Acta Psychologica 262 (2026) 106068

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Recognizing the unknown: Motor-response execution reflects the availability of positive evidence during recognition

Elisa Fiora [a], Michele Scaltritti [b], Simone Sulpizio [a,c,d,*]

[a] Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Milan, Italy
[b] Dipartimento di Psicologia e Scienze Cognitive, Università degli Studi di Trento, Rovereto, Italy
[c] Milan Center for Neuroscience (NeuroMI), Università degli Studi di Milano-Bicocca, Milan, Italy
[d] Mind and Behavior Technological Center, University of Milano-Bicocca, Milan, Italy

ARTICLE INFO

Keywords:
Memory-based recognition
Motor-response execution
Lexical decision
Decision making

ABSTRACT

Recognition is a core component of cognition, yet how we categorize a stimulus as unknown remains an open issue. In this study, we tested whether, during recognition, motor-response execution is affected by the presence vs absence of positive evidence for the to-be-recognized stimulus. Participants performed a visual lexical decision task with five categories of stimuli, varying along a word-likeness gradient: High-frequency words, low-frequency words, letter-substitution pseudowords (derived from real words by changing one letter), pseudowords (non-existing words not resembling any words), and illegal strings of letters. The effect of stimulus category was separately examined in the premotor and motor component of the response. Premotor times (and RTs) varied systematically with word-likeness: The more extreme categories of high-frequency words and illegal letter strings yielded the fastest responses, whereas the more ambiguous categories produced slower responses. In contrast, motor time, which indexed motor-response execution, was insensitive to word-likeness, but faster for stimuli providing positive evidence about their lexical status (i.e., words and illegal letter strings), compared to those that did not (i.e., pseudowords and letter-substitution pseudowords). These results point to a differentiation in the decisional dynamics modulating premotor and motor time, with motor-response execution being selectively sensitive to the presence vs absence of positive evidence, rather than to the more general dimension of word-likeness. More broadly, the findings highlight a differentiation in the processes underlying memory-based recognition depending on the presence vs absence of positive evidence, particularly with respect to the implementation of the corresponding action.

1. Introduction

Many goal-directed actions rely on our ability to recognize the stimuli we are interacting with. Whether clicking an icon on a desktop or making a high-stake decision, achieving these goals hinges on this fundamental skill. Recognition requires us to determine, based on the information we can retrieve from memory, whether we know a stimulus or not, ultimately determining the possibility to access related knowledge and, thus, to guide our decisions and actions.

A paradigmatic task to study decisions associated with recognition in long-term semantic memory is the lexical decision task (Balota & Chumbley, 1984; Norris, 2006; Rubenstein, Garfield, & Millikan, 1970). In its visual version, participants are presented with printed strings, and have to recognize them either as real words or nonwords, typically by pressing a button. Despite the apparent simplicity of the task and a rich tradition of studies, the process underlying the "recognition" of nonwords remains debated (e.g., Dufau, Grainger, & Ziegler, 2012; see also, e.g., Kelly & Tucker, 2022; Nenadić & Tucker, 2020; for the same issue in the auditory domain). What does it mean to "recognize" an item with no representation in memory? What type of evidence could support this type of response?

This uncertainty is testified by the heterogeneity of the computational solutions proposed to describe the decisional dynamics underlying nonword responses. A first issue concerns the presence vs absence of a specific decision variable for nonwords. Some models posit that both word and nonword responses are driven by a single decision variable,

* Corresponding author at: Simone Sulpizio, Dipartimento di Psicologia, Università degli Studi di Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 – Milano, Milan, Italy.
E-mail address: simone.sulpizio@unimib.it (S. Sulpizio).

https://doi.org/10.1016/j.actpsy.2025.106068
Received 21 July 2025; Received in revised form 2 December 2025; Accepted 2 December 2025

which accumulates noisy evidence toward one of the two competing alternatives (e.g. Ratcliff, Gomez, & McKoon, 2004; Wagenmakers et al., 2004; see also, Norris, 2006). Here, evidence toward one alternative corresponds to evidence against the other. Other implementations, instead, assume separate decision variables for word vs nonword decisions, which may either compete—for example via lateral inhibition (e.g., Davis, 2010; Dufau et al., 2012; Usher & McClelland, 2001)—or accumulate independently in a race dynamic (e.g., Brown & Heathcote, 2008). A second issue concerns the informational content driving the nonword response. One view envisages evidence for nonwords as subtractively derived from the available lexical evidence (Dufau et al., 2012). In this context, nonword evidence is indirect and depends on the strength of the match between the signal and the activated memory traces (i.e., lexical entries). Alternatively, nonword responses may be supported by direct evidence of a stimulus' nonlexical status (e.g., Norris, 2006; Wagenmakers et al., 2004), for example by accessing the degree of (mis)match between the signal and the activated representations when they do not correspond. In this case, the signal extracted from the input provides (partially) independent evidence for word vs. nonword responses (Brown & Heathcote, 2008; Wagenmakers et al., 2004), suggesting some differentiation in the underlying decision processes.

The variety of computational approaches to nonword recognition highlights a broader challenge: explaining how a response is generated despite the absence of a corresponding memory representation for the stimulus (cf. Hendrix & Sun, 2021; Yap, Sibley, Balota, Ratcliff, & Rueckl, 2015). This issue resonates with the classic distinction between local activation—the selective activation of a specific lexical representation—and global activation—the distributed activation across multiple lexical candidates (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Norris, 2006). While local activation can be taken as a signal of positive evidence for a lexical match, global activation reflects the more general extent to which a given stimulus resembles known words, even when no exact match is found. In this sense, responses to pseudowords often rely on a gradient of global activation rather than on local evidence from a stored representation. Potentially, the presence or absence of a stored representation may imply qualitative differences in the decisional dynamics involved in words vs nonword responses. An empirical hint in this direction comes from lexical decision studies encompassing motor-response execution in their chronometric measurements (Scaltritti, Greatti et al., 2023; Scaltritti, Giacomoni et al., 2023; Scaltritti, Greatti, & Sulpizio, 2024; Kamari Songhorabadi et al., 2025). For classic button-press responses, the electromyographic (EMG) signal of the muscle responsible of the button-press is used to divide the response time (RT) into a premotor time (PMT)—which extends from stimulus onset to the onset of motor activity—, and a motor time (MT)—which extends from the onset of motor activity until the actual button press (Botwinick & Thompson, 1966; Weiss, 1965).

The results from the lexical decision studies consistently show prolonged MTs for nonwords compared to words, a pattern also observed in object-decision tasks comparing real objects with non-existing pseudo-objects (Scaltritti, Greatti et al., 2023). Notably, across these previous studies, MTs displayed a series of phenomena that were independent from those observed on overall RTs and their premotor component (PMTs), suggesting that MT could grasp specific decisional components. In fact, unlike PMT, MT is not affected by lexical frequency (Scaltritti, Greatti et al., 2023; but see Dendauw et al., 2024), as the difference between high- and low-frequency words (e.g., *house* vs *sapphire*) remains bounded to the premotor interval. Moreover, the lexicality effect—i.e., the difference between words and nonwords—on MTs is impervious to other manipulations like speed-accuracy trade off (Scaltritti et al., 2024) and response bias (as the proportion of words/nonwords composing the experimental list; Kamari Songhorabadi et al., 2025), which instead affect the magnitude of the lexicality effect at the level of PMTs (see also, Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). This selectivity in the decisional phenomena that can reach the motor component suggests that MT may not reflect a mere propagation of premotor decisional dynamics, but instead captures a more qualitative difference in processing words vs nonwords.

A fundamental difference between words and nonwords may lie in the different types of evidence supporting the two decisional outcomes. Typically, only word responses can be based on conclusive positive evidence, via the activation of a stored representation in long-term memory that matches the stimulus signal (for a generalization to recognition memory including but not limited to lexical decision, see, e.g., Shiffrin, 2003). In contrast, the commitment to a nonword response seems based on negative—i.e., lack of positive—evidence (e.g., Davis, 2010; Dufau et al., 2012; Wagenmakers et al., 2004). As, by definition, nonwords lack memory representations, their recognition is based on a weak (or absent) match between the stimulus and any memory trace.

The lexicality effect on MT may thus stem from a qualitative difference between actions driven by decisions based on positive vs negative evidence, with the two processes being informed by (partially) different decision dynamics. To test this hypothesis, we implemented a lexical decision task designed to tease apart the dimension of interest—that is, the presence vs absence of positive evidence—from the stimulus category—that is, word vs nonword. Five types of stimuli were used, varying on a gradient in word-likeness: a) High-frequency words (the most word-like), b) low-frequency words, c) letter-substitution pseudowords (derived from real words of medium frequency by changing only one letter; e.g., *merenca* derived from *merenda* 'snack'), d) pseudowords (non-existing words not resembling any real words; e.g., *taderco*), and e) illegal strings of letters (unreadable sequence of letters; e.g., *lbqtrsv*). The latter are particularly relevant as they elicit (almost) no lexical activity (e.g., Yap et al., 2015) and can be recognized as nonwords using a very shallow processing based on admissibility/likelihood of letter sequences, thus bypassing any lexical processing (e.g., Cohen et al., 2002; Shulman & Davidson, 1977; Stone & Van Orden, 1992; Vinckier et al., 2007). Importantly, illegal letter strings can be positively identified as nonwords, due to the presence of orthotactic violations and illegal grapheme sequences. In contrast, pseudowords and letter-substitution pseudowords may fail to provide definitive confirmation of their non-lexical status—and thus to univocally map toward the nonword response—due to the inherent ambiguity related with a) the absence of a memory representation to be matched against, and b) their similarity with words prompting a (reduced) lexical activation. For what concerns words, albeit evidence toward the word response may accumulate with a different rate for high- and low-frequency words (e.g., Gomez & Perea, 2014; Yap, Balota, Sibley, & Ratcliff, 2012), for both types of stimuli the final commitment to the word response is based on the presence of positive evidence—i.e., the match with a trace stored in memory. Both categories can thus be positively identified as words.

Based on the above considerations, and under the hypothesis that MT reflects the different decisional dynamics underlying responses based on the presence vs absence of positive evidence, we predicted a dissociation between RTs/PMTs and MTs. RTs/PMTs are expected to replicate the well-established pattern of results documented in the literature (e.g., Andrews, 1997; Peressotti, Job, Rumiati, & Nicoletti, 1995; Ratcliff et al., 2004). Specifically, RTs/PMTs should vary as a function of word-likeness, with the two more extreme categories being the fastest (e.g., Berberyan, van Rijn, & Borst, 2021; Marinelli, Traficante, & Zoccolotti, 2014; Ratcliff et al., 2004), followed by the less word-like categories of low-frequency words and pseudowords, which have provided mixed results in terms of their relative speed (e.g., Barca & Pezzulo, 2012; Ratcliff et al., 2004; Wagenmakers et al., 2008). Letter-substitution pseudowords, instead, should be the slowest category, being characterized by a high word-likeness but requiring a nonword response. Differently, MTs are expected to vary only as a function of the availability of positive evidence, independently of word-likeness. Specifically, MTs should be longer for both pseudowords and letter-substitution pseudowords than for high- and low-frequency words. Critically, for illegal letter strings, albeit requiring a no-response like pseudowords and letter-substitution pseudowords, MTs were expected to be as fast as for high- and low-frequency words (and thus faster than both pseudowords and letter-substitution pseudowords), as in all these cases the decision could be based on the presence of positive evidence toward the to-be-delivered response. Finally, similar MTs are expected for high- and low-frequency words (replicating previous findings, Scaltritti, Greatti et al., 2023), and for pseudowords and letter-substitution pseudowords, which remain comparable due to the absence of definitive positive evidence.

To further corroborate our manipulation and assess potential converging evidence of a dissociation between MTs and RTs/PMTs, we also analyzed multiple indexes of response accuracy, including partial errors (i.e., sub-threshold muscular activation of the response hand associated with the incorrect response occurring before the correct response is delivered; Eriksen, Coles, Morris, & O'hara, 1985; Hasbroucq, Possamaï, Bonnet, & Vidal, 1999), and variations in correct responses and incorrect activations (partial or overt errors) as a function of response speed (e.g., Fluchère et al., 2018; Luce, 1986; Ollman, 1977). These indexes are sensitive to stimulus lexicality (Scaltritti et al., 2021; Scaltritti, Greatti et al., 2023; Scaltritti et al., 2024) and are used to offer a fine-grained characterization of the different stimulus categories with respect to their position in the word-likeness gradient. Specifically, for the highly ambiguous category of low-frequency words, incorrect activations should be evenly distributed across response latencies, whereas for pseudodowords they should be maximal in fast responses because of lexical capture phenomena (e.g., Scaltritti et al., 2021; Scaltritti, Greatti et al., 2023; Scaltritti, Giacomoni et al., 2023). Pseudowords might show an intermediate pattern, with incorrect activations being maximal in fast responses (Scaltritti, Giacomoni et al., 2023; Scaltritti, Greatti et al., 2023) but present across all response latencies because of their ambiguity. Finally, the comparison between incorrect and correct activations should reveal differences in the possibility to correct ongoing errors, as a function of stimulus category and response latency. Successful corrections are expected to be maximal for high-frequency words and illegal letter strings, representing the two extremes of the word-likeness gradient.

2. Method

2.1. Participants

The sample size was determined based on recent recommendations in the field (Brysbaert, 2019) and previous investigations (Scaltritti, Giacomoni et al., 2023; Scaltritti, Greatti et al., 2023; Scaltritti et al., 2024). Forty-three participants (35 females; $M_{age} = 23.4$; $SD_{age} = 3.45$) took part in the study. All participants reported to be Italian native speakers, had normal or corrected-to-normal vision, and reported no history of neurological or learning disorders. Handedness was assessed using the Edinburgh Handedness Inventory (Oldfield, 1971). Thirty-nine participants were classified as right-handers ($M = 81, 91$; $SD = 32.18$), 1 as a mixed right-hander (handedness scores = 20) and 1 as a mixed left-hander (handedness scores = -88.24). Data from two participants were excluded: one due to self-reported ongoing use of antidepressants, and one who reported having corrected-to-normal vision but failed to wear corrective lenses during testing. The final sample thus included 41 participants. Participants received either € 15 or course credits for compensation, according to their preference, and provided signed informed consent before the beginning of the experimental procedure. The study was evaluated by the local commission for minimal-risk studies of the Psychology Department at the University of Milano-Bicocca (protocol nr.: RM-2024-763). Materials and raw data are publicly available via the Open Science Framework (https://osf.io/zt8eh).

2.2. Stimuli

One set of 200 high-frequency words (e.g., *inverno* 'winter') and one set of 200 low-frequency words (e.g., *califfo* 'caliph') were selected from the PhonItalia database (Goslin, Galluzzi, & Romani, 2014). High- and low-frequency words had a log frequency equal/higher than 3 and lower than 0.7, respectively. Three sets of 200 nonwords each were created. One set was composed of illegal letter strings (e.g., *lbqtrsv*), which were created by assembling consonants to form orthographically and phonologically illegal sequences. One set was composed of pseudowords consisting of readable sequences of letter without clear resemblance with any real Italian word (e.g., *taderco*), which were assembled ensuring to be orthographically and phonologically legal. The last set was composed of letter-substitution pseudowords, which were created by changing one letter from real words of medium-to-low lexical frequency taken from the Phonitalia database (e.g., *merenca*, derived from *merenda* 'snack'). The five sets were comparable in terms of number of letters (all $ps > 0.2$, resulting from independent-sample two-tailed t-tests), but differed on both OLD20 (which is the orthographic Levenshtein distance to the twenty closest neighbors, Yarkoni, Balota, & Yap, 2008) and bigram frequency in order to maximize the difference between the categories of stimuli, and increase categories 'prototypicality'. The characteristics of the stimuli are reported in Table 1. Note that, with respect to pseudowords and low-frequency words, letter-substitution pseudowords were lower in OLD20 and higher in bigram frequency, which means that they were orthographically more similar to words than the other two categories. Moreover, with respect to pseudowords, low-frequency words were lower in OLD20 (i.e., more similar to words) and in bigram frequency (i.e., more similar to illegal letter strings). This absence of systematicity in the orthographic measures guarantees that pseudowords and low-frequency words could not be discriminated only on the basis of visuo-orthographic plausibility.

A further set of 200 words (100 of high frequency and 100 of low frequency) was selected from the PhonItalia database as fillers to keep the proportion of words/pseudowords equal.

2.3. Apparatus and procedure

At the beginning of the experiment, participants read and signed the informed consent, and completed a questionnaire collecting demographic and health-related information. Then, the EMG electrodes were installed (for details, see below).

The experimental procedure and the acquisition of behavioral data were controlled via the E-Prime 3 software (version 3.0.3.214, Psychology Software Tools) running on a desktop PC.

Participants sat in front of the screen at approximately 50 cm, holding two cylindric buttons (connected to a Black Box Toolkit module), with their thumbs resting on the buttons (the response force threshold for the handheld buttons corresponded to ~490–550 g). Participants were instructed to categorize each stimulus as a word or a nonword, using their thumbs to respond. Response speed and accuracy were equally emphasized.

The experimental session was divided into two main blocks, each one associated with a different stimulus (word vs. nonword) response (right vs. left hand) mapping. This ensured that, within each category of stimuli, each participant performed an equal number of responses with the two hands. Half of the participants started responding to words with their right thumb, and the other half with their left thumb. Within each block, half of the trials were words (100 high-frequency words, 100 low-frequency words, and 100 fillers, half of which were of high frequency and half of low frequency) and half nonwords (100 trials for each category) The assignment of stimulus subsets to each block was also counterbalanced across participants. Each block started with 20 practice trials (not used in the main experiment) to familiarize with the procedure. Participants could take a self-terminated break every 120 trials.

Each trial started with a fixation cross displayed in the center of the

Table 1
Means (and SDs) of the psycholinguistic variables for the 5 sets of stimuli used in the experiment.

Variables	HF	LF	S PW	PW	LS
Frequency (log)	4.56 (1.04)	0.31 (0.34)	–	–	–
Nr. of letters	7.15 (1.57)	7.11 (1.38)	7.17 (1.28)	7.02 (1.38)	7.13 (1.52)
OLD20	1.92 (0.51)	2.38 (0.61)	2.07 (0.47)	2.59 (0.65)	4.12 (1.03)
Big. freq. Sum	927,797.68 (243,488.18)	496,763.8 (185,885.59)	858,045.62 (256,081.78)	627,964.54 (256,511.96)	124,955.2 (93,297.47)
Big. freq. Mean	152,697.2 (17,087.80)	79,658.52 (19,765.66)	138,470.36 (27,394.47)	102,609.3 (28,768.45)	19,634.54 (13,705.05)

Note. HF = High Frequency words; LF = Low Frequency words; S PW = letter-substitution pseudowords; PW = pseudowords; LS = letter stings; Nr. of letters = Number of letters; Bigr. freq. Sum = summed bigram frequency; Bigr. freq. Mean = mean bigram frequency. For words, variables were extracted from the PhonItalia database (Goslin et al., 2014). For nonwords, the OLD20 was computed on the PhonItalia database using the vwr package (Keuleers, 2013) in R. Bigram frequency variables were drawn from the same database using a custom-made script.

screen for a randomly sampled duration (700, 750, 800, 850). Then, the target string appeared and remained on the screen until participants' response or for a maximum of 1500 ms. A blank screen lasting 400 ms was used as inter-trial interval. Stimuli were presented in Courier New font 25, in black against a grey background (RGB: 190, 190, 190). Error and time-out feedback messages ("*ERROR*"; "*TOO SLOW*") were displayed for 500 ms, in red font, only during the practice phase.

The full session, including electrode preparation, lasted approximately 90 min.

2.4. EMG recording and processing

EMG activity of the *flexor pollicis brevis* was recorded using an MP160 BioPac System (connected to a laptop) in combination with BioNomadix wireless devices (attached to the participants forearm via Velcro strap). The signal was sampled at 1000 Hz and a 10 Hz high-pass filter. For data acquisition, two electrodes were placed on the thenar eminences of each hand, ~2 cm apart, and a ground electrode was placed on the pisiform bone of each wrist. Before placing the electrodes, the skin was cleaned with isopropyl alcohol and a gently abraded using Nuprep gel (Weaver and Company, Aurora, CO, USA).

Offline signal preprocessing was conducted in MATLAB (version 2018b, MathWorks Inc., Natick, MA, USA), with EEGLAB (version 14_1_2b; Delorme & Makeig, 2004) and ERPLAB functions (Lopez-Calderon & Luck, 2014), and custom scripts. A 50 Hz notch filter was applied to the continuous data before segmenting it into epochs ranging from −500 ms to 2100 ms relative to stimulus onset (time 0). EMG onset within each epoch was determined using the integrated profile method (Liu & Liu, 2016), which computes the cumulative absolute EMG signal and then subtracts it from the straight line connecting the first and last data points (equivalent to the cumulative sum of a uniform distribution). The EMG onset was defined as the sample where this difference reached its minimum.

To support the detection of artifacts and partial responses, within each epoch, windows of EMG activity were identified based on consecutive samples in which the rectified EMG signal exceeded the threshold of 3.5 SDs above the mean absolute activity computed in the baseline period (from −500 to 0 ms). Windows separated by an interval shorter than 25 ms were merged, whereas those shorter than 50 ms or beginning after the button-press were excluded. Epochs in which 2 or more windows of activity were detected were automatically flagged. All epochs were visually inspected, and only those in which the EMG onset accurately aligned with the last window of activity representing the response-related EMG burst were retained, thereby excluding artifacts due to false starts, noise, and signal drift. On average, 2.13 % (SD = 1.74) of trials per participant were discarded using these criteria. The same preprocessing pipeline was applied to the hand not involved in the button-press, to detect partial errors and partial correct responses. Epochs with one or more windows of activity within these channels were flagged by the algorithm. Then, all epochs were visually inspected and partial responses were considered valid if the covert EMG activation was clustered in a visually clear burst and its onset was accurately detected. On average, partial errors occurred in 4.31 % of all trials (SD = 3.65 %). Partial correct responses were very rare (M = 0.34 %, SD = 0.27 %) and thus not further analyzed.

2.5. Measures

2.5.1. Chronometric measures

Analyses were conducted exclusively on pure-correct trials, defined as correct response without any covert activation in the non-response hand. RTs were fractionated into two components: PMTs (from stimulus onset to EMG burst onset) and MTs (from EMG burst onset to button press). Each of these chronometric measures was analyzed separately.

2.5.2. Response accuracy

Accuracy analyses included both correct and incorrect responses, regardless of whether covert EMG activation occurred in the non-responding hand. Trials in which participants failed to respond within the allotted time (4.5 %) were removed from the analysis.

To examine how accuracy varied as a function of response latency, conditional accuracy functions (CAFs) were computed. Within each participant and condition, trials were divided into five quintiles as a function of PMTs (with cut-off at 0.2, 0.4, 0.6, 0.8, and 1). Thus, the first quintile included the fastest 20 % of the trials, the second quintile the next fastest 20 % of the trials, and so on until the fifth quintile, which included the slowest 20 % of the trials. Quintiles served as a fixed factor in the analysis. Conditional incorrect activations functions (CIAFs) were computed in a parallel manner, except that partial error latencies were included in the data, thus enabling assessment of changes in the proportion of incorrect EMG activations as a function of (partial) response latency.

Partial errors analyses were conducted on correct responses only, in order to evaluate the likelihood of covert incorrect activations across experimental conditions.

2.6. Statistical analyses

Chronometric measures were analyzed using linear mixed-effects (LMEs, gaussian family) models, whereas binomial measures of response accuracy and partial errors were analyzed via generalized mixed-effects models (GLMEs, binomial family), using the lme4 library (version 1.1–37; Bates, Mächler, Bolker, & Walker, 2015) in R (version 4.5.1; R Core Team, 2025). Within GLMEs, the maximum number of iterations was increased (2^5) and a *bobyqa* algorithm was used for model optimization.

In all analyses, the fixed effect of stimulus category (high frequency words vs low frequency words vs letter-substitution pseudowords vs pseudowords vs illegal letter strings) was assessed by comparing the model in which it was present vs absent, and was considered significant only when the F-test for linear models (calculated with the lmerTest package, Kuznetsova, Brockhoff, & Christensen, 2017) and likelihood ratio test for generalized models revealed a significant increase in goodness of fit from its inclusion. We started fitting the model of maximal complexity including by-participants and by-items random intercepts, the by-participants random slope, and the correlations between random slopes, and random slopes and intercepts. In case of failure to converge, models were progressively simplified by first removing correlations and then the terms associated with the smallest amount of explained variance. For CAFs and CIAFs, fixed effects also included quintiles and their interaction with stimulus category. Possible non-linear trends in modelling the quintile variable were assessed by using second order polynomials. In these analyses, the random-effect structure was limited to random intercepts to aid convergence. For all the analyses, follow-up pairwise comparisons were conducted on estimated marginal means using the *emmeans* package (version 1.11.2, Lenth & Piaskowski, 2025). A Bonferroni correction was applied to the resulting *p*-values to control for Type I error inflation and degrees of freedom were approximated using the Satterthwaite method.

3. Results

3.1. Chronometric measures

The effect of stimulus category was significant in all the analyses (RTs: $F(4, 51.11) = 9.70, p < .001$; PMTs: $F(4, 50.74) = 9.93, p < .001$; MTs: $F(4, 965.92) = 12.48, p < .001$). Parameters of the fixed effects are reported in Table 2 (see Supplemental Material Table S1 for the random effects). As visible in Fig. 1, both RTs and PMTs closely followed the word-likeness gradient of the stimuli. Latencies were shorter for the more extreme categories of the continuum, high frequency words and illegal letter strings, whereas they were longer for more ambiguous stimuli such as low frequency words, letter-substitution pseudowords and pseudowords. In detail, pairwise comparisons (Table 3) revealed that RTs were shorter for illegal letter strings compared to all the other categories. High frequency words were the second fastest category, followed by low frequency words and pseudowords, which did not differ from one other, and letter-substitution pseudowords, which were the slowest ones. Results from PMTs revealed the same pattern (Table 3; Fig. 1). MTs, instead, were selectively modulated by the presence vs absence of positive evidence, with longer latencies for letter-substitution pseudowords and pseudowords (which both lack of positive evidence) compared to both high and low frequency words, as well as illegal letter strings (for all of which the responses could be delivered on the basis of positive evidence), which in turn yielded comparable response durations. Indeed, pairwise comparisons showed comparable MTs for high frequency words, low frequency words, and illegal letter strings. These stimuli all yielded faster MTs compared to both pseudowords and letter-substitution pseudowords, which did not differ from one each other.

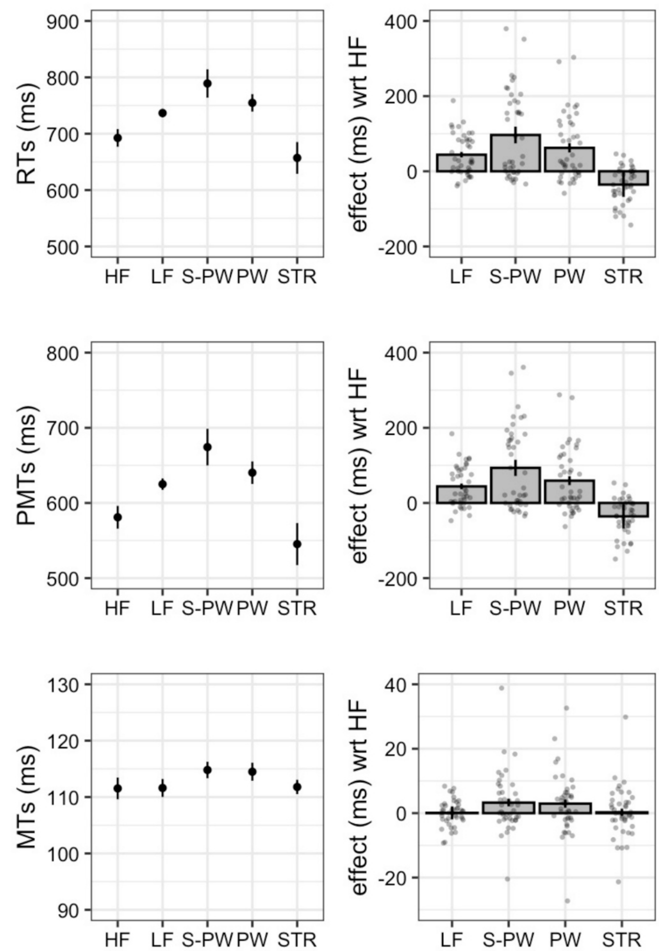

Fig. 1. Results for the chronometric measures.
Note. Results for reaction times (RTs; first row), premotor times (PMTs; second row), and motor times (MTs, third row). The first column reports mean chronometric measures by stimulus category (HF = high frequency words; LF = low frequency words; S-PW = letter-substitution pseudowords; PW = pseudowords; STR = illegal letter strings). The second column reports the mean effect for each category with respect to (wrt) high-frequency words. Points represent individual difference scores. For all plots, error bars display 95 % confidence intervals, adjusted for within-participants variables following Morey (2008).

3.2. Response accuracy

Parameters of the fixed effects for the analyses of correct responses and partial errors are reported in Table 4 (see Supplemental Material Table S2 for the random effects). Pairwise comparisons are reported in Table 5. The effect of stimulus category was significant in both analyses (correct responses: $\chi^2\ [4] = 105.64, p < .001$; partial errors: $\chi^2\ [4] = 37.04, p < .001$). Participants were most accurate in categorizing illegal letter strings, then, in order, high frequency words, pseudowords, and low frequency words and letter-substitution pseudowords, with the latter two categories not differing from each other. Moreover, letter-substitution pseudowords elicited more partial errors than high frequency words, low frequency words and illegal letter strings; also, pseudowords elicited more partial errors than illegal letter strings. Results are also represented in Fig. 2 (upper panels).

Parameters of the fixed effects for the CIAF and the CAF analysis, are reported in Table 6. For both the CIAF and the CAF analysis, the non-linear interaction between stimulus category and quintile was significant (CIAF: $\chi^2\ [8] = 172.18, p < .001$; CAF: $\chi^2\ [8] = 74.54, p < .001$). As visible in Fig. 2 (lower panels), for CIAFs the interaction reveals that,

Table 2
Fixed effects for LME models on chronometric measures.

	RTs				PMTs				MTs			
Fixed effects	Est	SE	t	p	Est	SE	t	p	Est	SE	t	p
Intercept	692.41	11.85	58.45	< 0.001	580.95	11.45	50.72	< 0.001	111.40	3.79	29.34	< 0.001
LF words	45.60	8.63	5.28	< 0.001	45.45	8.56	5.30	< 0.001	0.19	0.66	0.29	0.77
S Pseudowords	97.43	17.90	5.44	< 0.001	94.11	17.30	5.43	< 0.001	3.37	0.66	5.09	< 0.001
Pseudowords	62.22	13.68	4.54	< 0.001	59. 18	13. 20	4.48	< 0.001	3.09	0.64	4.80	< 0.001
Letter strings	−35.33	7. 58	−4.66	< 0.001	−35.68	7. 81	−4.56	< 0.001	0.19	0.67	0.72	0.47

Note. RTs = reaction times; PMTs = premotor times; MTs = motor times; SE = standard error; HF = high frequency; LF = low frequency; S = letter-substitution.

Table 3
Pairwise comparisons on chronometric measures.

	RTs			PMT			MT		
Fixed effects	Est	SE	z	Est	SE	z	Est	SE	z
HF – LF	−45.6	8.63	−5.28	−45.5	8.57	−5.30	−0.19	0.66	-0.29^{ns}
HF – S PW	−97.4	17.90	−5.44	−94.1	17.30	−5.43	−3.37	0.66	−5.09
HF – PW	−62.2	13.70	−4.54	−59.20	13.20	−4.48	−3.09	0.64	−4.80
HF – STR	35.3	7.58	4.66	35.7	7.82	4.56	−0.45	0.63	-0.72^{ns}
PW – LF	16.6	6.97	2.38^{ns}	13.7	6.75	2.03^{ns}	2.90	0.67	4.28
PW – S PW	−35.20	6.03	−5.84	−34.9	5.93	−5.89	−0.27	0.67	-0.40^{ns}
PW – STR	97.6	19.00	5.14	94.9	18.60	5.09	2.64	0.64	4.10
S PW – LF	51.80	10.73	4.83	48.7	10.34	4.71	3.17	0.69	4.58
S PW – STR	132.8	23.50	5.65	129.8	23.00	5.64	2.91	0.66	4.41
STR – LF	−80.90	14.50	−5.59	−81.1	14.60	−5.54	0.26	0.66	0.39^{ns}

Note. RTs = reaction times; PMTs = premotor times; MTs = motor times; SE = standard error; HF = high frequency words; LF = low frequency words; PW = pseudowords; S PW = letter-substitution pseudowords; STR = illegal letter strings; ns = not significant.

Table 4
Fixed effects for generalized LME models on accuracy measures.

	Accuracy				Partial Errors			
Fixed effects	Est	SE	z	p	Est	SE	z	p
Intercept	4.35	0.17	25.32	< 0.001	−3.46	0.12	−27.30	< 0.001
LF words	−2.26	0.16	−13.69	< 0.001	0.14	0.08	1.65	0.09
S Pseudowords	−1.91	0.19	−9.85	< 0.001	0.39	0.08	4.66	< 0.001
Pseudowords	−0.80	0.21	−3.79	< 0.001	0.16	0.08	1.91	0.05
Letter strings	0.69	0.21	3.269	0.001	−0.10	0.08	−1.15	0.24

Note. SE = standard error; S = letter-substitution; HF = high frequency; LF = low frequency.

Table 5
Pairwise comparisons on accuracy measures.

	Accuracy			Partial Errors		
Fixed effects	Est	SE	z	Est	SE	Z
HF – LF	2.26	0.16	13.69	−0.14	0.08	-1.65^{ns}
HF – S PW	1.29	0.19	9.85	−0.39	0.08	−4.66
HF – PW	0.80	0.23	3.79	−0.16	0.08	-1.91^{ns}
HF – STR	−0.69	0.21	−3.26	0.10	0.08	1.15^{ns}
PW – LF	1.46	0.21	6.73	0.01	0.08	0.21^{ns}
PW – S PW	1.12	0.14	7.95	−0.22	0.08	-2.76^{ns}
PW – STR	−1.49	0.22	−6.72	0.26	0.08	3.04
S PW – LF	0.33	0.19	1.78^{ns}	0.24	0.08	2.87
S PW – STR	−2.62	0.21	−12.23	0.49	0.08	5.75
STR – LF	2.95	0.22	13.2	−024	0.09	-2.75^{ns}

Note. SE = standard error; HF = high frequency words; LF = low frequency words; PW = pseudowords; S PW = letter-substitution; STR = illegal letter strings; ns = not significant.

although the fastest responses (first black point in each panel) are always associated with a relatively high proportion of incorrect activation, this pattern is enhanced for both pseudowords and low frequency words. Interestingly, the proportion of incorrect activations markedly decreased across quintiles for all categories, except for low-frequency words, which showed a more constant rate of incorrect activations across latencies. As shown in Fig. 2 (lower panels), the drop in the proportion of incorrect activation occurring between the first and the second quintile is reduced for low frequency words.

This pattern was complemented by CAFs analysis. The interaction revealed that: a) high frequency words and illegal letter strings were consistently categorized with a very high accuracy, irrespectively of response latency (see the flat trends represented in Fig. 2, lower panels, grey lines); b) pseudowords and letter-substitution pseudowords highlighted the presence of fast impulsive errors, as indicated by the steep increase in the proportion of accurate responses between the first and the second quintile. Notably, for low frequency words, accuracy tended to decrease across quintiles (see Fig. 2, lower panels, grey line), suggesting that, for these items, participants failed to correct inaccurate activations, irrespectively of their latency.

3.3. Post-hoc similarity analysis

To validate the different role of word-likeness across measures of PMTs and MTs, we a posteriori decided to further test the relation between stimuli properties and chronometric measures by applying a Representational Similarity Analysis (RSA) approach to our data (Kriegeskorte, Mur, & Bandettini, 2008). RSA is a statistical method that assesses second-order isomorphism. Instead of directly assessing the relationship between two measures, RSA computes a similarity metric within each measure and then compares the computed similarities. Despite being mainly popular in neuroimaging literature, its focus on second-order isomorphisms (i.e., similarity) makes this technique an extremely flexible tool to measure correspondences between different levels of analysis.

In the present context, we looked for isomorphism between stimulus word-likeness and PMTs/MTs. For word-likeness, a composite index considering both OLD20 and mean bigram frequency was used. The two measures—after being scaled—were considered as the axes of a bidimensional space and each stimulus was represented in this space as a point with its values of OLD20 and mean bigram frequency as coordinates. Then, the similarity matrix among stimuli was built by calculating the Euclidean distance among all the possible pairs of stimuli. For PMTs, within each participant, PMTs were ordered and then ranked from the fastest to the slowest. Then, ranks were averaged by stimuli and these values were used to calculate the Euclidean distance among all the possible pairs of stimuli. Ranks were preferred to raw data to overcome any distortion of the distance measures introduced by the potential non-linearity of the chronometric measures' distribution. Finally, the second-order similarity between the word-likeness similarity matrix and the PMTs similarity matrix was measured via a Spearman correlation between the two matrices. The same procedure was also applied to MTs to measure its similarity with that of word-likeness. In case the correlation coefficient was notable, its significance was tested by means of random permutations. Specifically, we calculated the Spearman correlation between the word-likeness similarity matrix and a

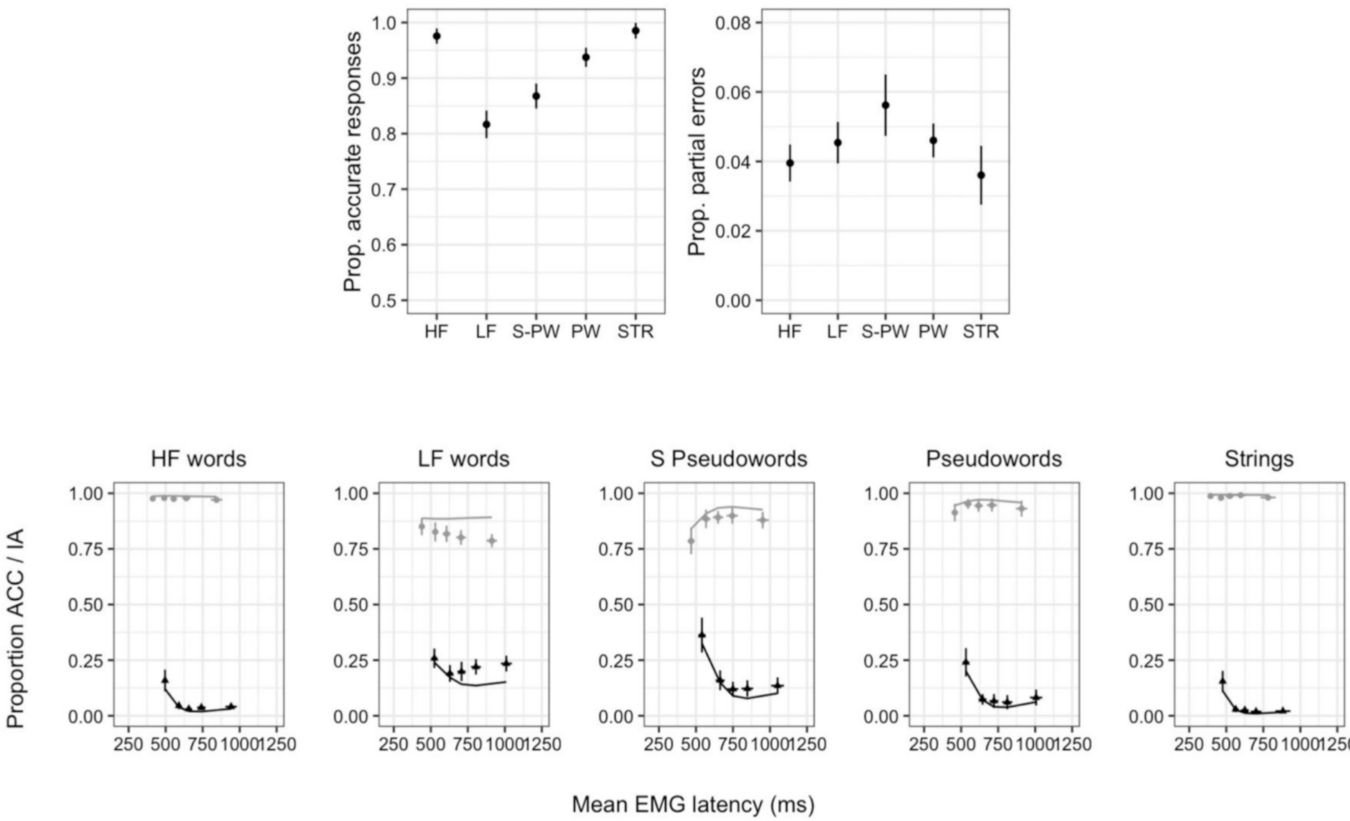

Fig. 2. Results for accuracy and partial errors (upper panels) and for CIAFs and CAFs (lower panels).
Note. Upper panels: HF = high frequency words; LF = low frequency words; S-PW = letter-substitution pseudowords; PW = pseudowords; STR = illegal letter strings. For both plots, error bars display 95 % confidence intervals, adjusted for within-participants variables following Morey (2008). Lower panels: Results for CIAFs (black) and CAFs (grey) for each stimulus category. On the y-axis, proportion of correct responses (ACC, for CAFs) and incorrect activations (IA, for CIAFs) are represented. Points represent empirical means, with vertical error bars highlighting 95 % confidence intervals, adjusted for within- participants variables following Morey (2008). Horizontal error bars reflect 95 % confidence interval of the average reaction time (RT) within each quintile. Lines represent models' predicted means.

Table 6
Fixed effects for generalized LME models on CIAFs and CAFs.

	CIAF				CAF			
Fixed Effects	Est	SE	z	p	Est	SE	z	p
Intercept	−3.28	0.12	−25.73	< 0.001	4.33	0.16	26.75	< 0.001
LF words	1.66	0. 09	16.75	< 0.001	−2.27	0.13	−16.78	< 0.001
1-L Pseudowords	1.40	0.10	13.99	< 0.001	−1.95	0.13	−14.45	< 0.001
Pseudowords	0.67	0.10	6.57	< 0.001	−1.09	0.14	−7.78	< 0.001
Illegal letter strings	−0.45	0.11	−4.04	< 0.001	0.58	0.16	3.56	< 0.001
Quintile, linear	−100.60	3.37	−29.76	< 0.001	−13.10	4.92	−2.66	0.007
Quintile, quadratic	95.01	3.37	28.13	< 0.001	−12.97	5.38	−2.40	0.01
LF words X Quintile, linear	61.22	4.11	14.87	< 0.001	16.05	5.87	2.73	0.006
1-L Pseudowords X Quintile, linear	−0.88	4.48	−0.19	0.84	73.64	6.47	11.38	< 0.001
Pseudowords X Quintile, linear	7.35	5.42	1.35	0.17	33.49	7.06	4.73	< 0.001
Illegal letter strings X Quintile, linear	−41.19	5.76	−7.15	< 0.001	9.73	7.80	1.24	0.21
LF words X Quintile, quadratic	−65.10	4.55	−14.30	< 0.001	16.88	7.83	2.15	0.03
1-L Pseudowords X Quintile, quadratic	−24. 40	5.82	−4.19	< 0.001	−32.53	6.72	−4.83	< 0.001
Pseudowords X Quintile, quadratic	−2. 41	4.80	−0.50	0.61	−29.79	6.92	−4.30	< 0.001
Illegal letter strings X Quintile, quadratic	11.90	6.22	1.91	0.05	−1.18	8.53	−0.13	0.88

permutated version of the chronometric similarity matrix (randomly permutating rows and columns). This operation was repeated 1000 times, and the *p*-value was evaluated as the proportion of random permutations that resulted in a larger correlation coefficient than the observed one.

For the analysis involving PMTs, the results revealed a correlation between how similar were in terms of word-likeness, and the similarity of behavioral response they elicited (*rho* = 0.27), which was statistically significant ($p < .001$). Differently, in the analysis involving MTs the correlation coefficient was negligeable (*rho* = −0.01). These results further indicate that the sensitivity to word-likeness is bounded to PMTs. The results of the analyses are represented in Fig. 3.

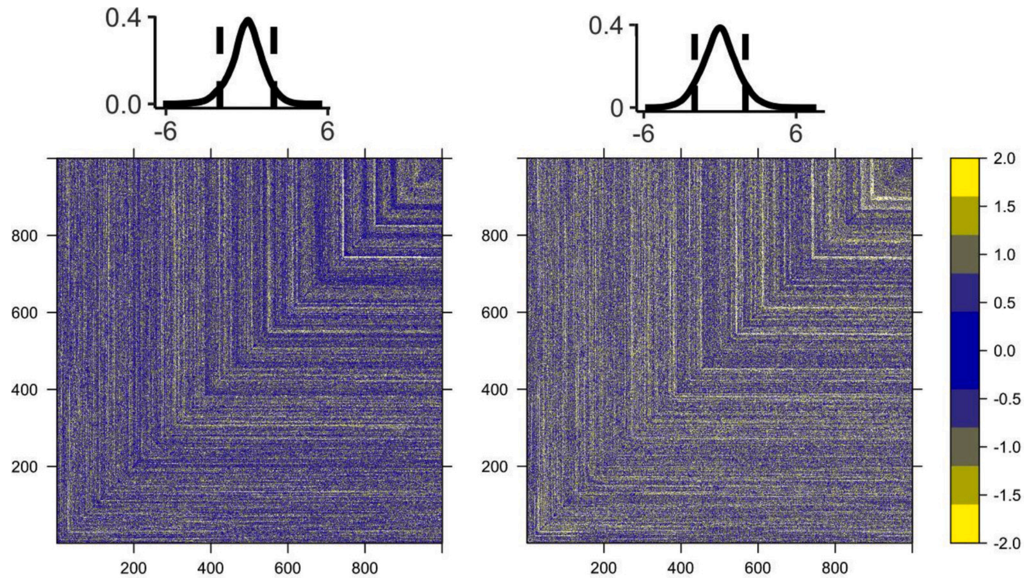

Fig. 3. Representation of the results of the similarity analysis between stimulus word-likeness and PMTs (left column) and MTs (right column).
Note. Visual representation of the similarity between stimulus word-likeness and PMTs/MTs. Euclidean distances (of word-likeness, PMTs, and MTs) used in the similarity analysis were first transformed into z-scores. Then, z-scores for word-likeness were subtracted from those of PMTs and MTs, thus obtaining two matrices of differential scores, which are respectively represented in the left and in the right matrix of the figure (numbers on the x- and y-axis represent stimuli). Shades of blue represent differential scores close to zero (i.e., higher similarity between word-likeness and chronometric measure), whereas shades of yellow represent differential scores farther from zero (i.e., lower similarity between word-likeness and chronometric measure). Note that in the two matrices, only scores between 2 and − 2 are represented, which constitute the vast majority of all the scores – see the two density distributions plotted in the upper part of the figure (PMT z-scores minus word-likeness z-scores on the left, MT z-scores minus word-likeness z-scores on the right), in which the two vertical dashed lines mark 2 and − 2 on the x-axis. The greater blueness of the left matrix visually represents the higher similarity of word-likeness with the PMTs, than with the MTs.

4. Discussion

The experiment investigated whether memory-based recognition relies on qualitatively different decisional dynamics depending on the availability of positive evidence toward the to-be-recognized stimulus. Using a visual lexical decision task, we exploited the lexicality effect—that is, the difference between word and nonword responses. Using different categories of stimuli that followed a word-likeness gradient, we isolated the manipulation of interest—that is, the presence vs absence of positive evidence—from the type of stimulus/response—that is, word vs nonword.

Chronometrically, RTs varied as a function of word-likeness, replicating well-established findings (e.g., Peressotti et al., 1995; Ratcliff et al., 2004; also, cf. Balota & Chumbley, 1984). Specifically, RTs were faster for the two extreme categories of the gradient, that are high-frequency words and illegal letter strings, with the latter showing the fastest responses (e.g., Berberyan et al., 2021; Marinelli et al., 2014). Letter-substitution pseudowords yielded the slowest RTs, consistent with their inherent ambiguity (highly word-like nonwords). Low-frequency words and pseudowords showed intermediate RTs and, intriguingly, did not differ from one each other. PMTs paralleled RTs and the results of the similarity analysis further strengthened the association between PMTs and word-likeness, which in this case was parametrically measured. Differently, MTs were insensitive to word-likeness: As predicted, high- and low-frequency words, and illegal letter strings displayed comparable MTs, and these were faster than those displayed by pseudowords and letter-substitution pseudowords. Additionally, pseudowords and letter-substitution pseudowords did not differ in MTs.

The pattern of MTs hence clearly diverges from the one displayed by PMTs. This differentiation speaks against the view of MT as a propagation of PMT (e.g., Servant, Logan, Gajdos, & Evans, 2021), and in favor of a partial dissociation of the decisional components modulating the two chronometric indexes, at least in the context of a recognition task (Scaltritti, Greatti et al., 2023; Scaltritti et al., 2024; Kamari Songhorabadi et al., 2025). In fact, while PMT is sensitive to word-likeness as well as to word frequency and overall stimulus lexicality (cf., e.g., Scaltritti, Greatti et al., 2023)—all phenomena that have been associated to the rate with which evidence accumulates toward response alternatives (e.g., Ratcliff et al., 2004)—MT is not. Instead, MT seems sensitive to the availability of positive evidence toward the to-be-delivered response. This sensitivity seems a categorical one, as no difference emerges among stimuli belonging to the same pole of the dimension discriminating items as a function of the presence (i.e., high- and low-frequency words, and illegal letter strings) vs absence (i.e., pseudowords, letter-substitution pseudowords) of positive evidence.

The comparison between pseudowords and low-frequency words, and the one between pseudowords and letter-substitution pseudowords, are particularly informative, as they reveal a sort of "double-dissociation" pattern. On the one hand, pseudowords and low-frequency words showed indistinguishable PMTs but different MTs. On the other, pseudowords and letter-substitution pseudowords showed indistinguishable MTs, but different PMTs. We will consider the comparison between pseudowords and low-frequency words at first. The indistinguishable PMTs suggest that, at least during the premotor part of the decision-making process, both require comparable processing times. The similarity is supported by the CAFs-CIAFs profiles (Ramdani et al., 2015, 2021; Scaltritti et al., 2024), which show a relatively high tendency for incorrect response activations in both categories, even in slower responses. Notably, for low-frequency words, incorrect activations were rarely counteracted by action control processes (e.g., van Den Wildenberg et al., 2010), as reflected in the relatively constant proportion of accurate responses across quintiles (see CAFs in Fig. 2). The absence of a clear involvement of action control suggests that, when errors occurred, participants genuinely misidentified low-frequency words as nonwords. This observation, coupled with the comparable PMTs yielded by the two categories of stimuli, suggests that low-frequency words and pseudowords are relatively similar in terms of word-likeness and, hence, in terms of the duration of evidence accumulation process (cf. Balota & Chumbley, 1984; Diependaele, Brysbaert, & Neri, 2012). Differently, pseudowords and low-frequency words clearly differ in terms of the availability of positive evidence, which is only present for low-frequency words, and might therefore underlie their difference in MTs.

A similar reasoning applies to the comparison between pseudowords and letter-substitution pseudowords. These two categories of stimuli clearly offer a different amount of evidence toward the two response alternatives, with letter-substitution pseudowords being harder to recognize than pseudowords. Indeed, letter-substitution pseudowords are the most word-like nonwords, as testified by their slower PMTs, highest proportion of partial errors (cf. Scaltritti et al., 2021; Scaltritti, Greatti et al., 2023), and higher proportion of fast impulsive incorrect activations (as revealed by CIAFs, Fig. 2), which are only partially corrected before the issuing of the final response (leading to lower accuracy for fast responses in the CAFs, Fig. 2). Despite all these differences, pseudowords and letter-substitution pseudowords were indistinguishable in terms of MT, possibly due to the lack of positive evidence in support of the responses to both categories.

A final comparison deserving consideration is the one between words and illegal letter strings. Specifically, relative to low-frequency words, both high-frequency words and illegal letter strings showed faster PMTs (and RTs, e.g., Berberyan et al., 2021; Marinelli et al., 2014) and an extremely high rate of error correction, as revealed by the inspection of the CAFs/CIAFs profile. Therefore, low-frequency words, on the one hand, and high-frequency words and illegal letter strings, on the other hand, show clearly different decisional dynamics, possibly because of their different level of discriminability related to their level of word-likeness. However, these three categories were indistinguishable in MTs, likely because all benefit from the presence of positive evidence supporting the response.

All the above considerations point to a qualitative difference in the decisional features shaping PMTs and MTs, challenging traditional lexical decision models. Notably, classic frameworks distinguish between local activation (i.e., activation of a specific lexical representation) and global activation (i.e., distributed activation across multiple lexical entries) (e.g., Coltheart et al., 2001; Grainger & Jacobs, 1996; Norris, 2006; see also Yap et al., 2015). In this view, pseudowords primarily elicit global activation, with little or no local activation due to the absence of a corresponding lexical entry. Therefore, one might hypothesize that local activation indexes the presence of positive evidence (the main driver of word responses), whereas global activation reflects its absence (the main driver of pseudowords responses), thus accounting for the current MT findings. However, in this framework, global activation scales with word-likeness: the more word-like the pseudoword, the greater the global activation, and consequently, the more difficult it becomes to reject it as a nonword. This prediction aligns with the graded modulations observed in PMTs (letter-substitution pseudowords $<$ pseudowords $<$ illegal letter strings), but fails to explain the pattern detected in MTs, which varies categorically as a function of the presence versus absence of positive evidence. In other words, the distinction between the presence vs absence of positive evidence in shaping the response cannot be fully reduced to the local/global activation dichotomy, as global activation remains tied to word-likeness, a dimension that does not appear to be necessarily involved in the mere categorical distinction between decisions made on the basis of the presence vs absence of positive evidence.

More generally, this qualitative difference in the decisional features driving the modulation of PMT (word-likeness) and MT (presence of positive evidence) revealed by our results challenges models positing a single decisional variable as the determinant of lexical decision performance (e.g., Ratcliff et al., 2004). In this perspective, it remains unclear how distinct portions of response latencies could differentially reflect sensitivity to word-likeness (PMT) vs presence of positive evidence (MT). Conversely, the divergence in factors shaping PMT and MT appears more consistent with models assuming different weighting of evidence for words and nonwords (e.g., Wagenmakers et al., 2004) or even (partially) different processes underlying word and nonword responses (e.g., Brown & Heathcote, 2008; Davis, 2010). From a computational standpoint, the hypothesis of separate evidence accumulation processes feeding distinct accumulators—one for each decision alternative—also provides a closer approximation to brain physiology, where neurons represent evidence in favor of specific choices (Shadlen & Kiani, 2013).

Notably, in our view, the critical difference between words and pseudowords processing lies in the type of evidence that can be collected—i.e., presence vs absence of positive evidence—, rather than in the type of response that has to be delivered—i.e., word vs nonword. Possibly, in the absence of positive evidence for any response alternative, the decision system may initiate a response before reaching a final commitment, under the influence of urgency signals that evolve over time (Cisek, Puskas, & El-Murr, 2009). This could delay the commitment toward a decision until after response execution has started, with the residual lack of positive evidence lengthening execution times. In this scenario, an extended MT would provide additional opportunity for the evaluation of the ongoing decision. Note that when positive evidence is lacking, elapsed time itself might be informative (cf. Dufau et al., 2012), potentially increasing the likelihood of a nonword response as a function of the time spent without collecting conclusive positive evidence toward a response.

Alternatively, one may assume a late verification process that evaluates the stimuli against a set of activated candidates. A similar process was proposed by Paap, Newsome, McDonald, and Schvaneveldt (1982) in the context of their activation-verification model of word recognition (see also Martínez-Tomás, Baciero, Lázaro, & Hinojosa, 2025; Perea, Rosa, & Gómez, 2005; Yap et al., 2015). In this model, after the analysis of the stimulus leads to the activation of a set of lexical candidates from memory, a late verification process would lead to the recognition of a single entry based on the degree of fit between the stimulus and a specific candidate. A word is recognized when the fit exceeds a certain decision criterion (note that, although originally developed for word recognition, this model can, in principle, extend to other memory-based recognition processes). Under the assumption that verification processes percolate into motor-response execution, in case of a readable nonword the commitment toward the correct response might require an exhaustive search that ultimately rejects all the activated lexical candidates. Such a process might take longer for nonwords than for words—where verification terminates upon identifying a match—and illegal letter string—for which the decision can be based on the detection of orthotactic violations and illegal grapheme sequences, without any (exhaustive) comparison between the stimulus and the activated candidates. Despite this interpretation is intriguing, it seems unlikely. In our experiment, pseudowords and letter-substitution pseudowords differed in terms of OLD20, with the latter being associated to more lexical candidates than the former. Within the perspective of verification processes percolating into response execution, this feature fits poorly with the comparable MTs detected across these two categories of stimuli. In order to measure comparable MTs, the two categories of readable nonwords would need to activate (approximately) the same number of lexical candidates, which does not seem to be the case, given the different psycholinguistic properties of the two categories of nonwords. In sum, if the duration of the verification process is a function of the number and the strength of activated lexical candidates, we should have found a difference in terms of MTs between pseudowords and letter-substitution pseudowords, which was clearly not the case. Alternatively, we must conclude that the duration of verification processes is insensitive to the number of activated candidates, but this would be in contrast with Paap et al.'s original proposal.

A final hypothesis ascribes the lexicality-induced MT modulation to the relative difference between response alternatives when positive evidence is present vs absent (cf. Abrams & Balota, 1991). Specifically, when no positive evidence is available, the system might reduce the relative difference between the two response alternatives by applying a penalty weight to the currently more plausible alternative. This mechanism would temporarily reduce the relative difference between the two alternatives, thus delaying the final commitment toward a decision until the information extracted from the stimulus offers a “smoking gun” toward any response alternatives. Optimally, such a weighing mechanism would come into play after a certain amount of time, ensuring that some degree of evidence has already been gathered. Potentially, this mechanism can begin after (or unfold on par with) motor-response execution, thus trading the cost of a longer processing time with the benefit of having more space for the final commitment (for a recent proposal of a smoothing mechanism at the level of motor preparation, but with a single decisional process, see Dendauw et al., 2024).

In conclusion, our findings indicate that PMT and MT are modulated by (partially) different decisional dynamics, with MT being specifically sensitive to the availability of positive evidence. While the mechanism underlying this process remains to be determined, our data suggest a qualitative distinction between responses to stimuli whose recognition can benefit of the presence of positive evidence toward the to-be-delivered response—such as words and illegal letter strings—and those that do not—such as pseudowords and letter-substitution pseudowords. More broadly, this points to a fundamental differentiation in the decisional processes underlying memory-based recognition in the presence vs absence of supporting evidence, particularly with respect to the implementation of the corresponding action.

CRediT authorship contribution statement

Elisa Fiora: Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Michele Scaltritti:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Simone Sulpizio:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to proofread the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

The research was funded by the European Union–Next Generation EU – PRIN 2022 PNRR (DD 1409, 14/09/22) – PNRR – M4 – C2 – INV1.1 – PRIN – Project Title [Functional characterization of decisional components in motor responses for young and older adults] – Grant Number [P2022C39ZH] – CUP [E53D23019540001].

Declaration of competing interest

Nothing to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2025.106068.

Data Availability

Materials and raw data are available at https://osf.io/zt8eh

References

Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2*, 153–157.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4*, 439–461.

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*, 340–357.

Barca, L., & Pezzulo, G. (2012). Unfolding visual lexical decision in time. *PLoS One, 7*(4), Article e35932.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Berberyan, H. S., van Rijn, H., & Borst, J. P. (2021). Discovering the brain stages of lexical decision: Behavioral effects originate from a single neural decision process. *Brain and Cognition, 153*, Article 105786.

Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology, 71*, 9–15.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153–178.

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments= A tutorial of power analysis with reference tables. *Journal of Cognition, 2*, 16.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in chang- ing conditions: The urgency-gating model. *Journal of Neuroscience, 29*, 11560–11571.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., & Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain, 125*, 1054–1069.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*(3), 713–758.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source tool- box for analysis of single-trial EEG dynamics including inde- pendent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.

Dendauw, E., Evans, N. J., Logan, G. D., Haffen, E., Bennabi, D., Gajdos, T., & Servant, M. (2024). The gated cascade diffusion model: An integrated theory of decision making, motor preparation, and motor execution. *Psychological Review, 131*, 825–857.

Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology, 3*, 348.

Dufau, S., Grainger, J., & Ziegler, J. C. (2012). How to say “no” to a nonword: A leaky competing accumulator model of lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1117–1128.

E-Prime (Version 3.0.3.214) [Computer software]. Pittsburgh, PA: Psychology Software Tools.

Eriksen, C. W., Coles, M. G., Morris, L. R., & O’hara, W. P. (1985). An electromyographic examination of response competition. *Bulletin of the Psychonomic Society, 23*, 165–168.

Fluchère, F., Burle, B., Vidal, F., van den Wildenberg, W., Witjas, T., Eusebio, A., & Hasbroucq, T. (2018). Subthalamic nucleus stimulation, dopaminergic treatment and impulsivity in Parkin- son’s disease. *Neuropsychologia, 117*, 167–177.

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *Quarterly Journal of Experimental Psychology, 67*, 2455–2466.

Goslin, J., Galluzzi, C., & Romani, C. (2014). Phonitalia: A phonological lexicon for Italian. *Behavior Research Methods, 46*, 872–886.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*(3), 518–565.

Hasbroucq, T., Possamaï, C. A., Bonnet, M., & Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: An electromyographic study in humans. *Psychophysiology, 36*, 522–526.

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*, 157–183.

Kamari Songhorabadi, S., Sulpizio, S., & Scaltritti, M. (2025). Dissociating premotor and motor components of response times: Evidence of independent decisional effects during motor-response execution. *Psychonomic Bulletin & Review, 32*, 1890–1900.

Kelly, M. C., & Tucker, B. V. (2022). The recognition of spoken pseudowords. *Language, Cognition and Neuroscience, 37*, 1169–1190.

Keuleers, E. (2013). *vwr: Useful functions for visual word recognition research* (R Package Version 0.3.0). https://CRAN.R-project.org/ package=vwr.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2*, 249.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.

Lenth R, Piaskowski J (2025). Emmeans: Estimated marginal means, aka least-squares means. R package version 2.0.0, https://rvlenth.github.io/emmeans/.

Liu, J., & Liu, Q. (2016). Use of the integrated profile for voluntary muscle activity detection using EMG signals with spurious background spikes: A study with incomplete spinal cord injury. *Biomedical Signal Processing and Control, 24*, 19–24.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213.

Luce, R.D. Response times: Their role in inferring elementary mental organization (Oxford Univx. Press, New York, 1986).

Marinelli, C. V., Traficante, D., & Zoccolotti, P. (2014). Does pronounceability modulate the letter string deficit of children with dyslexia? A study with the rate and amount model. *Frontiers in Psychology, 5*, 1353.

Martínez-Tomás, C., Baciero, A., Lázaro, M., & Hinojosa, J. A. (2025). What do pseudowords tell us about word processing? An overview. *Frontiers in Language Sciences, 4*, Article 1504770.
Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Tutorial in Quantitative Methods for Psychology, 4*, 61–64.
Nenadić, F., & Tucker, B. V. (2020). Computational modelling of an auditory lexical decision experiment using jTRACE and TISK. *Language, Cognition and Neuroscience, 35*(10), 1326–1354.
Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*, 327.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.
Ollman, R. T. (1977). Choice reaction time and the problem of distinguishing task effects from strategy effects. In S. Domic (Ed.), *Attention & performance VI* (pp. 99–113). Hillsdale, NJ: Erlbaum.
Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89*, 573–594.
Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics, 67*, 301–314.
Peressotti, F., Job, R., Rumiati, R., & Nicoletti, R. (1995). Levels of representation in word processing. *Visual Cognition, 2*, 421–450.
R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation of statistical computing. http://www.R-project.org/.
Ramdani, C., Carbonnell, L., Vidal, F., Béranger, C., Dagher, A., & Hasbroucq, T. (2015). Dopamine precursors depletion impairs impulse control in healthy volunteers. *Psychopharmacology, 232*, 477–487.
Ramdani, C., Sagui, E., Schmid, B., Castagna, O., Davranche, K., Vidal, F., & Hasbroucq, T. (2021). Action monitoring fails when motor execution is too fast: No time for correction. *Journal of systems and integrative Neuroscience, 7*.
Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.
Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*, 487–494.
Scaltritti, M., Giacomoni, F., Job, R., & Sulpizio, S. (2023). Redefining the decisional components of motor responses: Evidence from lexical and object decision tasks. *Journal of Experimental Psychology: Human Perception and Performance, 49*, 835–851.
Scaltritti, M., Greatti, E., & Sulpizio, S. (2023). Electrophysiological evidence of discontinuities in the propagation of lexical decision processes across the motor hierarchy. *Neuropsychologia, 188*, Article 108630.
Scaltritti, M., Greatti, E., & Sulpizio, S. (2024). Decisional components of motor responses are not related to online response control: Evidence from lexical decision and speed-accuracy tradeoff manipulations. Memory & Cognition, 1–15.
Scaltritti, M., Job, R., & Sulpizio, S. (2021). Selective suppression of taboo information in visual word recognition: Evidence for cognitive control on semantics. *Journal of Experimental Psychology: Human Perception and Performance, 47*, 934–945.
Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology: General, 150*, 2435.
Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron, 80*(3), 791–806.
Shiffrin, R. M. (2003). Modeling memory and perception. *Cognitive Science, 27*(3), 341–378.
Shulman, H. G., & Davidson, T. C. (1977). Control properties of semantic coding in a lexical decision task. *Journal of Verbal Learning and Verbal Behavior, 16*, 91–98.
Stone, G. O., & Van Orden, G. C. (1992). Resolving empirical inconsistencies concerning priming, frequency, and nonword foils in lexical decision. *Language and Speech, 35*, 295–324.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*, 550–592.
van Den Wildenberg, W. P., Wylie, S. A., Forstmann, B. U., Burle, B., Hasbroucq, T., & Ridderinkhof, K. R. (2010). To head or to heed? Beyond the surface of selective action inhibition: A review. *Frontiers in Human Neuroscience, 4*, 222.
Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., & Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron, 55*, 143–156.
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*, 140–159.
Wagenmakers, E. J., Steyvers, M., Raaijmakers, J. G., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology, 48*, 332–367.
Weiss, A. D. (1965). The locus of reaction time change with set, motivation and age. *Journal of Gerontology, 20*, 60–64.
Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English lexicon project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53–79.
Yap, M. J., Sibley, D. E., Balota, D. A., Ratcliff, R., & Rueckl, J. (2015). Responding to nonwords in the lexical decision task: Insights from the English lexicon project. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 597–613.
Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review, 15*, 971–979.