UvA-DARE (Digital Academic Repository)

Modulation of the word frequency effect in recognition memory after an unrelated lexical decision task

Neville, D.A.; Raaijmakers, J.G.W.; van Maanen, L.

DOI

[10.1016/j.jml.2019.05.004](https://doi.org/10.1016/j.jml.2019.05.004)

Publication date

2019

Document Version

Final published version

Published in

Journal of Memory and Language

License

Article 25fa Dutch Copyright Act

[Link to publication](#)

Citation for published version (APA):

Neville, D. A., Raaijmakers, J. G. W., & van Maanen, L. (2019). Modulation of the word frequency effect in recognition memory after an unrelated lexical decision task. *Journal of Memory and Language*, 108, [104026]. <https://doi.org/10.1016/j.jml.2019.05.004>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Modulation of the word frequency effect in recognition memory after an unrelated lexical decision task

David A. Neville^{a,*}, Jeroen G.W. Raaijmakers^{b,c}, Leendert van Maanen^{b,c}

^a Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, the Netherlands

^b Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

^c ABC, Amsterdam Brain and Cognition Center, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:

Word frequency effect
Memory
Recognition
Lexical decision
Linear ballistic accumulator

ABSTRACT

The natural language frequency of a word is known to influence the ability to perform recognition judgments based on either semantic or episodic memory, an effect commonly known as the word frequency effect (WFE). For episodic recognition specifically, the WFE presents a mirrored pattern with higher hit rates and lower false alarm rates for low frequency words compared to high frequency words. Interestingly, the use of certain study tasks such as judgements of concreteness has been shown to reduce or even abolish the hit rate advantage of low frequency words.

In the present study we tested the hypothesis that prior exposure to an unrelated lexical decision task leads to a modulation of the WFE in a subsequent episodic recognition test. Results showed an increase in hit rates and false alarm rates for high frequency words and a decrease in false alarm rates for low-frequency words. Analysis of behaviour with the linear ballistic accumulator model furthermore revealed that this effect was driven by changes in the rate of evidence accumulation of the decision process and in the response threshold. We discuss the relevance of these findings for current theories of memory.

Introduction

The frequency with which a word occurs in natural language (normative or environmental word frequency) is known to influence recognition memory performance, a phenomenon commonly referred to as the *word frequency effect* (WFE). The WFE has been investigated in the context of both semantic (Gardner, Rothkopf, Lapan, & Lafferty, 1987) and episodic (Glanzer & Adams, 1985) recognition memory tasks. For judgments that require the retrieval of information from lexical-semantic memory (e.g. 'word' or 'non-word'), previous studies have shown that responses for high frequency words are faster and more accurate than for low frequency words (Balota & Chumbley, 1984; Forster & Davis, 1984; Grainger, 1990; Scarborough, Cortese, & Scarborough, 1977). For judgments requiring instead the retrieval of information from episodic memory (e.g. 'old' or 'new'), the opposite pattern has been observed with responses for low frequency words being faster and more accurate than those for high frequency words (Guttentag & Carroll, 1997; Malmberg, Steyvers, Stephens, & Shiffrin, 2002; Yonelinas, 2002). In the context of episodic recognition the effect produced by manipulating environmental word frequency has a mirrored pattern with higher hit rates (HRs, correctly responding 'old' to an

old word) and lower false alarm rates (FARs, incorrectly responding 'old' to a new word) for low frequency words (LF) compared to high frequency words (HF) (Glanzer & Adams, 1985), also referred to as the word frequency mirror-effect (WFME) (Balota, Burgess, Cortese, & Adams, 2002; Criss, 2010). Interestingly, depending on the nature of the study task used for the episodic recognition test, some components of WFE (e.g. hit rate advantage for low frequency words) can be attenuated or even abolished. For example, Hirshman and Arndt (1997) found that performing judgements of concreteness at study led to similar hit rates for low and high frequency words but not when performing commonness judgments. Similarly, a study by Criss and Shiffrin (2004) showed that study tasks requiring judgements of animacy, pleasantness, frequency, spelling and vowel-counting also led to similar hit rates for low and high frequency words in episodic recognition (e.g. old-new judgements).

On the basis of these findings it could be argued that the memory processes underlying behavior are sensitive and adaptive to the past processing history (i.e. a study task), becoming increasingly optimized over time for the subsequent processing of specific types of information (i.e. a first task which emphasizes semantic features could lead to a more efficient processing of semantic features in a second task). In the

* Corresponding author at: Donders Centre for Cognitive Neuroscience, Radboud University Nijmegen, Kapittelweg 29, 6525EN Nijmegen, the Netherlands.

E-mail address: dneville78@gmail.com (D.A. Neville).

present study we focused precisely on this hypothesis and tested whether presenting an unrelated lexical decision task would modulate the WFE in a subsequent episodic recognition test. On the basis of previous findings our prediction was that the unrelated lexical decision task would lead to a reduction or elimination of the hit-rate advantage for low frequency words in the following recognition test.

In order to draw inferences on the memory processes underlying behavioral performance, a formal analytical approach is needed (Anders, Riès, van Maanen, & Alario, 2015; McKoon, Ratcliff, & Dell, 1986; Ratcliff & McKoon, 1986). This argument is based on the fact that classic behavioral priming effects used to measure memory performance (i.e. repetition priming) are composite effects in the sense that they map to different components of the memory process (Forster & Davis, 1984; Ratcliff, Hockley, & McKoon, 1985). Hence a perfect empirical separation of memory processes based on pure task performance can be problematic and unreliable (McKoon et al., 1986). In this study, we made the assumption that the memory processing components can be approximated by the parameters of sequential sampling models of choice-reaction time data (Ratcliff & Smith, 2004) and investigated changes in these components within the analytical framework of the linear ballistic accumulator model (LBA, Brown & Heathcote, 2008).

The LBA model is comparable to the Diffusion Decision Model of choice reaction time data (Donkin, Brown, Heathcote, & Wagenmakers, 2011) and it has been successfully applied for the analysis of behavior in a number of behavioral tasks (e.g., Boehm, Van Maanen, Forstmann, & Van Rijn, 2014; Donkin & Van Maanen, 2014; van Maanen et al., 2011). In this model, the process of producing a behavioral response (e.g., ‘word’ during lexical decision) is described by a total of five parameters. An accumulator is assumed for each response option that integrates evidence over time in a linear and deterministic fashion (e.g., an accumulator for the ‘word’ response and one for the ‘nonword’ response for lexical decision). The rate and variability at which information is accumulated are captured by the *drift rate* (v) and the *inter-trial variability* (s) parameters that determine, respectively, the mean and standard deviation of a Gaussian distribution governing the accumulation process. Once the amount of accumulated evidence crosses a specific boundary, the response associated with that accumulator is produced. The starting point of the accumulation process on each trial is drawn from a uniform distribution bounded between 0 and the A parameter with the minimal distance over which evidence is cumulated being captured by the χ parameter. The amount of evidence needed to produce a response, the response boundary, is then defined as the sum of the *response threshold* (χ) and of the *starting point* (A) parameters. The fifth and last parameter captures the contribution of non-decision processes and thus is referred to as *non-decision time* (t_0) (see Fig. 1.A). For each set of values of the model parameters the LBA model predicts complete reaction time distributions for each possible response (e.g., correct and incorrect responses).

In this study, we used the linear ballistic accumulator model (LBA) to obtain a parametric description of the components underlying the decision making process for each task and therefore a quantification of the contribution of memory processes to behavioral performance. Already in 1978 Ratcliff proposed a formal explanation of old-new recognition judgments in the framework of the drift diffusion model (Ratcliff, 1978) and in this study we followed a similar approach using the analytical framework of the LBA model to explain changes in the memory processing components producing the WFE.

To summarize, our experimental approach relied on testing whether previously completing a lexical decision task would lead to the changes in the WFE effect in a later unrelated recognition test. To gain insight into the potential changes of the WFE, we analyzed the data from both lexical decision and recognition tasks with the linear ballistic accumulator model obtaining a parametric description of the memory processing components deployed in each task. The estimated model parameters were then tested for significant differences with standard analyses of variance (ANOVA). Finally, the behavioral results were

Fig. 1. (A) The linear ballistic accumulator model with two response accumulators. (B) Task-order conditions of the present study.

interpreted in relation to the model parameters of the LBA model and current memory theory.

Experiment

The goal of the present study was to test whether previous semantic processing on an unrelated task would affect memory performance on a later episodic recognition test and identify any changes in the associated processing components. As the study task for the recognition test we used a familiarity judgement task and different lists of words were presented between lexical decision, on the one hand, and the familiarity judgement and recognition test, on the other hand. Three lists of words were rotated across task conditions: lexical decision (word items), recognition (old items), recognition (new items). Half of participants completed first the familiarity judgement task followed by the recognition test on the items from the immediately previous familiarity judgement task and finally the (unrelated) lexical decision (LD last condition). The other half of subjects instead completed first the unrelated lexical decision task followed by the familiarity judgment task and the recognition test (LD first condition). Subjects were randomly assigned to either one of the two task sequences (Fig. 1B).

Normative word frequency was manipulated to elicit the WFE effect with a contrast between high frequency words (HF) and low frequency words (LF). The stimuli were presented during lexical decision and familiarity judgement once, twice or four times and presented only once during recognition. Since it is known that the repetition of a stimulus (presenting it multiple times) has a different impact on episodic processing than on semantic processing, the manipulation of both word frequency and presentations provided a better approach to tease apart the separate contributions of episodic and semantic processing to behavioral performance.

Methods

Subjects

36 students (mean age, 21.6 years; SD, 2.6 years; 25 females) from the University of Amsterdam participated in the experiment for course credit or monetary compensation. All participants were non-dyslexic native Dutch speakers with normal, or corrected-to-normal, vision.

Materials

A total of 360 Dutch words were selected from the CELEX (Baayen, Piepenbrock & Van Rijn, 1993) and WebCelex (Baayen, Piepenbrock, & Gulikers, 2001) databases. From the total pool three lists of words (List A, List B, List C) were constructed with each of them consisting of an equal number of high and low frequency items. For list A, the word length of high frequency items (HF) varied between 4 and 6 letters ($M = 4.78$, $SE = 0.11$) and the frequency varied between 40 and 649 per million ($M = 134.70$, $SE = 17.46$). For the low frequency items (LF) of list A, the word length varied between 4 and 6 letters ($M = 5.45$, $SE = 0.10$) and the frequency varied between 0 and 5 per million ($M = 0.3$, $SE = 0.13$). For list B, the word length of HF items varied between 4 and 6 letters ($M = 4.75$, $SE = 0.10$) and the frequency varied between 40 and 734 per million ($M = 133.70$, $SE = 17.27$). For the LF items of list B, the word length varied between 4 and 6 letters ($M = 5.52$, $SE = 0.08$) and the frequency varied between 0 and 2 per million ($M = 1.08$, $SE = 0.13$). For list C, the word length of HF items varied between 4 and 6 letters ($M = 4.93$, $SE = 0.11$) and the frequency varied between 41 and 273 per million ($M = 99.54$, $SE = 7.33$). For the LF items of list C, the word length varied between 4 and 6 letters ($M = 5.62$, $SE = 0.08$) and the frequency varied between 0 and 2 per million ($M = 0.17$, $SE = 0.06$). Words lists A, B and C can be found in the Appendix A (Table A.1). A filler list of 120 non-words was also constructed for the lexical decision with word length varying between 4 and 6 letters ($M = 4.83$, $SE = 0.07$) and non-words differing from existing Dutch words at most for one letter.

Apparatus

Stimuli were presented on a CRT monitor at a refresh rate of 100 Hz and resolution of 1024x768 using Presentation® software (Version 0.70, Neurobehavioral Systems, Inc., Berkely, CA, www.neurobs.com). Font size was 25 points of the *arial* type. Participants were seated approximately 70 cm from the screen and were tested individually.

Design

A Latin square design was used in the present experiment to ensure that each of the three word lists was used in each of three task-set conditions: lexical decision (test items), recognition (test items) and recognition (novel items). Word frequency was manipulated with two levels: high frequency and low frequency and items presentations were varied with three levels: one, two and four item presentations. Participants were assigned randomly to either one of the two lexical decision task order conditions.

Procedure

Three tasks were used in the present experiment: Lexical decision, Familiarity Judgment and Recognition. Lexical decision was always administered as an unrelated task (different list of items) whereas the familiarity judgment task was used as the study task for the following

recognition test. Items were presented (once, twice or four times) in lexical decision and familiarity judgement tasks but not in recognition. Non-words were presented only once in the lexical decision task. Participants were randomly assigned to either one of two lexical decision task-order conditions. In the first task order condition (lexical decision last) participants completed first the familiarity rating task followed by recognition test on the words presented in the previous familiarity task and novel words. Afterwards they were presented with a lexical decision task involving a completely different set of items. In the second task order condition (lexical decision first), participants completed first the lexical decision task followed by the familiarity judgement and the recognition task.

For all of the tasks, instructions were provided on screen and subsequently repeated by the experimenter. After each task, a short break of 5 min was given during which participants played the classic version of the Nintendo SUPER MARIO® videogame. In lexical decision, subjects were first presented with a fixation cross for 1000 ms, followed by a forward mask¹ for 300 ms and lastly the target item presented for 1000 ms or until a response was made. Word and Non-word responses were expressed by using the left and right 'SHIFT' keys, respectively. In recognition, the task structure was identical. The only difference was that subjects had to decide whether the presented word was old or new. Response mappings were kept consistent across tasks. For the familiarity rating the task structure and timings were similar to lexical decision and recognition with the only difference being the response modality since subjects had to evaluate how familiar the presented word was on a scale from 1 to 7. Seven equally sized response boxes were presented on screen together with the target and subjects expressed their responses by pressing twice the appropriate numerical key. After a first key press, the corresponding box would be highlighted to give participants the option of changing their answer. A second key press would confirm the highlighted answer (see Fig. 2 for details of all tasks). In all of the tasks subjects were not informed about the correctness of their responses (no feedback was provided). At the end of the experimental session subjects were given a short exit questionnaire and debriefed regarding the nature of the experiment.

LBA analysis

The analysis of behavioral judgments with the LBA model was carried out by fitting the data from the lexical decision and recognition tasks from both lexical decision task orders separately for each participant. LBA models with different combinations of fixed/free parameters across word frequency and presentations conditions were fitted to the choice reaction time data. For each model the relative Bayesian Information Criterion score (BIC, Schwarz, 1978) was computed to obtain a measure of the model's goodness of fit penalized by the complexity of the model (number of free parameters to be estimated from the data). The best fitting model was then identified by comparing the rank-ordered BIC scores of each model averaged over participants and selecting the model with the lowest BIC score.

Fitting of the LBA model to the data was done via quantile maximum probability estimation (Heathcote, Brown, & Mewhort, 2002). The quantiles were 0.1, 0.3, 0.5, 0.7 and 0.9 of the RT distributions for correct and incorrect responses expressed in milliseconds. All of the analyses were conducted in R® (R Core Team, 2018) using customized scripts. To ensure identifiability of the model parameters, we followed common practice (Donkin, Averell, Brown, & Heathcote, 2009) and constrained the drift rates for the correct and incorrect accumulators to sum to 1. Furthermore, to capture potential bias effects in both lexical decision and recognition we followed previous studies (Forstmann, Brown, Dutilh, Neumann, & Wagenmakers, 2010; van Maanen,

Fig. 2. Task structures for familiarity rating (A), lexical decision (B), and recognition tasks (B).

¹ The forward mask was presented for consistency with other tasks being tested during the same session, however it is irrelevant for the present study.

Forstmann, Keuken, Wagenmakers, & Heathcote, 2015) and estimated separate thresholds for each response accumulator.

Behavioral results

Analysis of behavioral data from linguistic experiments is often confronted with potential confounds due to individual differences between participants and between word stimuli. In linear mixed effect regression analysis (LME) these potential confounds are taken into account in the form of random effects which are entered in the analysis along with the fixed effects of the experimental manipulations (Baayen, Davidson, & Bates, 2008; Brysbaert, 2007). In the present experiment we analyzed data from all three tasks using the following LME regression model:² main effects for *Word Frequency*, *Presentations*, *Lexical decision task order*, *List*, two-way interaction effects between *Word Frequency* and *Presentations*, and *Word Frequency* and *Lexical decision task order*. The model also included random effects for both participants and stimuli and the contrasts for *Lexical decision task order* were drawn as the difference between the lexical decision first condition minus the lexical decision last condition. Out of the 36 subjects tested, one participant was detected as statistical outlier due to relatively low accuracy performance in both lexical decision and recognition tasks (accuracy < 70%) and was therefore excluded from all of the following analyses. Including this participant in the analyses produced qualitatively similar results.

Lexical decision

Mean proportions of correct word responses are reported graphically (see Fig. 3) as well as hit rates (HRs) and false alarm rates (FARs) (see Fig. 4). For the non-word items in the LD last condition the mean proportion of correct responses was 0.88 (SEM = 0.01) with a mean reaction time of 613.23 ms (SEM = 2 ms), and for the LD first condition the mean proportion of correct responses was 0.89 (SEM = 0.01) with a mean reaction time of 585.97 ms (SEM = 3 ms). Both accuracy and reaction time data were analyzed with the same LME model described above. For the accuracy data ANOVA of the model estimates indicated main effects for *word frequency* ($\beta_{\text{freq}} = -0.62$, $F_{(1, 617)} = 369.69$, $p < 0.001$), for *presentations* ($\beta_{\text{pres}} = 0.01$, $F_{(1, 7414)} = 55.94$, $p < 0.001$), an interaction of *word frequency* with *presentations* ($\beta_{\text{freq} \times \text{pres}} = 0.04$, $F_{(1, 7546)} = 38.41$, $p < 0.001$) and an interaction of *word frequency* with *lexical decision task order* ($\beta_{\text{freq} \times \text{tskord}} = 0.04$, $F_{(1, 9530)} = 10.80$, $p < 0.01$) with effect size ($R^2 = 0.48$). For the analysis of reaction time data for correct word responses results revealed main effects of *word frequency* ($\beta_{\text{freq}} = 155.92$, $F_{(1, 1423)} = 340.23$, $p < 0.001$), of *presentations* ($\beta_{\text{pres}} = -9.93$, $F_{(1, 1591)} = 228.44$, $p < 0.001$), of *lexical decision task order* ($\beta_{\text{tskord}} = 39.23$, $F_{(1, 32)} = 6.02$, $p < 0.05$), an interaction between *word frequency* and *presentations* ($\beta_{\text{wfreq} \times \text{pres}} = -16.58$, $F_{(1, 1717)} = 47.65$, $p < 0.001$) and an interaction between *lexical decision task order* and *word frequency* ($\beta_{\text{tskord} \times \text{wfreq}} = -15.94$, $F_{(1, 7101)} = 10.50$, $p < 0.01$) with effect size ($R^2 = 0.30$). Overall the results indicate that subjects were faster and more accurate in processing high frequency words compared to low frequency words and that the overall speed of processing for correct word responses increased over item presentations. More importantly, word frequency interacted with lexical decision task order and presentations indicating that previously completing the familiarity judgment and recognition tasks led to more

² For all of the LME analyses we used the lmerTest package (Kuznetsova, Brockhoff, & Bojensen Christensen, 2014) which performs standard linear mixed effect regression analyses (Baayen et al., 2008) with calculation of p-values and degrees of freedom for the denominator of the F-statistic based on the Satterthwaite approximation method (Schalje, McBride, & Fellingham, 2002).

Fig. 3. Mean proportion of correct responses (left) and reaction times (right) with normalized standard errors for lexical decision task (Loftus & Masson, 1994) in (A) lexical decision last (LD last) and (B) lexical decision first (LD first) task-order conditions.

Fig. 4. Mean hit rates (HRs) and false alarm rates (FARs) with standard errors of the mean (SEM) for lexical decision task.

accurate and faster responses in the following lexical decision, an indication of the adaption of the underlying memory processes.

Familiarity rating

Mean rating scores and standard errors were computed for each level of the item presentation manipulation for both low frequency and high frequency words (Fig. 5). ANOVA of the LME model estimates for the rating scores revealed a main effect of *presentations* ($\beta_{\text{pres}} = 0.01$, $F_{(1, 9279)} = 3.95$, $p < 0.05$) and of *word frequency* ($\beta_{\text{wfreq}} = -3.15$, $F_{(1, 613)} = 316.81$, $p < .0001$) and an interaction between *word frequency* and *lexical decision task order* ($\beta_{\text{wfreq} \times \text{taskord}} = 0.25$, $F_{(1, 9524)} = 25.43$, $p < .0001$) with effect size ($R^2 = 0.67$) indicating that participants rated low frequency items significantly lower than high frequency items and that over presentations the ratings slightly increased for low frequency items. Notably, there was also an interaction between *lexical decision task order* and *word frequency* with low frequency items being rated slightly higher when not preceded by the unrelated lexical decision task. Similar analyses ran for the reaction times revealed main effects of *presentations* ($\beta_{\text{pres}} = -25.30$, $F_{(1, 9279)} = 3.95$, $p < 0.05$) and of *word frequency* ($\beta_{\text{wfreq}} = -1.15$, $F_{(1, 613)} = 316.81$, $p < .0001$) and an interaction between *word frequency* and *lexical decision task order* ($\beta_{\text{wfreq} \times \text{taskord}} = 0.25$, $F_{(1, 9524)} = 25.43$, $p < .0001$) with effect size ($R^2 = 0.67$) indicating that participants reacted faster to high frequency items than low frequency items and that over presentations the reaction times slightly increased for low frequency items.

Fig. 5. Mean ratings (left) and reaction times (right) with normalized standard errors (Loftus & Masson, 1994) for Familiarity Judgement task in lexical decision last (A) and lexical decision first (B) task-order conditions.

1539) = 82.19, $p < 0.001$), *word frequency* ($\beta_{\text{wfreq}} = 767.12$, $F_{(1, 3071)} = 120.41$, $p < 0.001$), *lexical decision task order* ($\beta_{\text{tskord}} = 290.78$, $F_{(1, 32)} = 5.57$, $p < 0.05$), interaction effects between *presentations* and *word frequency* ($\beta_{\text{wfreq} \times \text{pres}} = -94.51$, $F_{(1, 1668)} = 35.21$, $p < 0.001$) and between *lexical decision task order* and *word frequency* ($\beta_{\text{wfreq} \times \text{tskord}} = 149.19$, $F_{(1, 9491)} = 22.29$, $p < 0.001$) with effect size ($R^2 = 0.39$). These results indicate that low frequency items were rated significantly slower than high frequency items and with an overall decrease in reaction time over presentations. There was also a large effect of lexical decision task order with reaction times being significantly slower when not preceded by lexical decision. In summary, these results indicate that the manipulation of word frequency had an effect on judgements of familiarity with participants rating high frequency words as more familiar compared to low frequency words. Furthermore, the effect of lexical decision task order indicates that processing of semantic information was faster when preceded by lexical decision suggesting also in this case a change in processing components due to the previous processing history.

Recognition

Mean proportion of correct responses (i.e., hits), standard errors and reaction times for high and low frequency items were calculated for items presented once, twice and four times (see Fig. 6) as well as mean hit rates (HRs) and false alarms rates (FARs) (see Fig. 7). For the analysis of accuracy data, we used exactly the same LME model as in the previous analyses of familiarity ratings and lexical decisions. ANOVA of the model estimates yielded a main effect of *presentations* ($\beta_{\text{pres}} = 0.05$, $F_{(1, 1356)} = 92.12$, $p < 0.001$), a main effect of *word frequency* ($\beta_{\text{wfreq}} = -0.02$, $F_{(1, 2819)} = 17.24$, $p < 0.001$) plus an interaction of *lexical decision task-order* with *word frequency* ($\beta_{\text{wfreq} \times \text{tskord}} = 0.01$, $F_{(1, 3930)} = 33.66$, $p < 0.001$) with effect size ($R^2 = 0.21$). For the analysis of reaction time data for correct word responses results indicated main

Fig. 6. Mean proportion of correct responses (left) and reaction times (right) with normalized standard errors for recognition task (Loftus & Masson, 1994) in lexical decision last (A) and lexical decision first (B) task-order conditions.

Fig. 7. Mean hit rates (HRs) and false alarm rates (FARs) with standard errors of the mean (SEM) for recognition task.

effects of *word frequency* ($\beta_{\text{wfreq}} = 120.58$, $F_{(1, 2284)} = 59.60$, $p < 0.001$) and of *presentations* ($\beta_{\text{rep}} = -5.60$, $F_{(1, 1183)} = 30.95$, $p < 0.05$), an interaction between *word frequency* and *presentations* ($\beta_{\text{wfreq} \times \text{rep}} = -8.07$, $F_{(1, 1183)} = 5.46$, $p < 0.05$) and an interaction between *word frequency* and *lexical decision task order* ($\beta_{\text{wfreq} \times \text{tskord}} = -38.64$, $F_{(1, 2850)} = 20.98$, $p < 0.001$) with effect size

Fig. 8. Differences in d' primes measures for recognition between lexical decision first and lexical decision last task order conditions.

($R^2 = 0.24$). When the recognition test was preceded only by the study phase, we found consistently with previous studies a higher accuracy performance for LF words over HF words. Similarly, for HRs and FARs, we found the classic mirrored pattern with higher HRs and lower FARs for low frequency words compared to high frequency ones. When instead a lexical decision task (on different words) was presented before the study phase of the recognition test, accuracy performance was higher for HF words compared to LF words. In terms of HRs and FARs, the introduction of the lexical decision task led in episodic recognition to higher HRs and FARs for high frequency words and lower FARs for low frequency words. Also, increasing the number of presentations led to progressively higher accuracy, lower reaction times and higher HRs indicating an overall increase in performance. To better quantify the effect of *lexical decision task order* on memory discriminability, we also calculated for each task the difference in d' prime measures (Stanislaw & Todorov, 1999) between *lexical decision task order* conditions. The d' for each task and task-order conditions was calculated by z-scoring HRs and FARs and then computing the difference as in $d' = z(\text{HR}) - z(\text{FA})$. Next, we calculated the change in d' measures for each task by computing the difference in d' measures among task-order conditions. Results yielded a large increase in recognition in d' for HF words repeated once and twice and very small effects for all other conditions (see Fig. 8 for details).

Overall the results indicate that engaging first in a lexical decision task leads to better performance for HF words over LF ones in a subsequent (unrelated) recognition task which could suggest increased emphasis (i.e. better processing) of semantic memory processes during recognition as also suggested by the increase in d' measure. Engaging first in the recognition memory test instead led to the standard advantage of LF words over HF words in recognition and to a moderate increase in accuracy for LF words with an overall decrease in reaction times for lexical decision. The present pattern of results could be explained by changes in the decision process but it could be also be attributed, in part, to a general effect of practice and/or changes in decision strategy. In order to separate changes in the decision process from task related effects due to strategy or practice we analyzed the choice reaction time data from lexical decision and recognition tasks with the linear ballistic accumulator model.

In the LBA model, changes in the decision process (e.g. quality of information) and the effect of general practice or strategies map to

different model parameters (components). The increase in HR in recognition for HF words after lexical decision could be explained in the LBA model by a change in the decision process which would map to changes in the rate of evidence accumulation (drift rate). The effects instead of a change of strategy of the effect of general practice would map to changes the response threshold parameter and/or in the starting point of the accumulation process. For lexical decision, the increase in performance observed was subtler with only a moderate decrease in reaction times, suggestive of an effect of practice which could be explained by changes in the response threshold and/or starting point parameters but not in the drift rate parameter. Similarly, the increase in accuracy performance and decrease in reaction times for recognition could be attributed to a change in the decision process but also, partially, to a change in the response strategy due to practice.

To quantitatively test whether the modulation of the HRs and FARs components of the WFE was explained by changes in the decision process or by task-set related effects, we analyzed choice-reaction time data from the recognition task and the lexical decision task with the linear ballistic accumulator model (LBA).

LBA analysis results

Comparison of BIC scores for different LBA models indicated a model with v (drift rate) free to vary across both *word frequency* and *presentations* conditions and all other parameters fixed across both *word frequency* and *presentations* conditions as the most parsimonious account of the data for each of the tasks in each of the lexical decision task order conditions (see Table A.2 in the Appendix A for details). For ease of reading mean parameter estimates of the best fitting LBA model are reported in graphical format (see Figs. 10, 11 and 12) alongside the predicted mean reaction times for each quantile interval averaged across participants (see Fig. 9).

After having estimated the parameters of the LBA model for each task, we tested for significant differences across the factors of the experimental design (*word frequency*, *presentations*, *task* and *lexical decision task order*) in the drift rate parameter by means of a repeated measures ANOVA analysis. Similarly, significant differences across tasks and task-orders factors in response threshold, starting point, noise of the accumulation process and non-decision time parameters were tested by means of regular ANOVA analyses. Since the *response threshold* parameter was separately estimated for correct and incorrect accumulators, the analysis was carried out on the difference between the estimates for both recognition and lexical decision tasks.

Drift rate (v)

Results of the ANOVA analysis showed a main effect of *word frequency* ($\beta_{\text{wfreq}} = -0.54$, $F_{(1, 416)} = 203.49$, $p < .001$), a main effect of *presentations* ($\beta_{\text{repeats}} = 0.09$, $F_{(1, 416)} = 59.57$, $p < .001$),

an interaction between *word frequency* and *task* ($\beta_{\text{wfreq} \times \text{task}} = 0.57$, $F_{(1, 416)} = 157.55$, $p < .001$), an interaction between *presentations* and *task* ($\beta_{\text{repeats} \times \text{task}} = -0.03$, $F_{(1, 416)} = 4.40$, $p < .05$) and an interaction between *word frequency* and *lexical decision task order* ($\beta_{\text{wfreq} \times \text{taskord}} = 0.02$, $F_{(1, 416)} = 9.83$, $p < .01$).

Response threshold (χ)

For the response bias results indicated a main effect of *task* ($\beta_{\text{task}} = 90.94$, $F_{(1, 416)} = 82.64$, $p < .001$), a main effect of *task-order* ($\beta_{\text{task-ord}} = -15.60$, $F_{(1, 416)} = 26.61$, $p < .001$) and an interaction between *task* and *lexical decision task order* ($\beta_{\text{task} \times \text{taskord}} = -45.94$, $F_{(1, 416)} = 9.44$, $p < .01$).

Fig. 9. Mean reaction times (x-axis) for each quantile interval (y-axis) averaged across participants for lexical decision task (A) and recognition (B) in lexical decision last and lexical decision first task order conditions.

Non-decision time (t_0)

For non-decision time, results showed only a main effect of *task* ($\beta_{\text{task}} = -40.63$, $F_{(1, 416)} = 4.10$, $p < .05$).

Noise of the accumulation process (s)

For drift rate variability, ANOVA results indicated a main effect of *task* ($\beta_{\text{task}} = -0.01$, $F_{(1, 416)} = 15.77$, $p < .001$) and a main effect of

lexical decision task order ($\beta_{\text{task-ord}} = -0.01$, $F_{(1, 416)} = 8.65$, $p < .01$).

Starting point (A)

For the starting point of the accumulation process, we also found a main effect of *task* ($\beta_{\text{task}} = 34.62$, $F_{(1, 416)} = 24.50$, $p < .001$) and a main effect of *lexical decision task order* ($\beta_{\text{task-ord}} = -20.74$, $F_{(1, 416)} = 10.60$, $p < .01$)

Previous studies have shown that changes in the *drift rate* parameter

Fig. 10. Mean estimated drift rate (v) parameter (arbitrary units) averaged across participants with standard errors of the mean (SEM) for the lexical decision task (A) and recognition (B) in lexical decision last and lexical decision first task order conditions.

Fig. 11. Mean estimated threshold (χ) parameter in arbitrary units with standard errors of the mean (SEM) averaged across participants for incorrect (dark grey) and correct (light grey) responses in lexical decision task (A) and recognition (B) in lexical decision last and lexical decision first task-order conditions.

reflect changes in the quality of information being sampled by the decision process (Criss, 2009, 2010; Criss, Wheeler, & McClelland, 2013; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Global theories of

Fig. 12. Mean estimated inter-trial variability (s), starting point (A) and non-decision time (t_0) parameters in arbitrary units with standard errors of the mean (SEM) averaged across participants in lexical decision task (A) and recognition (B) in lexical decision last and lexical decision first task-order conditions.

memory such as SAM (Search of Associative Memory) (Raaijmakers & Shiffrin, 1980) and REM (Retrieving Effectively from Memory) (Shiffrin & Steyvers, 1997) assume that information is stored in memory in the form of “memory images” (or memory traces). The retrieval of information from long-term memory is characterized as a cue-dependent reconstructive process where the cue is matched against each of the memory images. The higher the match between the cue and a memory image, the higher the probability that the memory image will be retrieved. In this theoretical framework, it follows that priming effects such as presentations or semantic priming can be explained by changes in the matching process with an increase in the feature overlap between the cue and the target memory (i.e. studied item or semantic associate) and a decrease in the feature overlap between the cue and the other (non-relevant) images present in long-term memory. A change in the quality of information can therefore be interpreted in these theories as an increase in the matching process between the cue and the memory image which in turn leads to an increase in behavioral performance.

In line with this view we found main and interaction effects of word frequency and of task on the drift rate parameter, indicating that the amount of evidence provided by the stimulus, and specifically by its normative frequency, changes as a function of the memory task. In other words, the behavioral WFE is explained by the model in terms of a change in the rate of evidence accumulation which can be interpreted as an increase in matching features between the cue and the target memory image. However, the most important result of this analysis is the interaction of lexical decision task order with drift rate which indicates a modulation of the WFE as a function of lexical decision task order. Changes in the threshold parameter have been shown by previous studies to be dependent on task-set differences and the main effect of task is in line with this view. Interestingly, we also found an effect of lexical decision task order along with an interaction with task indicating that engaging in further semantic processing leads to a change in response threshold in an unrelated subsequent episodic task, that is a change in response strategy. For the non-decision time parameter, results showed only a main effect of task indicating a difference in non-decisional processes across lexical decision and recognition tasks which could be explained by differences in the encoding processes (i.e. extraction of features from the percept). For the drift rate variability parameter, results showed a main effect of task with an interaction with

lexical decision task order indicating a decrease in the noisiness of the accumulation process after additional semantic processing. Last, for the *starting point* parameter we found a main effect of *task* and of *lexical decision task order* with no interaction suggesting that the change in the amount of starting evidence for the accumulation process elicited by the previous processing history was comparable across lexical decision and recognition tasks.

In summary, the findings of the present study show that additional (unrelated) semantic processing can lead to a modulation of the WFE in recognition. Furthermore, the modulation can be explained by changes in the underlying memory processing components and specifically by changes in the amount of evidence being accumulated (drift rate) from the stimulus features (normative word frequency). Results showed also an interesting interaction of task-order and task for the response threshold component but not for the starting point of accumulation. This further supports the interpretation that interactions between task-sets lead to selective changes in the memory components underlying the decision process.

In the general discussion section, we elaborate on the relevance of these findings for the existing literature and current theories of memory processing.

General discussion

The most important result of the present study is the modulation of the WFE in recognition following an unrelated lexical decision task. In the lexical decision last condition (i.e., when recognition was only preceded by the study phase), we found the typical pattern for the word frequency mirror effect (WFME) for episodic recognition with higher hit rates and lower false alarm rates for low frequency words compared to high frequency ones. In the lexical decision first condition instead, we observed an increase in the proportion of correct responses for the high frequency words. Specifically, we observed an increase in hit rates and false alarm rates for high frequency words as well as a reduction of false alarm rates for low-frequency words. This pattern of results indicates an increase in the probability of responding 'old' for high frequency items and could be explained by the fact that the previous completion of the lexical decision task leads to an emphasis (increased processing) on semantic processing of subsequent stimuli, which could translate to an increase in the storage of features for high frequency items during the study task. Such an interpretation is in line with the results of the LBA model on the drift rates which indicate a change in the quality of information (e.g. processing of additional features). Furthermore, the LBA analysis also revealed a decrease in response threshold which is consistent with the idea of change in response strategy with subjects adopting a more liberal criterion (less evidence needed to make a decision) after lexical decision. The adoption of a more liberal criterion that is selective for word frequency could also explain the decrease in FARs for low frequency words where fewer matching features would be expected for low frequency words and thus a lower probability of saying "old" would be predicted for novel words. Similarly, the small but significant decrease in familiarity ratings for the low frequency words after lexical decision could be a reflection of discounted (less informative) low-frequency features again resulting from increased emphasis on semantic processing and specifically on high-frequency features.

We will now discuss in turn the behavioral patterns and the results of the model-based analysis in light of existing findings and theories of memory processing.

In lexical decision, the observed behavioral pattern shows that the manipulation of normative word frequency led to a higher performance for HF words across both lexical decision task orders, which is in line with what has been reported in previous studies (Balota & Chumbley,

1984; Forster & Davis, 1984; Grainger, 1990; Scarborough et al., 1977). For episodic recognition, when not preceded by the unrelated lexical decision task, we replicated previous findings of the word frequency effect with an advantage of LF words over high frequency words (Guttentag & Carroll, 1997; Malmberg et al., 2002; Scarborough et al., 1977; Yonelinas, 2002). When instead the recognition task was preceded by lexical decision, a modulation of the WFE was obtained with an increase in hit rates and false alarm rates for HF words and a decrease in false alarm rates for LF words. We now turn to the results of the model-based analysis to discuss the processing components underlying the modulation of the WFE and the corresponding changes in behavioral performance.

A first important result revealed by the LBA analysis is the clear effect of *normative word frequency* on *drift rate* indicating as in previous studies a change in the quality of information for the decision process (Criss et al., 2013; Criss, 2009, 2010; Wagenmakers et al., 2008). A novel finding of the present study in this respect is the clear double dissociation of the *drift rate* over *normative word frequency* and *task sets*. In lexical decision, the rate of evidence accumulation (*drift rate*) for the high frequency words was higher than for low frequency words and this pattern reversed in recognition with an advantage of low frequency words over high frequency ones. This double dissociation of drift rates is consistent with previous findings indicating an advantage for processing high frequency words over low frequency words for semantic tasks (Roger Ratcliff, Gomez, & McKoon, 2004; Roger Ratcliff, Thapar, Gomez, & McKoon, 2004) and a reversed pattern for episodic recognition tasks (Roger Ratcliff, Thapar, & McKoon, 2004). Similarly, a previous study with the drift diffusion model contrasting semantic (animate vs. inanimate) and episodic (old vs. new) types of memory retrieval (experiment 1, Spaniol, Madden, & Voss, 2006) also found a dissociation of memory retrieval types in drift rates, with overall smaller drift rates for episodic retrieval. Crucially we found that the double dissociation of drift rates over retrieval types (*task sets*) changed only in episodic recognition with an increase in drift rates for high frequency words after the unrelated lexical decision.

A second interesting result of the LBA analysis is the interaction between *task-set* and *lexical decision task order* for the *response threshold* (χ) parameter which similar to other studies (Forstmann et al., 2010) might reflect changes in task related strategies. In recognition we found a lower threshold for the correct response accumulator when the critical test was preceded by lexical decision and familiarity tasks. A lower threshold indicates a more liberal decision strategy and therefore shorter response times since less information needs to be accumulated before producing a response. The decrease in response threshold for recognition in combination with the increase in drift rates for high frequency words supports the idea of emphasized semantic processing after lexical decision with more efficient processing of semantic features and therefore better recognition performance for high frequency words.

A third result less important but worth of mentioning is the difference between task sets in the non-decision time parameter which indicates that the non-decision processes (extraction and encoding of the stimulus features, motor preparation) in recognition were significantly faster than in lexical decision.

Overall the results of the model-based analysis corroborate and extend previous findings by showing that the word frequency effect is explained by changes in the rate of evidence accumulation (drift rate) and that the modulation in episodic recognition after an unrelated lexical decision can also be explained by changes in the same processing component (drift rate). The change in quality of information (evidence accumulation) can be interpreted as an increase in the storage of information (features) in memory leading ultimately to better retrieval and we will now discuss how in this perspective the observed pattern of results can be explained by models of memory which make the

following set of assumptions: (1) features are encoded in memory probabilistically with sometimes the wrong feature being encoded, (2) retrieval of information from memory occurs by selecting the memory image which has the highest number of matching features with the cue and (3) previous processing history can lead to selective changes in the semantic processing components (e.g. increased storage or diagnostic value of high-frequency features).

Current models of memory generally explain repetition-priming effects in terms of retrieval of information from long-term memory (Raaijmakers, 2005; Schooler, Shiffrin, & Raaijmakers, 2001; van Maanen, van Rijn, & Taatgen, 2012; Wagenmakers, Huber, Raaijmakers, Shiffrin, & Schooler, 2003). The SAM-REM theory in particular (for a review, see Raaijmakers, 2008) provides an appropriate framework for the present study since it has been shown in the past to account for repetition priming effects in both lexical decision (Wagenmakers et al., 2004) and episodic recognition (Shiffrin & Steyvers, 1997). In the SAM-REM theory memory traces are distinguished into semantic and episodic traces and each memory trace is assumed to contain a mixture of three types of information (features): *content* features (i.e., semantic, phonological, and orthographic information), *contextual* features (i.e., “physical, spatial/temporal, environmental, physiological, and/or emotional states in which the item was experienced”, Malmberg & Shiffrin, 2005, pp. 323) and *inter-item* features (e.g., associative information between items).

In this framework, the changes in the rate of evidence accumulation (drift rate) could be explained by a change in the *diagnostic value* of the contextual features extracted during encoding. In a standard episodic recognition test, the low frequency features extracted during encoding would yield increased diagnostic evidence at test due to fewer matches with the other existing images in memory. In lexical decision, low frequency features would predict a lower performance since the few matches with semantic memory traces would bear low diagnostic value. In episodic recognition, instead the low frequency features would predict higher performance since they would be highly diagnostic for deciding whether the same word was presented before or not. In this theoretical framework, the modulation of the WFE could be explained by an increase in the diagnostic value of matching/mismatching features, resulting from either an increase in the quality of the encoding process (more features added to the memory trace) or by a different deployment of attentional resources (a higher weight is assigned to semantic features). Both of these effects could result from engaging in the previous unrelated lexical decision task and leading to ‘an emphasis’ in the processing of semantic features. Malmberg and colleagues (Malmberg, Zeelenberg, & Shiffrin, 2004) have shown that the disruption of the mirrored-patterned word-frequency effect caused by Mizadolan can be explained in the REM model by assuming a decrease in the accuracy with which features are stored in memory (i.e. c parameter). Similarly, the increase in performance for HF words we observed in the present study could be explained in the REM model by assuming an increase in the c parameter for HF words leading in turn to more features being stored correctly and therefore to a higher hit rate (an implementation for exploratory purposes of the REM model can be found at the GitHub page of the first author at <https://github.com/neurobayesian/REM>).

The results of the present study are consistent with the interpretation that the previous presentation of lexical decision led to better processing of high frequency features in the following recognition, indicated by a positive change in memory sensitivity (d') and in the quality of the encoding process (drift rate). In other words, a selective adaptation of memory processing components to the previous processing history.

Insights into the memory processing components underlying the

modulation of the WFE effect were made possible by adopting the analytical framework of sequential sampling models such as the diffusion decision model or the linear ballistic accumulator model. As previously argued by Criss (2010), quantitative analyses of choice reaction time data provide a unique approach for gaining a deeper understanding of the mechanisms (e.g. processing components) underlying behavioral performance. Crucially they provide a feasible approach to investigate effects in memory performance that do not divide perfectly over empirical manipulations (Forster & Davis, 1984; McKoon et al., 1986; Ratcliff et al., 1985).

The present study offers an intriguing pattern of results; however, some limitations of the current design warrant a replication study to better evaluate the generalizability of the observed effect. Specifically, it would be important to replicate the modulation effect with only single words presentations (no *presentations* manipulation) and with low frequency words slightly more frequent. For the present study it could be argued the *presentations* manipulation introduced a certain amount of episodic processing during lexical decision (e.g. participants might initially mistake a low frequency word for a non-word and over repetitions infer that it is actually a word). An online norming study conducted on the same material and with a comparable population ($N = 31$) indicated that only very few words (~2%) were completely unknown to participants and therefore potentially processed in lexical decision as non-words. The fact that these items were consistently regarded as non-words across the levels of the *presentations* manipulation discounts the hypothesis for the present study that subjects relied on episodic processing (e.g. whether a word was presented or not) to solve the lexical decision task. This could be further verified in future studies by using a more optimal design as highlighted above. It would be also insightful to replicate the modulation effect in recognition using an intentional task as study phase (instead of the familiarity task) which would provide an important term of comparison for evaluating the generalizability of the modulation effect in recognition by contrasting intentional versus incidental types of encoding.

In conclusion, the present study augments the existing literature on semantic and episodic memory by showing for the first time a modulation of the WFE effect in episodic recognition when an unrelated lexical decision task was previously presented. Results of the model-based analysis further revealed that this behavioral effect was explained by changes in the quality of information (e.g. modulation of the drift rate) of the decision process and in global models of memory an increase in evidence accumulation reflects an increase in the retrieval of information from long-term memory (more matching features). Overall in the context of the present study this view suggests that the changes at the level of behavior in the WFE are a reflection of the underlying memory processes dynamically adapting to the previous processing history.

Declaration of Competing Interest

The authors declared that there is no conflict of interest.

Acknowledgements

We would like to thank Amy Criss and Jeff Starns for very useful comments on previous versions of this manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A

See [Tables A.1](#) and [A.2](#).

Table A.1

Stimulus lists rotated across lexical decision (word), recognition (old) and recognition (novel) conditions.

HIGH FREQUENCY WORDS			LOW FREQUENCY WORDS		
List A	List B	List C	List A	List B	List C
aantal	aarde	actie	aalbes	abces	absint
adres	auto	adem	agnost	amulet	aqua
arts	avond	arbeid	asurn	atrium	azalea
bank	baan	baby	baai	baret	bidon
begin	baas	basis	boleet	blaam	biogas
beleid	beeld	begrip	bonsai	buste	boiler
beurt	bezoek	belang	census	canvas	cynici
buik	blik	besef	chic	chalet	creool
burger	bodem	boord	cultus	chroom	carbid
dame	boek	cliënt	debet	delict	dadel
dienst	boog	dood	dictee	dukaat	daluur
eeuw	broek	effect	doerak	dorpel	diode
einde	brood	eiland	elixer	enzym	elpee
feest	deel	fase	epiek	ethica	enigma
gang	ding	fles	errata	ellips	epoxy
gebouw	dokter	geluid	fagot	fysici	factum
geld	dorp	hart	fluum	fatum	fanaat
gemak	feit	hoek	geode	focus	fineer
geval	figuur	hulp	gesp	garde	gluten
glas	film	indruk	grein	gesel	glycol
groep	gebaar	keel	hofnar	gondel	gotspe
hand	gebied	kerk	hyena	hetze	hobo
hemel	gedrag	keuze	idylle	hesp	hortus
hond	geluk	kleur	imam	idioom	hospik
hout	geur	koffie	intern	imker	indigo
jeugd	geweld	licht	jargon	jingle	intro
kaas	greep	markt	jolijt	kobalt	ivoor
kant	haar	massa	kelp	korpus	koliek
keuken	helpt	meneer	kits	kervel	knook
klas	hotel	meter	korvet	legaat	lamet
land	idee	middag	laesie	lagune	lupine
leger	jaar	model	lende	lotus	leviet
lijf	kader	natuur	lychee	module	meiose
maan	keer	olie	makron	musket	mensa
macht	klap	orde	merkel	mout	merite
mate	kring	plan	monist	neon	myopie
melk	lezer	plek	narcis	nering	nimbus
moeite	muur	prins	nuclei	novum	noga
ogen	naam	rand	oblaat	octaaf	nomade
omvang	nacht	reeks	odeur	oratie	octaan
papier	niveau	regen	pekton	odium	odeon
pijn	oosten	ruimte	pipet	pias	oedeem
proces	paar	schip	prefix	plasma	pincet
rijk	prijs	slot	quant	primus	plasma
roman	punt	spel	radial	quota	poëet
school	raam	staat	rappel	radium	pagode
sfeer	schaal	stille	residu	remise	retina
stap	schoot	tafel	sabel	robijn	riedel
strand	slag	trap	saki	sokkel	ragôut
taak	stem	type	seroen	stigma	sores
thee	stof	veld	tarot	teint	silex
titel	studie	verzet	taxus	tombe	serum
vinger	stukje	vloer	toupet	quilt	tyfoon
volk	term	voet	umlaut	unicum	tricot
vraag	vader	vriend	vizier	uterus	teneur
warmte	vlak	waarde	voorn	vaccin	vondel
wereld	week	wagen	wentel	veto	vullis
wezen	winter	wijn	woord	vector	xenon
zand	zaak	zuiden	zilt	wimpel	zefier
ziel	zijde	zuster	zodiac	zwier	zenit

Table A.2

Averaged rank-ordered BIC scores (from best to worst) for the top ten LBA models.

MOD	BIC	RANK	A	CHI	V	S	TO
49	2483.63	259.5	1	1	4	1	1
50	2490.73	305.17	1	1	4	1	2
53	2495.56	368.11	1	1	4	2	1
305	2496.42	369.61	2	1	4	1	1
113	2502.46	437.5	1	2	4	1	1
306	2504.78	462.5	2	1	4	1	2
51	2502.15	465.72	1	1	4	1	3
561	2504.9	489.83	3	1	4	1	1
54	2508.35	514.61	1	1	4	2	2
309	2509.52	528.89	2	1	4	2	1

References

- Anders, R., Riès, S., van Maanen, L., & Alario, F. X. (2015). Evidence accumulation as a model for lexical selection. *Cognitive Psychology*, 82, 57–73. <https://doi.org/10.1016/j.cogpsych.2015.07.002>.
- 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>.
- Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database [Computer software]*. Philadelphia, PA: Data Consortium, University of Pennsylvania.
- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, 46(1), 199–226. <https://doi.org/10.1006/jmla.2001.2803>.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 340–357. <http://www.ncbi.nlm.nih.gov/pubmed/6242411>.
- Boehm, U., Van Maanen, L., Forstmann, B., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage*, 96, 95–105. <https://doi.org/10.1016/j.neuroimage.2014.03.063>.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57(3), 153–178. <https://doi.org/10.1016/j.cogpsych.2007.12.002>.
- Brysbaert, M. (2007). *The language-as-fixed-effect-fallacy: Some simple SPSS solutions to a complex problem*. London: Royal Holloway, University of London. <http://www.pc.rhul.ac.uk/staff/M.Brysbaert/files/TheLanguageAsFixedEffectFallacyVersion20.pdf>.
- Criss, A. H. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology*, 59(4), 297–319. <https://doi.org/10.1016/j.cogpsych.2009.07.003>.
- Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 484–499. <https://doi.org/10.1037/a0018435>.
- Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2013). A differentiation account of recognition memory: Evidence from fMRI. *Journal of Cognitive Neuroscience*, 25(3), 421–435. <https://doi.org/10.1162/jocn.a.00292>.
- Donkin, C., Averell, L., Brown, S., & Heathcote, A. (2009). Getting more from accuracy and response time data: Methods for fitting the linear ballistic accumulator. *Behavior Research Methods*, 41(4), 1095–1110. <https://doi.org/10.3758/BRM.41.4.1095>.
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, 18(1), 61–69. <https://doi.org/10.3758/s13423-010-0022-4>.
- Donkin, C., & Van Maanen, L. (2014). Piéron's Law is not just an artifact of the response mechanism. *Journal of Mathematical Psychology*, 62–63, 22–32. <https://doi.org/10.1016/j.jmp.2014.09.006>.
- Forster, K. I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 680–698. <https://doi.org/10.1037//0278-7393.10.4.680>.
- Forstmann, B. U., Brown, S., Dutilh, G., Neumann, J., & Wagenmakers, E.-J. (2010). The neural substrate of prior information in perceptual decision making: A model-based analysis. *Frontiers in Human Neuroscience*, 4(May), 40. <https://doi.org/10.3389/fnhum.2010.00040>.
- Gardner, M. K., Rothkopf, E. Z., Lapan, R., & Lafferty, T. (1987). The word frequency effect in lexical decision: Finding a frequency-based component. *Memory & Cognition*, 15(1), 24–28. <https://doi.org/10.3758/BF03197709>.
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13(1), 8–20.
- Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29, 228–244. <http://www.sciencedirect.com/science/article/pii/0749596X9090074A>.
- Guttentag, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*, 37(4), 502–516. <https://doi.org/10.1006/jmla.1997.2532>.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review*, 9(2), 1–31. <https://doi.org/10.3758/BF03196299>.
- Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology Learning Memory and Cognition*, 23(6), 1306–1323. <https://doi.org/10.1037/0278-7393.23.6.1306>.
- Kuznetsova, A., Brockhoff, P. B., & Bojensen Christensen, R. H. (2014). Package 'lmerTest'.
- Lofthus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476–490. <https://doi.org/10.3758/BF03210951>.
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 322–336. <https://doi.org/10.1037/0278-7393.31.2.322>.
- Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, 30(4), 607–613. <http://www.ncbi.nlm.nih.gov/pubmed/12184562>.
- Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by Midazolam. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(2), 540–549. <https://doi.org/10.1037/0278-7393.30.2.540>.
- McKoon, G., Ratcliff, R., & Dell, G. S. (1986). A critical evaluation of the semantic-episodic distinction. *Journal of Experimental Psychology: Learning Memory and Cognition*, 12(2), 295–306.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In Gordon H. Bower (Ed.), *Psychology of learning and motivation* (pp. 207–262). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60162-0](https://doi.org/10.1016/S0079-7421(08)60162-0).
- Raaijmakers, J. G. W. (2005). Modeling implicit and explicit memory. *Human learning and memory: Advances in theory and application: The 4th Tsukuba international conference on memory* (pp. 85–105). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers. <http://books.google.com/books?hl=en&lr=&id=GMB5AgAAQBAJ&oi=fnd&pg=PA85&dq=Modeling+Implicit+and+Explicit+memory&ots=qsLetn0ZhC&sig=IFf2F8Nlp23hrFSdb3eEwA7HdCo>.
- Raaijmakers, J. G. (2008). Mathematical models of human memory. In H. L. Roediger (Vol. Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference: 4 vols.*
- Ratcliff, Roger (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108. <https://doi.org/10.1037/0033-295X.85.2.59>.
- Ratcliff, Roger, Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, 111(1), 159–182. <http://psycnet.apa.org/journals/rev/111/1/159/>.
- Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General*, 114(4), 435–450. <http://www.ncbi.nlm.nih.gov/pubmed/2934497>.
- Ratcliff, R., & McKoon, G. (1986). More on the distinction between episodic and semantic memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(2), 312–313.
- Ratcliff, Roger, & Smith, P. (2004). A comparison of sequential sampling models for two-choice. *Psychological Review*, 111(2), 333–367.
- Ratcliff, Roger, Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, 19(2), 278–289. <http://psycnet.apa.org/journals/pag/19/2/278/>.
- Ratcliff, Roger, Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, 50(4), 408–424. <https://doi.org/10.1016/j.jml.2003.11.002>.
- Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3(1), 1–17. <https://doi.org/10.1037/0096-1523.3.1.1>.
- Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(4), 512–524. <https://doi.org/10.1198/108571102726>.
- Schooler, L. J., Shiffrin, R. M., & Raaijmakers, J. G. W. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological review*.

- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>.
- Shiffrin, R., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*. <http://www.springerlink.com/index/C6G85741088R3247.pdf>.
- Spaniol, J., Madden, D., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology. Learning Memory and Cognition*, 32(1), 101–117. <http://psycnet.apa.org/journals/xlm/32/1/101>.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, and Computers*, 31(1), 137–149. <https://doi.org/10.3758/BF03207704>.
- van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(48), 17488–17495. <https://doi.org/10.1523/JNEUROSCI.2924-11.2011>.
- van Maanen, L., Forstmann, B. U., Keuken, M. C., Wagenmakers, E.-J., & Heathcote, A. (2015). The impact of MRI scanner environment on perceptual decision-making. *Behavior Research Methods*(2012), <https://doi.org/10.3758/s13428-015-0563-6>.
- van Maanen, L., van Rijn, H., & Taatgen, N. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science*, 36(1), 62–101. <https://doi.org/10.1111/j.1551-6709.2011.01213.x>.
- Wagenmakers, E. M., Huber, D., Raaijmakers, J. G. W., Shiffrin, R. M., & Schooler, L. J. (2003). REMI and ROUSE: Quantitative models for long-term and short-term priming in perceptual identification, 1–25.
- Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language*, 58(1), 140–159. <https://doi.org/10.1016/j.jml.2007.04.006>.
- Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, 48(3), 332–367. <https://doi.org/10.1016/j.cogpsych.2003.08.001>.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517. <https://doi.org/10.1006/jmla.2002.2864>.