more details). RT was defined as the average RT between the consecutive responses, separately for each trial of the respective tasks (note that RTs were used rather than the absolute number of generated words to match the response format with the ACT). The total administration time for verbal fluency tasks was approximately 10 minutes in each session.

### Associative Chain Test (ACT)

Lexical-semantic retrieval and functions were then assessed using the ACT (Marko, Cimrová, & Riečanský, 2019; Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a). In this experimental paradigm, individuals continuously produced chains of words according to three specific retrieval conditions, each including a short practice trial and two test trials per session (i.e., there were four test trials for each condition). In the *associative* condition, participants were instructed to generate chains of words so that each new response in the chain was semantically related to the previous one (e.g., *Tree* [starting word] ← *Apple* ← *Fruit* ← *Sugar*, etc.) for 40 seconds. Subjects were instructed to respond with the first association that spontaneously comes to their mind, which is considered to involve automatic retrieval processes (Marko, Michalko, & Riečanský, 2019; Marron et al., 2018). In the *dissociative* condition, participants were instructed to produce a chain of unrelated words for 50 seconds, where each new response should not relate to the previous one (e.g., *Table* [starting word] ← *Engine* ← *Owl* ← *Pen, etc.*), and reminded that responding with a related word was considered an error. In contrast to the associative condition, retrieving unrelated words (dissociates) requires substantially more time and additional demands on inhibition (Allen et al., 2008; Collette et al., 2001; Marko, Cimrová, & Riečanský, 2019; Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a). Finally, in the alternating (*associative–dissociative)* condition, the participants alternated between delivering associations and dissociations (i.e., the retrieval rule switched after each response; *Tree* [starting word] ← *Apple* ← *Engine* ← *Car* ← *Envelope, etc.) for 60 seconds.* Notably, compared with the previous *fixed* retrieval conditions, the *alternating* condition imposes additional demands on flexible switching (Marko, Cimrová, & Riečanský, 2019; Marko, Michalko, & Riečanský, 2019; Marko & Riečanský, 2021a).

The ACT included two main factors (*retrieval type* [associative vs. dissociative] × *chain type* [fixed vs. alternating]). For each factor combination, the responses were screened by two trained independent raters for errors and repetitions (9.5%), which were excluded prior to the statistical analysis. Also, a set of responses from a factor combination had to contain at least three appropriate and unique responses to be included in the analyses (see the Supplemental materials for more details). Then, for each factor combination and trial, the performance was estimated as the average RT between the consecutive responses. Finally, two additional measures of controlled lexical-semantic retrieval were extracted: IC (the difference between dissociative and associative retrieval RT in fixed chains) and SC (the difference between dissociative RT in the alternating vs. fixed chains; see the Supplemental materials for more details). The total administration time for ACT was approximately 10 minutes in each session.

### *Statistical Analysis*

The data were processed and analyzed in RStudio (RStudio Team, 2021) and JASP (JASP Team, 2020). A descriptive statistical analysis of the raw data indicated the presence of outlying observations (values exceeding median ±1.5 interquartile range) and thus, the RTs from verbal fluency tasks and ACT were winsorized using a 20% quantile two-sided trimming (i.e., 20% of responses from each tail of the distribution), separately for each participant and condition/factor combination. Thereafter, the individual RT values for each task and condition were averaged and all performance measures were calculated. In analysis of variance (ANOVA), the effect sizes were estimated using $\eta_p^2$ and post hoc *p*-values were corrected with Holm adjustment wherever appropriate. The respective 95% confidence intervals (CI) indicate the estimated differences between the retrieval conditions. The ANOVA reports were complemented with Bayes factor ($BF_{10}$), that is, the ratio of the likelihood of a hypothesized model (or effect) against a null or simpler hypothesis, indicating the strength of evidence in favor of it (van Doorn et al., 2021). SEM was employed to test the main hypotheses. As the assumption of multivariate normality was violated in the models, all SEMs were estimated using a robust diagonally weighted least squares (DWLS) procedure supplied with robust standard errors (Mîndrilă, 2010). The following indices and indicative criteria were used to assess the model fit: $\chi^2$ goodness of fit test ($p > .05$), Comparative Fit Index (CFI ≥ .95), Tucker Lewis Index (TLI ≥ .95), RMSEA ≤ .08, and Standardized Root Mean Square Residuals (SRMR ≤ .08). Mean explained variance among the indicators (mean $R^2$) was calculated for each model. Finally, the hypothesized differences among the correlations between estimated latent scores were assessed using *z* tests (for overlapping and non-overlapping correlation pairs; Hittner et al., 2003; Silver et al., 2004) and corrected using Holm adjustment.

## Results

### *Differences Between the Retrieval Tasks*

For verbal fluency tasks, a one-way repeated measures ANOVA indicated a robust main effect of *retrieval* (category vs. letter) on performance (averaged across respective trials), $F(1,174) = 87.14$, $p < .001$, $\eta_p^2 = .334$. As expected, the average retrieval latency was shorter for category ($M = 2.175$ seconds, $SE = 0.047$ seconds) than letter ($M = 2.630$ seconds, $SE = 0.047$ seconds) retrieval, the mean difference was 0.455 seconds, 95% CI [0.359, 0.551]. This difference was further supported by a Bayes factor analysis, yielding $BF_{10} = 1.066e+^{14}$, which indicated very strong evidence favoring the hypothesized difference (for more details see Figure 1A).

For ACT, a 2 × 2 repeated measures ANOVA revealed a significant main effect of *retrieval* (associative vs. dissociative), $F(1,146) = 418.60$, $p < .001$, $\eta_p^2 = .741$, *chain* (fixed vs. alternating), $F(1,146) = 106.50$, $p < .001$, $\eta_p^2 = .422$, as

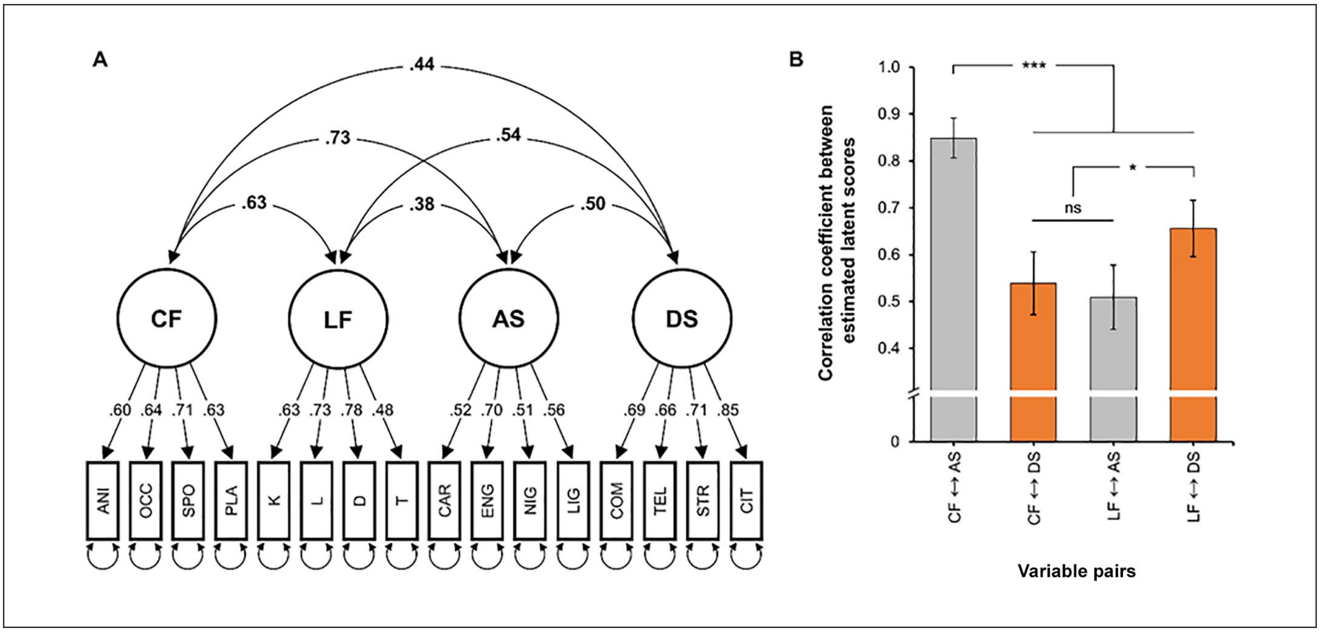


**Figure 2.** Correlated Factor Model (A) and Pair-wise Correlations Between the Estimated Latent Retrieval Scores (B).
*Note.* Panel A: standardized estimates; all covariates (curved arrows) and regression loads (straight arrows) were significant ($p < .001$). Panel B: bivariate correlation coefficients (Pearson's r) between latent scores from the unconstrained model ($CFM_{unc}$). ns = non-significant; Error bars represent ±SE; Latent factors/retrieval scores: CF = category; LF = Letter; AS = Associative; DS = Dissociative.
$^{*}p < .05$ and $^{***}p < .001$ (adjusted using Holm correction) indicate significant differences between the correlation coefficients.

well as their interaction, $F(1,146) = 15.56$, $p < .001$, $\eta_p^2 = .228$. Post hoc tests with Holm adjustments showed that, for fixed chains, the retrieval latency was shorter in associative ($M = 2.702$ seconds, $SE = 0.074$ second) than dissociative ($M = 3.677$ seconds, $SE = 0.074$ second) retrieval, indicating that the IC of 0.975 second, 95% CI [0.761, 1.189], was significant ($p_{holm} < .001$). Furthermore, a comparison of the dissociative retrieval between fixed ($M = 3.677$ seconds, $SE = 0.074$ second) and alternating ($M = 4.484$ seconds, $SE = 0.074$ second) chains showed that the SC of 0.808 second, 95% CI [0.627, 0.989], was also significant ($p_{holm} < .001$). Notably, for associative performance, the difference between fixed and alternating condition was 0.157 second, 95% CI [−0.024, 0.338], that is, substantially smaller, but significant ($p_{holm} = .022$; for more details see Figure 1B.). The Bayes factor analysis for ACT showed very strong evidence for the effect of *retrieval* ($BF_{10} > 9.350e+^{74}$), *chain* ($BF_{10} > 3.852e+^{13}$), and their interaction ($BF_{10} > 2.749e+^{6}$) in the ANOVA model. The BF associated for the respective post hoc *t*-tests pertaining to the *inhibition* (dissociative fixed vs. associative fixed; $BF_{10} > 9.016e+^{22}$) and SC (dissociative alternating–dissociative fixed; $BF_{10} > 6.096e+^{21}$) were also extremely high. The evidence for a SC in the associative condition was substantially weaker, yet still favoring the difference (associative alternating vs. associative fixed; $BF_{10} = 38.47$). Taken together, the frequentist and Bayesian analyses converge on the expected pattern of differences among the retrieval conditions, which warrants testing the main hypotheses.

## *Correlated Factor Model (CFM)*

A CFA model was specified to include four latent factors (*category, letter, associative*, and *dissociative*)[1] each allowed to covary with each other (i.e., unconstrained CFM; $CFM_{unc}$) and loading onto the four corresponding observed measures (see Figure 2A). The analysis revealed that the full $CFM_{unc}$ yielded appropriate fit with the data (see Table 1; fit indices for measurement models are reported in Table 3S in the Supplemental materials). For $CFM_{unc}$, all covariances and regression weights were significant (all *p*s $< .001$; for more details see Figure 2A and the Supplemental materials file). Importantly, as expected, the model showed that the covariance between *associative* and *category* factor was stronger than the covariance between *dissociative* and *category* factor, whereas the covariance between *associative* and *letter* factor was weaker than that between *dissociative* and *letter* factor.

Regarding the first hypothesis, these parametric differences were tested by comparing the fit of $CFM_{unc}$ against two additional models that included constraints on the covariation structure. The first model ($CFM_{cat}$) was

**Table 1.** Summary of the Structural Equation Modeling.

| Model | $\chi^2$ (*df*) | *p* | $\chi^2/df$ | CFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|
| $CFM_{unc}$ | 48.99 (98) | >.999 | .500 | 1.000 | 1.000 | <.001 | .052 |
| $CFM_{cat}$ | 62.02 (99) | >.999 | .626 | 1.000 | 1.000 | <.001 | .060 |
| $CFM_{let}$ | 53.32 (99) | >.999 | .539 | 1.000 | 1.000 | <.001 | .056 |
| General Factor | 168.92 (104) | <.001 | 1.624 | .945 | .937 | .063 | .104 |
| Bifactor model | 62.54 (88) | .982 | .711 | 1.000 | 1.000 | <.001 | .062 |

*Note.* CFI = Comparative Fit Index; CFM = Correlated Factor Model; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residuals; TLI = Tucker Lewis Index.

constrained by setting the $cov_{ASC\longleftrightarrow CAT}$ and $cov_{DSC\longleftrightarrow CAT}$ parameters to be equal, whereas the second model ($CFM_{let}$) was constrained by setting the $cov_{ASC\longleftrightarrow LET}$ and $cov_{DSC\longleftrightarrow LET}$ parameters to be equal. The resulting $\chi^2$ test indicated that the fit for both constrained models was significantly worse compared with the $CFM_{unc}$ ($\Delta\chi^2 = 13.027$, $df = 1$, $p < .001$ for $CFM_{cat}$ and $\Delta\chi^2 = 4.328$, $df = 1$, $p = .037$ for $CFM_{let}$), which favors the expected parametric differences.

Furthermore, individuals' latent retrieval scores were estimated from $CFM_{unc}$ to test the pair-wise differences between their correlation coefficients. The results showed that *category* retrieval score was significantly more strongly related to the *associative* than the *dissociative* retrieval score ($\Delta r = .310$, $z = 7.068$, $p_{holm} < .001$), whereas the *letter* retrieval score was more strongly related to the *dissociative* than the *associative* retrieval score, ($\Delta r = .147$, $z = 2.684$, $p_{holm} = .022$; for more comparisons and details see Figure 2B), which is in line with the findings from the SEM above and the hypothesis.

## Bifactor Model

Regarding the second hypothesis, we tested whether *dissociative* and *letter* retrieval involves specific processes beyond those shared across all four retrieval tasks. To this aim, we first specified a simple CFA model including a single (general) factor, which loads onto all observed measures across all four retrieval conditions. Compared with other models, the general factor model showed the worst fit (see Table 1), suggesting that one factor does not adequately account for the differences among the retrieval tasks. Following that, we estimated a bifactor CFA model, where the observed measures were set to load onto a general factor and four orthogonal (i.e., mutually uncorrelated) specific factors, one for each retrieval task (see Figure 3A), which showed an appropriate fit for the observed data. Importantly, only the general and the dissociative-specific factor showed significant variances and consistently significant regression loadings (all $p < .005$), whereas the variances for letter, category, and associative specific factors were not significant ($p = .066$, $p = .174$, and $p = .214$, respectively) and their regression loadings were weaker and less consistent (for more details see Figure 3A and the Supplemental materials). Taken together, when accounting for the general factor (i.e., the variance shared across all retrieval tasks), only the *dissociative-specific* factor showed a consistent explanatory value, whereas the significance for letter-specific factor was marginal.

## Exploratory Analyses

Next, we calculated IC and SC for each chain and participant (see Figure 1B and the Supplemental materials) and used these scores in an exploratory CFA including the two respective latent factors (IC and SC), each indicated by four observed measures. The model showed a good approximation of the data, $\chi^2 = 25.142$, $df = 19$, $p = .156$, CFI = .974, TLI = .961, RMSEA = .047, except for SRMR = .081, which was slightly above the standard criterion. Furthermore, all regression weights were significant ($p < .001$ and $p < .05$ for IC and SC, respectively). Importantly, the covariation between IC and SC was negative, moderately strong, and significant, $cov_{IC\longleftrightarrow SC} = -.592$, $p < .001$ (see the Supplemental materials for more details). Notably, the model explained considerable variance among the indicators (mean $R^2 = .313$), however, the explanatory power was stronger for IC (mean $R^2 = .453$) than SC factor (mean $R^2 = .172$). Finally, participants' IC and SC latent scores were estimated from the model and assessed for their association with the verbal fluency latent scores and their contrast (i.e., the "cost" associated with letter relative to category retrieval). The analysis showed that the IC correlated positively with the *letter* (Pearson's $r = .265$, $p_{holm} = .007$) but not with CF factor score ($r = -.044$, $p_{holm} = .604$), whereas SC correlated negatively with the *letter* ($r = -.211$, $p_{holm} = .048$) but not with CF factor score ($r = -.089$, $p_{holm} = .584$). In addition, while higher IC predicted higher costs associated with letter compared with category retrieval ($r = .380$, $p_{holm} < .001$), SC did not show a significant relationship with this contrast ($r = -.130$, $p_{holm} = .376$).

# Discussion

Lexical-semantic memory retrieval represents a core human cognitive function whose impairments manifest throughout

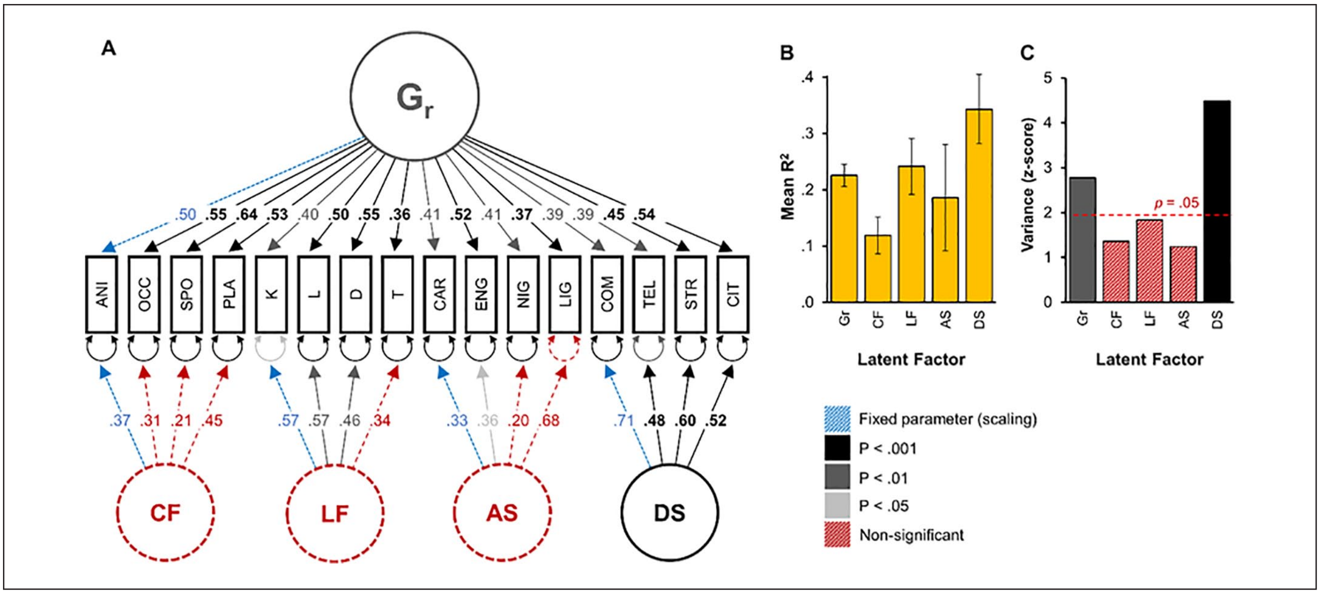


**Figure 3.** Bifactor Model.

*Note.* Panel A: Standardized parameter estimates (factor variances, regression loads, and error terms) with color-coded significance levels (bottom right). Note that only the $G_r$ and DCS yielded consistently significant loadings and variance terms. Panel B: Mean explained variance (mean $R^2$) for the indicators by latent factors. Error bars represent $\pm SE$. Note that DCS explained the most variance of the respective indicators. Panel C: Estimated z-scores for the variance parameter of the latent factors. Latent factors: $G_r$ = general; CF = category; LF = letter; AS = associative; DS = dissociative.

several pathologies and neuropsychiatric conditions (Henry et al., 2004; Raucher-Chéné et al., 2017). Hence, a major effort should focus on the understanding of the available methods as well as developing new tools capable of identifying diagnostically relevant sources of such impairments. Category and letter verbal fluency tasks have been employed to discern retrieval impairments rooted in perturbations of lexical-semantic memory stores from those stemming from hindered executive capacities that regulate access to lexical-semantic knowledge (Birn et al., 2010; Troyer et al., 1997). Although experimental, patient, or lesion-based studies (Biesbroek et al., 2016; Rende et al., 2002; Tröster et al., 1998) have suggested specific neural and cognitive differences between the verbal fluency forms, their multifactorial nature warrants an explicit evaluation of these differences using psychometric methods, which is currently lacking. For this purpose, we subjected multiple standard measures of CF and LF together with novel associative-dissociative measures (ACT) to latent structural modeling. By these means, we first assessed the correlation structure of the tasks, hypothesizing both a degree of "unity" (shared variance), reflecting the common abilities employed in all tasks, as well as a degree of "diversity" (unique variance) concerning the relative engagement of automatic (associative) and controlled (dissociative) processes in the category versus letter retrieval (hypothesis 1). Moreover, we investigated whether letter retrieval and dissociative retrieval exert specific, and perhaps similar, demands on retrieval control (e.g., inhibition and/or response selection), manifested in retrieval-specific factors complementing the general factor (hypothesis 2 and follow-up exploratory analyses). Drawing upon such analyses, the present study evaluated the convergent and discriminant validity of the fluency and ACT measures, aiming to identify common and diverse processes they engage.

## *Commonalities and Differences Among the Retrieval Tasks*

In line with previous reports (Katzev et al., 2013; Rende et al., 2002), we found that the retrieval fluency constrained by phonologic (orthographic) criteria was substantially slower than that following semantic criteria (Figure 1A). Similar increments in retrieval latency were also present when individuals retrieved unrelated versus related words in ACT, which was further prolonged under the switching condition (Figure 1B). These findings therefore support the assumed processing costs in the letter and the dissociative retrieval condition. However, a deeper insight into the putative sources of these costs comes from the CFM (Figure 2). As expected, the model indicated that all four latent retrieval scores were correlated, that is, showing a moderate level of "unity" among them. This finding conforms the view that a set of common domain-general (e.g., processing speed, self-monitoring) and verbal abilities (e.g., knowledge, mental lexicon) underpins diverse forms of retrieval (Henry & Phillips, 2006), irrespective of the involved search process and other (specific) control demands. More importantly, the model showed that the category retrieval is more strongly coupled with the free-associative than the dissociative condition, whereas the letter retrieval yielded the opposite pattern. Such a pattern of parametric coupling and dissociation is thus consistent with the hypothesized differences regarding the engagement of automatic (free-associative) versus executively demanding (dissociative) processes between the fluency tasks. Notably, the constrained models (i.e., assuming equal covariance parameters) provided significantly worse fit to the data (see Table 1), further validating that the retrieval scores reflect partially distinct cognitive processes. In particular, the parametric coupling between the letter and the dissociative retrieval suggests that difficulty of these tasks (i.e., additional "costs") may stem from the need to suppress the natural mode of retrieving words following the established semantic structures (categories and associations), indicating their reliance on retrieval control. More generally, our results indicate that lexical-semantic retrieval likely involves a common mechanism, yet the profile of automatic and controlled processes involved vary substantially depending on the nature of the tasks (i.e., specific demands on the search and executive regulation of the stimulus-driven activations within lexical-semantic representations). Here, we conclude that inhibitory processes may represent a principal component (or dimension) that underpins the ability to control semantic memory retrieval when habitual responses are not appropriate. Thus, in line with the previous evidence (Baldo et al., 2006; Collette et al., 2009; Marko & Riečanský, 2021a), the relative involvement and efficiency of inhibitory control during retrieval may account for the prolonged retrieval latency as well as the specific activations within frontal brain areas in both the LF and dissociative retrieval tasks.

Important in this regard are also findings from the bifactor model, which indicated the inadequacy of one general factor ($G_r$) to reliably account for the variance within all retrieval measures (see Figure 3 and Table 1). Instead, we found that the dissociative (and marginally the letter-cued) specific factors remained significant, explaining the most variance (above the general factor), whereas the category and the associative specific factors did not yield significant variances, showed unreliable loadings, and weaker explanatory power. Corroborating the previous models, these results indicate that dissociative retrieval and, to a lesser extent, also LF involve specific processes beyond the general retrieval ability. Finally, the IC and the SC from ACT (see Figure 1) were distinctively correlated with LF, but not CF. This further supports the hypothesis that CF and LF reflect partially diverging modes of lexical-semantic retrieval: while the former follows the habitual structure of semantic representation and therefore is more semantically driven and automatic, the latter incorporates the need to disentangle from the semantic associates and hence is more executively driven and demanding.

Together, these findings coincide with the proposed neuropsychologic dissociation between CL and LF tasks (Baldo et al., 2010; Schmidt et al., 2019), that is, that the tasks are sensitive to damage of the temporal versus frontal lobes, respectively. However, a critical consideration stemming from our data is that this partitioning is not clear-cut, as proposed previously, but parametric in nature. Specifically, both fluency tasks likely employ the left-lateralized frontal-temporal brain network, yet differentially engaging temporal versus prefrontal brain regions. In this sense, poor performance on CF might predominantly indicate deteriorated associative representational system (e.g., in Alzheimer's disease or semantic dementia; Henry et al., 2004) underpinned by the temporal cortex, whereas poor letter-cued retrieval more finely exposes executive/frontal dysfunction (e.g., in attention-deficit hyperactivity disorder [ADHD]; Andreou & Trott, 2013) pertaining to impeded ability to inhibit associative intrusions and adapt to phonological criteria (Collette et al., 2001; Marko & Riečanský, 2021a). The increased need for this prefrontal control was apparent from the coupling of LF with the dissociative retrieval as well as with the ACT control measures (IC and SC) assumed to engage the executive brain network. Nevertheless, it is important to note that the prefrontal control may be engaged also in category retrieval when retrieving less typical exemplars from poorly organized semantic categories (Demetriou & Holtzer, 2017; Mayr & Kliegl, 2000). In those cases, working memory and interference control may support retrieving and selecting appropriate items (Badre & Wagner, 2007; Hirshorn & Thompson-Schill, 2006; Katzev et al., 2013).

## Further Methodological Considerations

To promote the theory and assessment of lexical-semantic retrieval, the specific features and utility of the novel measures from ACT should be contrasted against the typical fluency tasks. First, the performance in verbal fluency tasks substantially depends on the stimuli (e.g., categories), which, as mentioned above, may exert distinct cognitive demands. For instance, categories differ in their size, exemplar availability, and other features (living vs. inanimate), which affects the retrieval difficulty (Katzev et al., 2013; Mayr, 2002). Moreover, since a category includes a relatively narrow and finite set of possible responses, the retrieval demands tend to increase as the category becomes exhausted. Thus, although category-cued retrieval is quite effortless when using standard stimuli (i.e., "animals") and shorter task duration (i.e., up to 60 seconds), under some circumstances the score may reflect different (e.g., automatic vs. controlled) processes depending on the availability of category exemplars. In particular, when retrieving from poorly organized semantic categories, the involvement of inhibition may increase due to interference from among competing concepts or intrusive repetitions (for more details, see Michalko et al., 2022). Notably, such differences may undermine the comparability among different categories when carrying out repeated measurements. The associative retrieval in ACT does not considerably involve some of these features and limits. Instead, associative chains are (theoretically) infinite as they fixate only the initial word, enabling the participants to freely explore (and manifest) their associative structures following the path of "least resistance" during the whole trial (i.e., reflecting a spontaneous, effortless, and unconstrained mode of retrieval). Although the first word, indeed, may influence a few subsequent responses, the associative responses are far more unique than responses within a specific category, rendering associative chains more suitable for repeated measurement. Furthermore, while the responses from CF follow primarily taxonomic relations, the links between associative responses are more diverse (e.g., involving also thematic or ad hoc relations; Mirman et al., 2017).

Another important part of ACT is the dissociative production. We have argued that both the letter and the dissociative conditions require suppressing the natural way of retrieving words according to their meaning. Nevertheless, when attempting to assess such specific demands on inhibition, ACT may provide a more straightforward and interpretable behavioral index (i.e., the IC) than verbal fluency tasks (i.e., using the letter-cued task alone or the difference between letter and category retrieval score, where the former involves phonological and articulatory mechanisms, and the latter is mainly semantic). Moreover, ACT also enables for calculating the SC, a measure related to retrieval (in)flexibility, whose assessment via verbal fluency tasks is questionable (e.g., Abbott et al., 2015; Hills et al., 2015).

Finally, given that verbal fluency tasks are finite (and quite narrow, e.g., for "liquids" or "tools"), participants are usually able to generate only a limited number of category exemplars. This is not the case for ACT, which allows researchers to constrain the production not only by time (as in verbal fluency tasks) but also by the number of responses required and thus tailoring the reliability of the scores. Taken together, this methodological commentary reflects on the noteworthy differences between the verbal fluency tasks and ACT, highlighting that both paradigms involve interesting features that should be carefully considered in the light of specific research/design demands.

## *Limits and Future Directions*

Several limitations that could be the subject of future research work need to be mentioned. First, the correlational design of this study constrains the interpretation of the shared and unique underlying processes driving the performance in the respective tasks. Therefore, despite some convincing suggestions of the current data about distinct embedment of CF and LF in associative/habitual versus executive/controlled processes, future experimental studies may seek to elucidate the precise nature of these processes and how they contribute to the performance on these tasks. Such studies could employ dual-task paradigms to impose distinct forms of loads or interference (associative or executive) while performing on verbal fluency and ACT tasks.

A related limitation concerns the nature of the inhibition as measured in the dissociative tasks. From the current behavioral outcome, it is not clear whether the increased dissociative RT, as compared with the associative RT, reflects a *proactive* inhibitory (or interference control) mechanism, which prevents prepotent responses entering the mind/working memory, a *retroactive* inhibition, which is employed to suppress the inappropriate responses that have entered the mind and occupy the limited slots in working memory, or both (for more discussion see Marko & Riečanský, 2021a). Future studies employing ACT should consider assessing the rate of intrusions to better understand the nature of inhibitory mechanisms involved in the dissociative condition.

Another limitation stems from fixed administration order of the retrieval tasks. Since the assessment of lexical-semantic retrieval lasted no more than 25 minutes (including short breaks in-between the tasks/trials) in each session, we assume that an effect of progressively increasing fatigue was negligible. Furthermore, the word stimuli in the ACT were selected so that they did not belong to the categories used in the preceding fluency task, to minimize potential carry over effects.

Our sample size was reasonable to test the main hypotheses but the power to rigorously evaluate the misfit of the specific measurement models for each latent variable was likely suboptimal (see the Supplemental materials). Although the evaluation of such measurement models was not in the focus of the present study, this limitation may warrant further psychometric investigation.

Finally, the knowledge about the neural correlates of associative and dissociative retrieval (or the derived retrieval control measures) in ACT is still very limited and awaits further investigation (but see Collette et al., 2001; Marko, Cimrová, & Riečanský, 2019; Marko & Riečanský, 2021b; Marron et al., 2018). So far, it seems that the free-associative retrieval does not considerably engage lateral prefrontal-parietal cortical regions (the so-called semantic control network), which are active while producing semantically dissociated responses. Nevertheless, future neurophysiological or patient studies should aim to identify the brain circuits and pathways engaged in the executively demanding dissociative retrieval in ACT and whether these correspond to those reported in the studies of inhibitory control in lexical-semantic retrieval, specifically in LF task.

Despite these limitations, our data present a reasonable picture of parametric differences between category- and letter-cued retrieval and how these differences relate to associatively and executively driven semantic processing, providing a prospective look on the methodological alternative which can be used to study lexical-semantic retrieval.

## Conclusion

Using a novel lexical-semantic retrieval paradigm, we conducted a thorough psychometric testing of the proposed dissociation between CF and LF. Our results confirmed that CF is rooted mainly in free-associative processing, while LF demands greater exertion of executive capacities possibly acting to suppress habitual but inappropriate semantic activations. Importantly, however, the current study shows that this dissociation is partial rather than absolute, thus suggesting a considerable overlap between the cognitive and neural resources involved in these tasks. To our knowledge this is the first study that explicitly addressed and statistically demonstrated this parametric dissociation in a large sample of healthy adults using both correlated latent factor and bifactor modeling. Finally, our findings and methodological commentary highlight that ACT might be used as a viable and advantageous alternative or complementary measure to study associative and executive processes in lexical-semantic memory retrieval. Moreover, the evidence for a higher involvement of retrieval control in LF, compared with CF, is important for the practical application of the tasks in neuropsychological assessment targeting the functions of lexical-semantic control versus storage, respectively.

### Author contribution

**MM:** Conceptualization, methodology, software, data curation, formal analysis, visualization, writing—original draft preparation, writing—reviewing and editing, funding acquisition; **DM:** Data curation, visualization, writing—original draft preparation, writing—reviewing and editing; **JD:** Investigation, project administration, supervision, resources; **ZV:** Investigation, project administration, funding acquisition; **DJ:** Investigation; **IR:** Conceptualization, writing—reviewing and editing, supervision, funding acquisition.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic (project no. VEGA 2/0059/20), the Slovak Research and Development Agency (project no. APVV-19-0570), and the Slovak Psychiatric Association (research grant 02/2018, "Genetic variability of BDNF, mental reactivity and memory").

### ORCID iDs

Martin Marko https://orcid.org/0000-0003-1473-1616
Igor Riečanský https://orcid.org/0000-0002-4211-8993

### Ethics approval

Research was conducted in accordance with the Declaration of Helsinki and approved by the institutional review board. All participants gave written informed consent and received a financial reward for their participation. All procedures were carried out in accordance with the relevant guidelines and regulations.

### Transparency and open practices

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. The data and additional materials are also openly available in OSF at https://osf.io/vczy7/. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

### Note

1. Only the fixed-rule ACT measures were included in this model given that the alternating condition incorporates additional distinct processes related to switching, which significantly affected both retrieval conditions (see Figure 1)

### Supplemental Material

Supplemental material for this article is available online.

### References

Abbott, J. T., Austerweil, J. L., Griffiths, T. L., Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558–569.

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323–336. https://doi.org/10.1177/107319110100800308

Allen, P., Mechelli, A., Stephan, K. E., Day, F., Dalton, J., Williams, S., & McGuire, P. K. (2008). Fronto-temporal interactions during overt verbal initiation and suppression. *Journal of Cognitive Neuroscience*, *20*(9), 1656–1669. https://doi.org/10.1162/jocn.2008.20107

Andreou, G., & Trott, K. (2013). Verbal fluency in adults diagnosed with attention-deficit hyperactivity disorder (ADHD) in childhood. *ADHD Attention Deficit and Hyperactivity Disorders*, *5*(4), 343–351. https://doi.org/10.1007/s12402-013-0112-z

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883–2901. https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Baldo, J. V., Schwartz, S., Wilkins, D. P., & Dronkers, N. F. (2006). Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. *Journal of the International Neuropsychological Society: JINS*, *12*(6), 896–900. https://doi.org/10.1017/S1355617706061078

Baldo, J. V., Schwartz, S., Wilkins, D. P., & Dronkers, N. F. (2010). Double dissociation of letter and category fluency following left frontal and temporal lobe lesions. *Aphasiology*, *24*(12), 1593–1604. https://doi.org/10.1080/02687038.2010.489260

Biesbroek, J. M., van Zandvoort, M. J. E., Kappelle, L. J., Velthuis, B. K., Biessels, G. J., & Postma, A. (2016). Shared and distinct anatomical correlates of semantic and phonemic fluency revealed by lesion-symptom mapping in patients with ischemic stroke. *Brain Structure and Function*, *221*(4), 2123–2134. https://doi.org/10.1007/s00429-015-1033-8

Birn, R. M., Kenworthy, L., Case, L., Caravella, R., Jones, T. B., Bandettini, P. A., & Martin, A. (2010). Neural systems supporting lexical search guided by letter and semantic category cues: A self-paced overt response fMRI study of verbal fluency. *NeuroImage*, *49*(1), 1099–1107. https://doi.org/10.1016/j.neuroimage.2009.07.036

Chylová, M., Marko, M., Dragašek, J., Virčík, M., Rovný, R., Roháriková, V., Murínová, J., Cimrová, B., Katina, S., & Riečanský, I. (2017). Slovak adaptation of the Schizotypal Personality Questionnaire. *Československá Psychologie*, *61*(3), 267–280.

Collette, F., Germain, S., Hogge, M., & Van der Linden, M. (2009). Inhibitory control of memory in normal ageing: Dissociation between impaired intentional and preserved unintentional processes. *Memory*, *17*(1), 104–122. https://doi.org/10.1080/09658210802574146

Collette, F., Van der Linden, M., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2001). The functional anatomy of inhibition processes investigated with the Hayling task. *NeuroImage*, *14*(2), 258–267. https://doi.org/10.1006/nimg.2001.0846

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Demetriou, E., & Holtzer, R. (2017). Mild cognitive impairments moderate the effect of time on verbal fluency performance. *Journal of the International Neuropsychological Society*, *23*(1), 44–55. https://doi.org/10.1017/S1355617716000825

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dunn, K. J., & McCray, G. (2020). The place of the bifactor model in confirmatory factor analysis investigations into construct dimensionality in language testing. *Frontiers in Psychology*, *11*, 1357. https://doi.org/10.3389/fpsyg.2020.01357

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, *86*, 186–204. https://doi.org/10.1006/cogp.1999.0734

Gourovitch, M. L., Kirkby, B. S., Goldberg, T. E., Weinberger, D. R., Gold, J. M., Esposito, G., Van Horn, J. D., & Berman, K. F. (2000). A comparison of rCBF patterns during letter and semantic fluency. *Neuropsychology*, *14*(3), 353–360. https://doi.org/10.1037/0894-4105.14.3.353

Gray, K., Anderson, S., Chen, E. E., Kelly, J. M., Christian, M. S., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). "Forward flow": A new measure to quantify free thought and predict creativity. *American Psychologist*, *74*(5), 539–554. https://doi.org/10.1037/amp0000391

Grogan, A., Green, D. W., Ali, N., Crinion, J. T., & Price, C. J. (2009). Structural correlates of semantic and phonemic fluency ability in first and second languages. *Cerebral Cortex*, *19*(11), 2690–2698. https://doi.org/10.1093/cercor/bhp023

Heim, S., Eickhoff, S. B., & Amunts, K. (2009). Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *NeuroImage*, *48*(3), 616–624. https://doi.org/10.1016/j.neuroimage.2009.06.044

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, *42*(9), 1212–1222. https://doi.org/10.1016/j.neuropsychologia.2004.02.001

Henry, J. D., & Phillips, L. (2006). Covariates of production and perseveration on tests of phonemic, semantic and alternating fluency in normal aging. *Aging, Neuropsychology, and Cognition*, *13*(3–4), 529–551. https://doi.org/10.1080/138255890969537

Heretik, A., Ritomský, A., Novotný, V., Heretik, A., & Pečeňák, J. (2009). Restandardization of State-Trait Anxiety Inventory X-2–Trait-anxiety. *Československá Psychologie: Časopis pro Psychologickou Teorii a Praxi*, *53*(6), 587–599.

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in Cognitive Science*, *7*(3), 513–534. https://doi.org/10.1111/tops.12151

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547–2557. https://doi.org/10.1016/j.neuropsychologia.2006.03.035

Hittner, J. B., May, K., & Silver, N. C. (2003). A Monte Carlo evaluation of tests for comparing dependent correlations. *The Journal of General Psychology*, *130*(2), 149–168. https://doi.org/10.1080/00221300309601282

Hyafil, A., Summerfield, C., & Koechlin, E. (2009). Two mechanisms for task switching in the prefrontal cortex. *Journal of Neuroscience*, *29*(16), 5135–5142. https://doi.org/10.1523/JNEUROSCI.2828-08.2009

Iudicello, J. E., Ph, D., Kellogg, E. J., Weber, E., Grant, I., Drane, D. L., Ph, D., Cn, A., Woods, S. P., & Psy, D. (2012). Semantic cueing improves category verbal fluency in persons living with HIV infection. *Journal of Neuropsychiatry and Clinical Neuroscience*, *24*(2), 183–190.

JASP Team. (2020). JASP (version 0.14.1) [Computer software]. https://jasp- stats. org/

Jobst, L. J., Bader, M., & Moshagen, M. (2021). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*, *23*(1), 54–60. https://doi.org/10.1037/met0000423

Katzev, M., Tüscher, O., Hennig, J., Weiller, C., & Kaller, C. P. (2013). Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: The crucial role of task demands and individual ability. *Journal of Neuroscience*, *33*(18), 7837–7845. https://doi.org/10.1523/JNEUROSCI.3147-12.2013

Marko, M., Cimrová, B., & Riečanský, I. (2019). Neural theta oscillations support semantic memory retrieval. *Scientific Reports*, *9*, 17667. https://doi.org/10.1038/s41598-019-53813-y

Marko, M., Michalko, D., & Riečanský, I. (2019). Remote associates test: An empirical proof of concept. *Behavior Research Methods*, *51*(6), 2700–2711. https://doi.org/10.3758/s13428-018-1131-7

Marko, M., & Riečanský, I. (2021a). The left prefrontal cortex supports inhibitory processing during semantic memory retrieval. *Cortex*, *134*, 296–306. https://doi.org/10.1016/j.cortex.2020.11.001

Marko, M., & Riečanský, I. (2021b). The structure of semantic representation shapes controlled semantic retrieval. *Memory*, *29*(4), 538–546. https://doi.org/10.1080/09658211.2021.1906905

Marron, T. R., Berant, E., Axelrod, V., & Faust, M. (2020). Spontaneous cognition and its relationship to human creativity: A functional connectivity study involving a chain free association task. *NeuroImage*, *220*, 117064. https://doi.org/10.1016/j.neuroimage.2020.117064

Marron, T. R., Lerner, Y., Berant, E., Kinreich, S., Shapira-Lichter, I., Hendler, T., & Faust, M. (2018). Chain free association, creativity, and the default mode network. *Neuropsychologia*, *118*, 40–58. https://doi.org/10.1016/j.neuropsychologia.2018.03.018

Martin, A., Wiggs, C. L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsychologia*, *32*(12), 1487–1494. https://doi.org/10.1016/0028-3932(94)90120-1

Mayr, U. (2002). On the dissociation between clustering and switching in verbal fluency: comment on {Troyer}, {Moscovitch}, {Winocur}, {Alexander} and {Stuss}. *Neuropsychologia*, *40*(5), 562–566.

Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go? *Psychology and Aging*, *15*(1), 29–43. https://doi.org/10.1037/0882-7974.15.1.29

Michalko, D., Marko, M., & Riečanský, I. (2022). Executive functioning moderates the decline of retrieval fluency in time. *Psychological Research*. Advance online publication. https://doi.org/10.1007/s00426-022-01680-0

Mîndrilă, D. (2010). Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal for Digital Society*, *1*(1), 60–66. https://doi.org/10.20533/ijds.2040.2570.2010.0010

Mirman, D., Landrigan, J.-F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092

Monsch, A. U., Bondi, M. W., Butters, N., Paulsen, J. S., Salmon, D. P., Brugger, P., & Swenson, M. R. (1994). A comparison of category and letter fluency in Alzheimer's disease and Huntington's disease. *Neuropsychology*, *8*(1), 25–30. https://doi.org/10.1037/0894-4105.8.1.25

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia*, *12*(3), 323–330. https://doi.org/10.1016/0028-3932(74)90047-5

Raine, A. (1991). The SPQ: A scale for the assessment of schizotypal personality based on *DSM-III*-R criteria. *Schizophrenia Bulletin*, *17*(4), 555–564. https://doi.org/10.1093/schbul/17.4.555

Raucher-Chéné, D., Achim, A. M., Kaladjian, A., & Besche-Richard, C. (2017). Verbal fluency in bipolar disorders: A systematic review and meta-analysis. *Journal of Affective Disorders*, *207*(1), 359–366. https://doi.org/10.1016/j.jad.2016.09.039

Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology*, *16*(3), 309–321. https://doi.org/10.1037/0894-4105.16.3.309

RStudio Team. (2021). *RStudio: Integrated Development for R*. RStudio, PBC.

Schmidt, C. S. M., Nitschke, K., Bormann, T., Römer, P., Kümmerer, D., Martin, M., Umarova, R. M., Leonhart, R., Egger, K., Dressing, A., Musso, M., Willmes, K., Weiller, C., & Kaller, C. P. (2019). Dissociating frontal and temporal correlates of phonological and semantic fluency in a large sample of left hemisphere stroke patients. *NeuroImage: Clinical*, *23*, 101840. https://doi.org/10.1016/j.nicl.2019.101840

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

Silver, N. C., Hittner, J. B., & May, K. (2004). Testing dependent correlations with nonoverlapping variables: A Monte Carlo simulation. *The Journal of Experimental Education*, *73*(1), 53–69. https://doi.org/10.3200/JEXE.71.1.53-70

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

Tröster, A. I., Fields, J. A., Testa, J. A., Paul, R. H., Blanco, C. R., Hames, K. A., Salmon, D. P., & Beatty, W. W. (1998). Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. *Neuropsychologia*, *36*(4), 295–304. https://doi.org/10.1016/S0028-3932(97)00153-X

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.

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499–504. https://doi.org/10.1016/S0028-3932(97)00152-8

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: The role of controlled search. *Memory & Cognition*, *41*(2), 242–254. https://doi.org/10.3758/s13421-012-0261-x

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813–826. https://doi.org/10.3758/s13423-020-01798-5

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

World Medical Association. (2013). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *JAMA: Journal of the American Medical Association*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). SNAFU: The semantic network and fluency utility. *Behavior Research Methods*, *52*(4), 1681–1699. https://doi.org/10.3758/s13428-019-01343-w

Zhao, Q., Guo, Q., & Hong, Z. (2013). Clustering and switching during a semantic verbal fluency test contribute to differential diagnosis of cognitive impairment. *Neuroscience Bulletin*, *29*(1), 75–82. https://doi.org/10.1007/s12264-013-1301-7