AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

2020, Vol. 46, No. 11, 2034–2048
http://dx.doi.org/10.1037/xlm0000942

Should I Stay or Should I Go? An ERP Analysis of Two-Choice Versus Go/No-Go Response Procedures in Lexical Decision

Marta Vergara-Martínez
University of Valencia

Pablo Gomez
California State University, San Bernardino

Manuel Perea
University of Valencia and Nebrija University

Prior behavioral experiments across a variety of tasks have typically shown that the go/no-go procedure produces not only shorter response times and/or fewer errors than the two-choice procedure, but also yields a higher sensitivity to experimental manipulations. To uncover the time course of information processing in the go/no-go versus the two-choice procedures during visual word recognition, we examined the impact of a lexical factor (word frequency) in a lexical-decision task by tracking event-related potential (ERP) waves. If the differences across response procedures influence relatively early lexical processing stages, we would expect word frequency to induce differences across tasks in the early epochs of the ERP. Alternatively, if the differences across response procedures only occur at a postaccess response selection stage, we would only expect differences across procedures in late time windows of the ERP. Results showed that the word-frequency effect occurred earlier (starting around 200 ms poststimuli) in the go/no-go than in the two-choice response procedure. These results support the view of a largely flexible cognitive network in which a subtle manipulation of the response procedure can affect early components of processing.

Keywords: lexical decision, response procedure, word frequency, visual word recognition, ERPs

In cognitive neuroscience studies of language processing, researchers often have explored brain and mental processes within the chronometric tradition of cognitive psychology. In this endeavor, they have indistinctly employed two response procedures: go/no-go (GNG) and two-choice (2C). In the GNG procedure, participants are instructed to respond to a category of stimuli (e.g., words in a word/nonword discrimination task [i.e., lexical decision]) and to refrain from responding to the other category (e.g., nonwords). In the 2C procedure, participants are instructed to respond not only to the stimuli from one category but also to the stimuli from the other category (e.g., right-hand response for words and left-hand response for nonwords in a lexical-decision task).

In his seminal experiments, Donders (1868/1969) observed that response times (RTs) were longer in 2C than in GNG tasks (see also Broadbent & Gregory, 1962; Callan, Klisz, & Parsons, 1974; Gottsdanker & Shragg, 1985; Hackley, Schäffer, & Miller, 1990). This difference has been traditionally interpreted in terms of ancillary response processes: the addition of a "response selection" stage in the choice task would slow down performance, but the underlying central processes (e.g., lexical) would be unaffected by the response procedure (see Gordon, 1983, for a discussion in the context of lexical processing). However, the evidence from prior behavioral experiments is difficult to reconcile with the response selection hypothesis, as they have often reported that the GNG procedure is more sensitive to experimental manipulations than the 2C procedure. A number of lexical decision experiments have shown greater orthographic, lexical, and semantic effects in the GNG than in the 2C version of the lexical-decision task (e.g., orthographic processing: Perea, Mallouh, & Carreiras, 2014; lexical processing: Hino & Lupker, 1998, 2000; semantic processing: Perea & Rosa, 2003; Perea, Rosa, & Gómez, 2002). Furthermore, a larger sensitivity of the GNG over the 2C procedure has also been reported in other behavioral tasks (e.g., target detection in a scene: Bacon-Macé, Kirchner, Fabre-Thorpe, & Thorpe, 2007; same-different matching task: Grice & Reed, 1992; semantic categorization: Siakaluk, Buchanan, & Westbury, 2003).

The apparent gains in the detectability of a number of phenomena with the GNG procedure in behavioral experiments do suggest

This article was published Online First July 30, 2020.

Marta Vergara-Martínez, Interdisciplinary Research Structure- ERI-Lectura, University of Valencia; Pablo Gomez, Department of Psychology, California State University, San Bernardino; Manuel Perea, Interdisciplinary Research Structure- ERI-Lectura, University of Valencia, and Cognitive Science Center - C3, Nebrija University.

The research reported in this article has been partially funded by Grant PSI2017-86210-P (Manuel Perea) from the Spanish Ministry of Science, Innovation and Universities.

The raw data for all analyses are provided in the anonymized OSF website (Vergara-Martínez & Gomez, 2020; https://osf.io/gpc46/?view_only=b4885a92902c4324a7f939bbb414479f).

Correspondence concerning this article should be addressed to Marta Vergara-Martínez, Interdisciplinary Research Structure- ERI-Lectura, University of Valencia, Avenue Blasco Ibáñez, 21, 46010 Valencia, Spain. E-mail: Marta.Vergara@uv.es

that response demands in the GNG and 2C procedures may affect core components of processing rather than merely ancillary processes such as response selection or motor response execution. To examine this issue, we analyzed the time course of the most extensively studied lexical factor in visual word recognition: word frequency (WF; high-frequency words are identified faster and more accurately than low-frequency words; see Forster & Chambers, 1973; Preston, 1935; Rubenstein, Garfield, & Millikan, 1970; Solomon & Postman, 1952, for early evidence). We did so by tracking the event-related potential (ERP) signature of WF in GNG and 2C versions of the lexical-decision task (i.e., the most commonly used task in the field of visual-word recognition). In this setup, the WF effect was used as a marker for the activation of lexical properties. As initially reported by Hino and Lupker (1998, 2000; see also Gomez, Ratcliff, & Perea, 2007), the magnitude of the WF effect in the RTs (and error rates) is systematically greater in the GNG than in the 2C version of the lexical-decision task. Therefore, the ERP responses will allow us to track in detail whether the temporal dynamics of lexical access (as reflected by the ERP WF effect) change as a function of each task procedure.

WF ERP Effects and Task Demands

The brain's electrophysiological response to WF has been the topic of a large number of studies (see Hauk & Pulvermüller, 2004; Laszlo & Federmeier, 2014, for reviews). As a benchmark of lexical-semantic processing, many of these studies focused on the impact of WF in the N400 ERP component (Kutas & Federmeier, 2011). Low-frequency words elicit larger N400 amplitudes than high-frequency words (Barber, Vergara, & Carreiras, 2004; Kutas & Federmeier, 2011; Rugg, 1990; Vergara-Martínez, Comesaña, & Perea, 2017). This difference is assumed to reflect the strength of the memory traces regarding the specific characteristics of a word. However, it is not rare to observe WF effects in earlier time windows, both in EEG or magnetoencephalography (MEG) measures (Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006; Hauk & Pulvermüller, 2004; see Chen, Davis, Pulvermüller, & Hauk, 2015, for similar effects with MEG). Indeed, the results regarding the latency of the WF effect are rather heterogeneous. This is not surprising when one looks at the differences among the experimental setting within each experiment. For example, some early factorial experiments reporting early WF effects (Assadollahi & Pulvermüller, 2003; Sereno, Rayner, & Posner, 1998) did not control for potentially confounding variables which were later shown to impact very early stages of visual word recognition (e.g., bigram frequency, orthographic neighborhood; Laszlo & Federmeier, 2014). More recent regression designs that aim to overcome some of the limitations of factorial designs (see Hauk et al., 2006; Hauk, Pulvermüller, Ford, Marslen-Wilson, & Davis, 2009) also reported early latencies of WF (~150 ms). However, these studies have systematically used a very fast presentation rate of each item (100 ms), a factor found to impact the nature of word processing: the latency of WF effects gradually decreases as the presentation rate of words increases (Dambacher et al., 2012).

Importantly, the variability of the WF latency may result not only from potentially relevant characteristics such as the frequency range of the stimuli under analysis, the underestimation of highly correlated variables (e.g., bigram frequency), or differences in stimulus presentation rates, but also from differences in the nature of the experimental task. Regarding this last point, when looking at the neurophysiological reports of WF in the ERP literature, one can grasp a relation between the probability of obtaining early WF effects (i.e., less than 200 ms) and the type of task that has been employed. For example, most studies that have presented words embedded in sentences report WF effects starting around 300 ms (Dambacher & Kliegl, 2007; King & Kutas, 1998; Osterhout, Bersick, & McKinnon, 1997; Van Petten & Kutas, 1990; but see Penolazzi, Hauk, & Pulvermüller, 2007); however, when the words are presented in isolation and participants are instructed to sort the items into two binary categories where linguistic information is explicitly required (e.g., word vs. nonword; animal name vs. common word, etc.), the reported WF latencies have ranged from very short (<200 ms;: Braun, Hutzler, Ziegler, Dambacher, & Jacobs, 2009; Hauk et al., 2006; Hauk & Pulvermüller, 2004; Proverbio, Vecchi, & Zani, 2004; Sereno et al., 1998; Strijkers, Bertrand, & Grainger, 2015) to longer latencies similar to those experiments with words embedded in sentences (around 300 ms; Barber et al., 2004; Carreiras, Vergara, & Barber, 2005; Laszlo & Federmeier, 2014; Münte et al., 2001; Rugg & Doyle, 1992; Vergara-Martínez & Swaab, 2012).

Overall, one could argue that when the stimuli are processed in binary tasks, the participants' decision-making system optimizes the information available by amplifying the classification-relevant features while disregarding task-irrelevant features. Indeed, some contemporary models of word recognition like the Bayesian reader (Norris & Kinoshita, 2008) implement this intuition, which is also presented in the original diffusion model account of the lexical-decision task (see Figure 4 in Ratcliff, Gomez, & McKoon, 2004). Accordingly, studies that have explicitly manipulated the nature of the stimulus processing found an early latency of lexical factors (e.g., WF) in tasks such as semantic categorization, but not in tasks focused on irrelevant word parameters such as ink color categorization (see Strijkers et al., 2015). Note that in the Strijkers et al. (2015) experiment, the automaticity of processes underlying lexical access was put to test by changing the stimuli criteria on which participants' judgments are required (perceptual or semantic). According to this view, top-down modulation refers to intentionally driving attention to a specific set of features, following the instructions of the experimenter. In the present study, we go a step further: our aim is to assess the impact of a procedural factor—one that is not related to the stimuli, nor to the discrimination at hand—onto the automaticity of lexical access, by having participants performing exactly the same task (lexical-decision task: "is this a word or not?") with exactly the same stimuli (matched across a series of psycholinguistic variables) while changing the specific instructions to respond: GNG and yes/no. Hence this is a novel approach to assess the permeability of visual word processes.

GNG Versus 2C Tasks and Process Models

The processing of word stimuli in the GNG and 2C lexical-decision task is, in principle, quite similar: in the two procedures, participants are asked to classify strings of letters as either words or nonwords and, furthermore, they are asked to make a motor response to indicate that a string is a word. Thus, the only difference is the lack of an overt response to nonwords in the GNG procedure. Using the drift-diffusion model (DDM), Gomez et al. (2007) claimed that the evidence accumulation was constant across the two procedures and that the only difference had to do with the parameter T_{er}, a parameter that encompasses both the encoding and the response execution processes. The model, as implemented in that paper, is agnostic about whether an effect on that parameter reflects changes in the encoding (early processes) versus the response execution (late processes) stages; however, Gomez et al. (2007) interpreted it as a difference in the response execution phase. In this article we reexamine this issue with new evidence and pose the following questions: what are the consequences of the difference in the conflict between the responses to words (always overt in both GNG and 2C procedures) and to nonwords (overt only for the 2C procedure)? How does the subtle difference in the response procedure affect the core (lexical) processes? These are the main questions that we set out to address in the current experiment.

Regarding the first question, the differences in response execution between the two procedures may lead to two different types of response conflict—response conflict is defined as the simultaneous coactivation of incompatible responses. In the GNG procedure, one could boil down each trial to a resolution of the response conflict between executing an overt response and inhibiting it (Braver, Barch, Gray, Molfese, & Snyder, 2001), whereas the 2C procedure requires a selection between two competing overt alternatives. However, according to models on executive control of conflict (Jones, Cho, Nystrom, Cohen, & Braver, 2002; Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003), the nature of conflict between response representations is similar in both 2C and GNG procedures. In Jones et al.'s (2002) computational model of conflict, the difference between the GNG and the 2C procedures is that, unlike the 2C procedure, the decision layer unit corresponding to the no representation in the GNG procedure is not connected to a response execution unit. However, this decision layer unit can become active and suppress activity in the go (yes) decision unit. As a result, the activation of two alternative overt response representations in the 2C procedure would be analogous to the activation of the representation of an overt go response and a hidden no-go response in the GNG procedure. Similarly, the diffusion model of the GNG (Gomez et al., 2007) proposed an implicit decision for the no-go trials even in the absence of an explicit motor response. The similarities between the two tasks regarding conflict detection is reflected in terms of similar activation in the anterior cingulate cortex (a brain region sensitive to response conflict) in the two procedures (Braver et al., 2001; Jones et al., 2002). In short, there is ample consensus that the main difference between the two procedures cannot be described as one of conflict resolution. Instead, a more likely source of differences across procedures is response preparation. Although participants in both procedures prepare either for the go versus no-go response (GNG), or for the yes versus no response (2C), the trial dynamics for a no response within each procedure may have an impact on response preparation. While a yes response immediately terminates the trial (as is also the case for a no response in the 2C procedure), a no-go decision requires waiting until some time has passed (i.e., the deadline to respond). Hence, preparation toward a no response in the 2C procedure (i.e., a head-start) might lead to anticipation errors to word items; on the other hand, these anticipation errors cannot occur in the GNG procedure because there is no overt nonword response that terminates a trial, and participants can recover from a premature decision.

These ideas on response preparation may help shed some light on our second research question: does task procedure affect lexical processes? The differential response preparation requirements and constraints imposed by the 2C versus the GNG procedure might make the latter more sensitive to lexical variables such as WF. There is, indeed, extensive behavioral evidence for this pattern (Gomez et al., 2007; Hino & Lupker, 1998, 2000). However, it is unclear if the modulating role of procedure operates at an early processing component, or if it simply cascades into later components.

Critically, in a binary-response task such as lexical decision, the decision in a given trial results from the interaction between accumulation of evidence, the ability to suppress an incorrect tendency to respond (response conflict), and response preparation. In the case of a familiar, high-frequency word (or an illegal nonword), the accumulation of evidence toward the boundary is fast, and hence a yes (go) response and a no or no-go response would proceed similarly in both procedures. In this case, the premature activation of response representation would not tax one procedure over the other. Importantly, the scenario may be different when encountering a less familiar, low-frequency word or a wordlike nonword. As the accumulation of evidence would be slow, both alternative and incompatible responses would compete against each other, and conflict would arise. Furthermore, the premature activation of any response may lead to different types of errors if conflict resolution does not operate properly. Specifically, the premature activation of no responses would bring different consequences within each procedure. In the 2C procedure, it has the risk of incurring in an overt error (i.e., as the response finishes the trial). On the contrary, no counterpart is expected in the GNG procedure until the deadline (i.e., trial duration) is reached.

All in all, we can infer that the premature preparation of a given response compromises response efficiency to a smaller degree in the GNG than in the 2C procedure and, furthermore, this should be more pronounced for the less familiar, low-frequency words. Following this rationale, the 2C would be more resource demanding than the GNG procedure, as it requires a higher degree of control over response preparation—indeed, developing readers perform dramatically better (fewer errors, faster responses) in the GNG than in the 2C lexical decision (e.g., Perea, Soares, & Comesaña, 2013). Hence, if subtle aspects of information processing (i.e., lexical processing) are tolled to a larger degree in the 2C compared to the GNG, we would expect differences during lexical processing between the two procedures.

Assuming that the WF effect in the ERP responses reflects the difference in activation of lexical representations, its earliest latency can be considered a marker of lexical access (Hauk & Pulvermüller, 2004; see also Sereno et al., 1998). Thus, the differences in the latency and/or magnitude of the WF effect across response procedures would reveal that response procedure (GNG vs. 2C) has an impact on the core processing of meaningful stimuli, an outcome that would be consistent with fully flexible accounts of visual word recognition (see Carreiras, Armstrong, Perea, & Frost, 2014, for a review).

Preview of the Experiment

In the current experiment, we assessed the impact of response procedure on lexical processes with an ERP lexical-decision task in which the stimuli (half were high- and low-frequency words, and the other half were nonwords) were presented in two counterbalanced blocks (one with GNG responses and the other with 2C responses). We presented the same set of words—using counterbalanced lists—in the two response procedures (participants had to execute the same "word" responses in both the GNG and the 2C blocks). Although the same set of nonwords was also presented across the two response procedures, our comparisons focused on word stimuli because, unlike nonwords (refrain from a response in the GNG procedure vs. press "nonword" in the 2C procedure), words require exactly the same explicit overt response in both procedures. Therefore, any changes on the latency and/or size of the WF effect in the ERP responses shall be accounted for exclusively in terms of the differences between the response procedures.

Method

Participants

Twenty undergraduate/graduate students of the University of Valencia (10 women) participated in the experiment in exchange for course credit or for a small gift. All of them were native Spanish speakers with no history of neurological or psychiatric impairment, and with normal (or corrected-to-normal) vision. Ages ranged from 18 to 30 years (mean: 23 years, *SD*: 4.4). All participants were right-handed, as assessed with a Spanish abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971). The data from one participant were discarded because of excessive artifacts in the EEG recording of four peripheral electrodes (T7, T8, P7, P8) and in the reference (right mastoid channel) during the experimental session.

This research was approved by the Research Ethics Committee of the University of València and was in accordance with the declaration of Helsinki. All participants provided written informed consent before starting the experiment.

Materials

We selected a set of 160 five-letter Spanish words from the BuscaPalabras database (Davis & Perea, 2005). Eighty of these words were of high frequency (mean = 64; range = 25–245), and 80 of low frequency (mean = 7; range = 3.5–11). The details of the experimental stimuli are presented in Table 1. As can be seen in Table 1, words in the sets of high- and low-frequency were matched for number of letters, imageability, concreteness, mean positional bigram frequency, and orthographic neighborhood. For the purposes of the lexical-decision task, a matched set of 160 orthographically legal nonwords was carefully created (replacing 2–5 letters from the original words, depending on their length) using Wuggy (Keuleers & Brysbaert, 2010). Four lists of materials were constructed so that each word or nonword appeared once in each list but each time in a different block and task procedure. That is, each target stimulus (e.g., golfo [gulf]) was rotated across the combinations of block and task procedure (e.g., in list 1, it would be presented in the first block [with a GNG procedure], in list 2, it would be presented in the first block [with a 2C procedure], in list 3, it would be presented in the second block [with a GNG procedure], and in list 4, it would be presented in the second block [with a 2C procedure]; see Perea et al., 2014, for a similar procedure). The full list of words and nonwords is presented in the Appendix.

Design

Response procedure (GNG, 2C) and WF (low frequency, high frequency) were manipulated within participants. Half of the participants were presented with the GNG lexical-decision task in the first half of the experiment, whereas the other half were presented with the 2C lexical-decision task. Each participant was given a total of 320 experimental trials: 160 word trials and 160 nonword trials.

Procedure

Participants were seated comfortably in a dimly lit and sound-attenuated chamber. All stimuli were presented on a high-resolution monitor that was positioned at eye level 70 cm in front of the participant. The stimuli were displayed in white lowercase Courier 24-pt font against a dark-gray background where each character subtended about 0.4° of visual angle in height and 0.6° in width. Participants performed a lexical-decision task: they had to decide as accurately and rapidly as possible whether or not the stimulus was a Spanish word. In the GNG version of the lexical-decision task, participants were instructed to press the "YES" button for words, and refrain from responding if the stimulus was not a word. In the 2C version of the lexical-decision task, they pressed one of two response buttons (the YES button or the NO button). The hand used for each type of response was counterbalanced across subjects. RTs were measured from stimuli onset until the participant's response. The sequence of events in each trial was as follows: A fixation cross (+) appeared in the center of the

Table 1
Mean Values of Psycholinguistic Characteristics of Words Across Conditions (SDs in Parentheses) as Provided in the B-Pal Spanish Database (Davis & Perea, 2005)

	No. of letters	LEXESP freq.[a]	B-Pal freq.[a]	N[b]	Imageability[c]	Concreteness[c]	Mean log bigram frequency: Words	Mean log bigram frequency: Pseudowords
HF words	5	64 (43.37)	71.3 (75.11)	1.55 (1.32)	5.12 (.97)	4.77 (.95)	2.55 (0.24)	2.39 (.38)
LF words	5	7.11 (2)	8.22 (6.4)	1.55 (1.32)	4.90 (1.10)	5.15 (1.08)	2.55 (0.24)	2.44 (.33)

Note. LEXESP = Spanish database (Sebastián-Gallés, Martí, Cuetos, & Carreiras, 2000); B-Pal = BuscaPalabras; HF = high frequency; LF = low frequency.
[a] Frequency (freq.) per million. [b] Orthographic neighbors. [c] Range: 1–7.

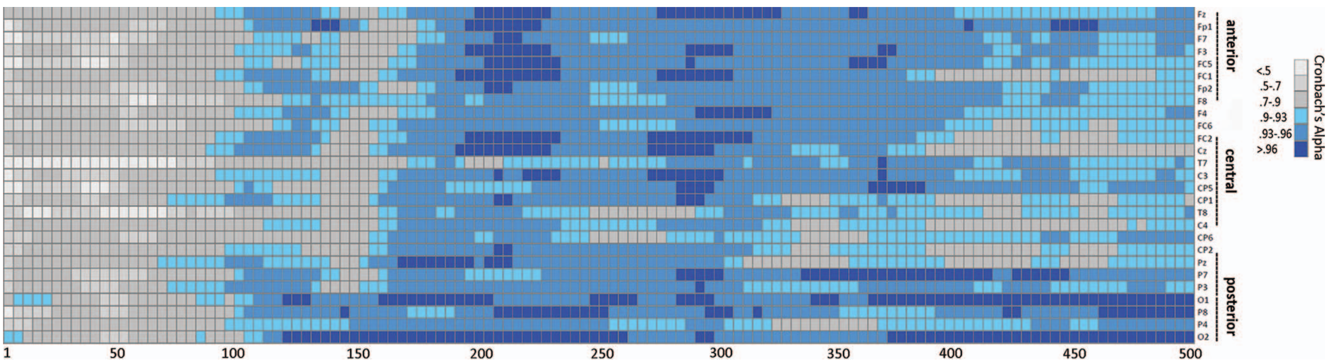

Figure 1. Reliability of averaged voltage measures at each time point and across each electrode location. Raster plots show color-coded Cronbach's alpha with blue indicating greater consistency (.9 or higher). Note that higher Cronbach's alpha values were obtained in temporal intervals and specific electrodes corresponding to N1, P2, N250, P300, and N400. See the online article for the color version of this figure.

screen for 600 ms. This was replaced by a 100-ms blank screen which was in turn replaced by a stimulus word or nonword that remained in the screen for 400 ms. After the participant responded (or after a maximum interval of 1,500 ms had passed), a blank screen of random duration (drawn from a uniform distribution with range 400–800 ms) was presented.

The experimental session was divided in two experimental blocks separated by a 5-min break. Each of the experimental blocks was preceded by 16 practice trials. There were brief 10-s breaks every 40 trials. Every 80 trials, there was a brief pause for resting and impedance checking. To minimize subject-generated artifacts in the EEG signal during the presentation of the experimental stimuli, participants were asked to refrain from blinking and making eye movements from the onset of each trial to the response. Each participant received the stimuli in a different random order. The whole experimental session lasted approximately 45 min.

EEG Recording and Analyses

The EEG was recorded from 29 Ag/AgCl electrodes mounted in an elastic cap according to the 10/20 system, referenced to the right mastoid and rereferenced offline to the averaged signal from two electrodes placed on the left and right mastoids.[1] The EEG recording was amplified and bandpass filtered between 0.01–100 Hz with a sample rate of 250 Hz by a BrainAmp (Brain Products, Gilching, Germany) amplifier. An offline low-pass filter between 0.01 and 20 Hz was applied to the EEG signal. Impedances were kept below 5 kΩ during the recording session. All single-trial waveforms were screened offline for amplifier blocking, drift, muscle artifacts, eye movements, and blinks. These artifacts were detected by means of a semiautomatic data inspection procedure applied to the full set of channels for each participant. The following parameters thresholds were set for automatic detection: gradient (100 μV/ms), amplitude difference (100 μV; 100ms interval), amplitude threshold (−100 μV/100 μV), and low activity (0.5 μV; 100-ms interval). Each epoch was also visually inspected for the presence of any undetected artifact. This was done for a 500-ms epoch with a 100-ms prestimulus baseline. Trials containing artifacts or incorrectly responded to were not included in the average ERPs or in the statistical analyses. These processes led to an average rejection rate of 9.2% (6.4% due to artifact rejection; 2.8% due to incorrect responses). There were no differences in the number of rejections due to artifacts across conditions (all *F*s < 1). Importantly, at least 30 trials were included for each condition in the average ERP data from each participant[2] (mean of the averaged trials per condition across participants: 36.3, *SD*: 2.6). ERPs were averaged separately for each of the experimental conditions, each of the subjects, and each of the electrode sites.

To assess the internal consistency of the ERP measures with the current setup of participants and items, we calculated Cronbach's alpha for the averaged data. This coefficient represents the consistency of items (e.g., conditions) across observations (e.g., participants; see Thigpen, Kappenman, & Keil, 2017) and allows assessing whether the average across a number of trials and across a number of participants results in a consistent pattern, different from surrounding noise. Cronbach's alpha was calculated for individual ERP voltages at each electrode and time point (1–500 ms poststimuli) using the four conditions resulting from the combination of WF and response procedure as items, and 19 participants as observations. According to Hinton, McMurray, and Brownlow (2014), a Cronbach's alpha exceeding .9 indicates excellent internal consistency, whereas coefficients between .7 and .9 indicate high internal consistency and coefficients from .6 to .7 indicate moderate internal consistency. A coefficient below .5 is considered poor. In the current experiment, this analysis yielded the highest consistency estimates (>.9) for those epochs corresponding to the series of ERPs that characterize visual word recognition processes (N1, P2, N250, P300, N400). The estimates are displayed in Figure 1. Thus, the experimental conditions used as replications (items) displayed a very high degree of consistency in estimating the underlying dimension (here, the average volt-

[1] Following a reviewer's suggestion, we ran parallel ANOVAs on the data using the average reference—note that the reported results were obtained with linked mastoids-reference. Results were essentially the same.

[2] To increase the signal-to-noise ratio in two subjects with a large number of eyeblinks, we applied the independent component analysis procedure to correct eye blinks using the Infomax algorithm (Jung et al., 2000).

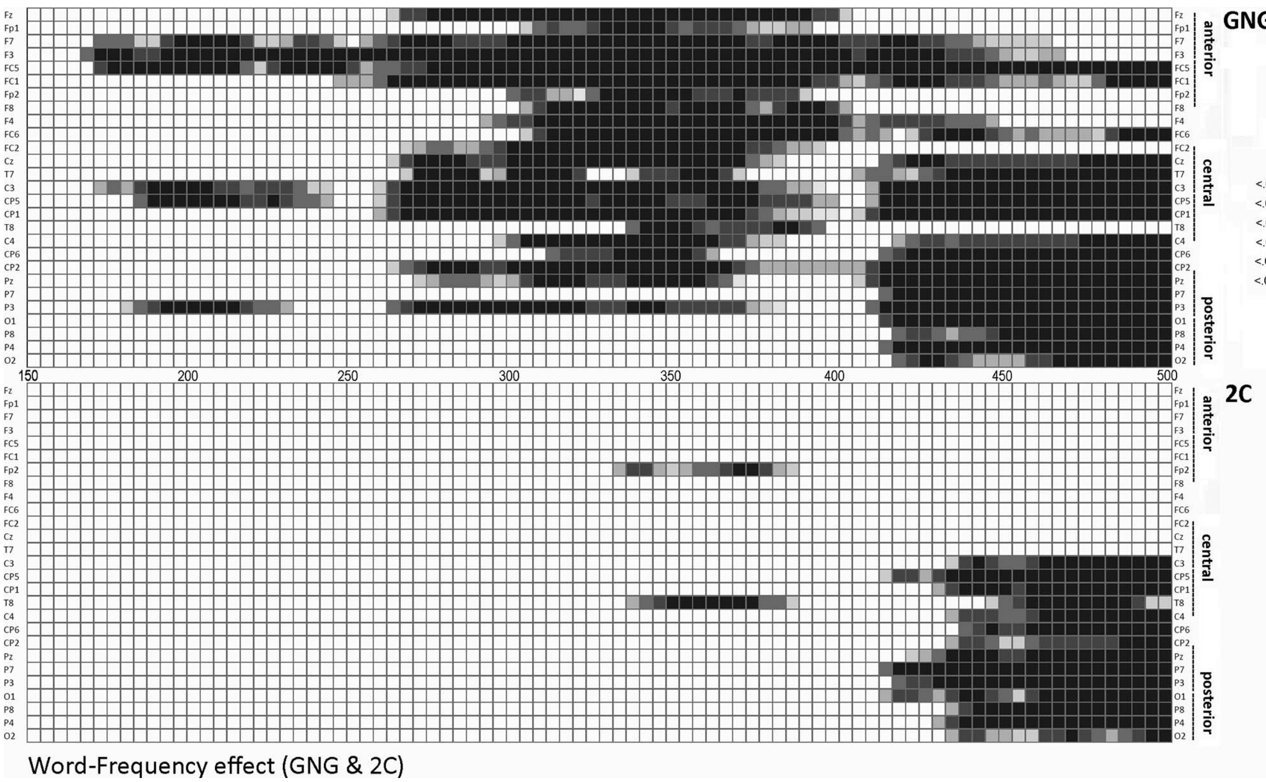

Figure 2. Results of the univariate statistical analyses of the time course of word-frequency in the two response procedures. The plots show the results of repeated-measures *t* tests at every 4-ms interval between 150 and 500 ms (no significant effects were obtained at earlier time-points) at all 27 scalp sites (listed in an anterior-posterior progression). To minimize false positive rates, the statistical significance level was set at $p < .05$ for a minimum of 15 contiguous data points (60 ms; Guthrie & Buchwald, 1991). *p* values are coded from lighter to darker.

age at a given electrode and time point) with the present number of items/condition and participants.

Since our main aim was to carefully track the time-course of the differences between experimental conditions, statistical analyses were performed on the mean ERP values in four time windows: 175–250 ms, 275–325 ms, 325–400 ms, and 400–500 ms. The selection of these epochs was informed by repeated measures *t* tests at every 4-ms interval between 1 and 500 ms at all 27 scalps sites for the factor WF (high vs. low) in the GNG and the 2C procedures separately. To minimize false positive rates, the statistical significance level was set at $p < .05$ for a minimum of 15 contiguous data points (60 ms; displayed in Figure 2; Guthrie & Buchwald, 1991; see also Vergara-Martínez et al., 2017, for a similar approach). We analyzed the topographical distribution of the ERP results by including the averaged amplitude values across three electrodes of nine representative scalp areas that result from the factorial combination of the factors laterality (left, central, right) and the anterior-posterior (AP) distribution (anterior, medial, posterior): left-anterior (FP1, F7, F3), left-medial (FC5, T7, C3), left-posterior (CP3, P7, P3), central-anterior (FZ, FC1, FC2), central-medial (CZ, CP1, CP2), central-posterior (PZ, O1, O2), right-anterior (FP2, F8, F4), right-medial (FC6, T8, C4); and right-posterior (CP6, P8, P4; Figure 3). This strategy was applied in each ERP analysis of the present experiment. For each time window, a separate repeated-measures analysis of variance (ANOVA) was performed, including the factors response procedure (GNG, 2C), frequency (high, low), laterality (left, central, right) and ap distribution (anterior, medial, posterior). In all analyses, list (List 1, List 2, List 3, List 4) was included as the between-subjects factor to extract the variance due to the counterbalanced lists (Pollatsek & Well, 1995). Effects for the topograph-

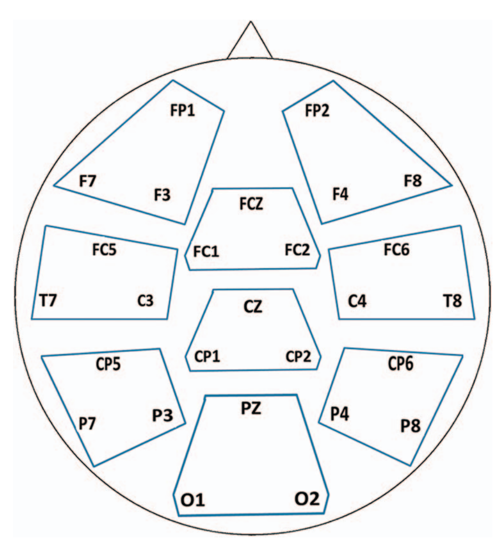

Figure 3. Schematic representation of the electrode montage. See the online article for the color version of this figure.

ical factors are reported when they interact with the experimental manipulations. Interactions between factors were followed up with comparisons using *F* tests.

In order to check for the differences regarding response preparation between the two tasks, we also performed an analysis on the readiness potential (or Bereitschaftspotential), a slow negative shift that begins 1–2 s prior to a voluntary action (Deecke, Scheid, & Kornhuber, 1969; for reviews, see Di Russo et al., 2017; Shibasaki & Hallett, 2006). Grand average response preparation ERPs were calculated in the −1,000–0 ms epoch preceding stimulus presentation and baselined from −1,000 to −800 ms. The ANOVA performed on the averaged voltage values across FP1 and FP2 in the −500–0 ms time window included the factors WF and response procedure.

Results

Behavioral Results

Error responses and lexical decision times less than 250 ms or greater than 1,500 ms were excluded from the RT analyses.[3] The mean lexical decision times and error rates per condition are displayed in Table 2. The participants' lexical decision times and percent errors per condition on word stimuli were submitted to separate ANOVAs with a 2 (response procedure: GNG, 2C) × 2 (WF: high-frequency, low-frequency) design—as stated above, List was also included as a dummy between-subjects factor. The ANOVA on the latency data showed that responses to high-frequency words were, on average, 50 ms faster than the responses to low-frequency words, $F(1, 15) = 94.3$, $p < .001$. Importantly, we found an interaction between WF and response procedure, $F(1, 15) = 9.62$, $p = .007$. This reflected that the effect of WF was greater in the GNG (60 ms) than in the 2C procedure (38 ms). In addition, there was an advantage of 33 ms in the GNG over the 2C procedure, although this difference did not reach statistical significance: $F(1, 15) = 2.1$, $p = .16$.

The ANOVA on the error data showed that participants committed fewer errors to high-frequency words than to low-frequency words (0.7% vs. 4.8%, respectively), $F(1, 15) = 24.7$, $p < .001$. Error rates for GNG and 2C were 2.1% versus 3.4%, respectively; $F(1, 15) = 3.76$, $p = .07$. There were no signs of an interaction between the two factors, $F < 1$.

In sum, as expected, the WF effect was greater in the GNG than in the 2C procedure (e.g., see Gomez et al., 2007; Hino & Lupker, 1998, 2000, for a similar pattern).

Table 2
Mean Lexical Decision Times (in ms) and Percentages of Error (in Parentheses) for Words and Nonwords in the Experiment

	Word frequency	
Task procedure	High frequency	Low frequency
Go/no-go task	567 (0.1)	628 (4.2)
Two-choice task	616 (1.4)	653 (5.5)

Note. The mean response time for nonwords in the two-choice lexical decision task was 724 ms. The percentage of errors for nonwords were 5.5% and 5.6% in the go/no-go and the two-choice tasks, respectively.

ERP Results

Figure 4 shows the ERP waves for the words (high- vs. low-frequency in the GNC vs. 2C procedures) in nine representative electrodes. The ERPs for the word stimuli produced an initial small negative potential peaking around 100 ms post-stimulus, which was followed by a much larger and slower positivity (P2) ranging between 150 and 250 ms (see anterior locations in Figure 4; note that the opposite pattern is shown in the most posterior locations). This positive deflection peaks earlier (P2: approx. at 250 ms) over most frontal-central areas compared to central-parietal scalp areas (P3: 300 ms) and carries a small negativity that is maximal around 200 ms poststimulus onset, and is more prominent over frontal and central electrode sites. Following these early potentials, a large and slow negativity extends approximately between 350 and 500 ms and is maximal around 400 ms post stimulus (i.e., the N400 component). In the two procedures, a substantial WF effect can be observed around 400 ms poststimuli, with larger negative values for low- than for high-frequency words (see Figures 4 and 5). Importantly, while the WF effect is already apparent in early time epochs in the GNG procedure (around 200 ms), the emergence of the WF effect occurs later in time in the 2C procedure. To capture the dissociation between the two response procedures regarding the time-course of the WF effect, we conducted ANOVAs in the following time epochs: 175–250 ms (N2), 275–325 ms (P3), 325–400 ms (N400a), and 400–500 ms (N400b). The results of the ANOVAs for each epoch are shown below.

175–250 ms epoch. The ANOVA showed that neither the main effect of WF nor the main effect of response procedure was significant (WF and response procedure: *F*s < 1). More importantly, there was a significant three-way interaction between WF, laterality, and response procedure, $(1, 15) = 6.30$, $p = .006$, $\eta_p^2 = .296$. Follow-up analyses showed a significant effect of WF in the GNG procedure, which was restricted to the left hemisphere, $F(1, 15) = 16.98$, $p = .001$: larger negative values were observed for low-frequency than for high-frequency words (WF effect in the central line: $F(1, 15) = 2.38$, $p = .14$; right hemisphere: $F < 1$). In contrast, there were no signs of a WF effect in the 2C procedure (all *F*s < 1).

250–325 ms epoch. The ANOVA showed a main effect of WF ($F(1, 15) = 7.94$, $p = .013$, $\eta_p^2 = .346$) that was modulated by an interaction between WF, laterality, and response procedure, $F(1, 15) = 5.51$, $p = .015$, $\eta_p^2 = .269$. Follow-up analyses showed that the effect of WF was significant in the GNG procedure, in both the left hemisphere: $F(1, 15) = 22.64$, $p < .001$; and in the medial line: $F(1, 15) = 19.74$, $p < .001$ (right hemisphere: $F(1, 15) = 4.37$, $p = .05$), with larger negative values for low-frequency than for high-frequency words. In contrast, there were no clear signs of a WF effect in the 2C procedure (left hemisphere and medial line: *F*s < 1; right hemisphere: $F(1, 15) = 1.9$, $p = .188$). The main effect of response procedure did not approach significance, $F(1, 15) = 1.19$, $p = .29$, $\eta_p^2 = .074$.

[3] The results were basically the same without such a cutoff: only one observation was greater than 1,500 ms (1,525 ms).

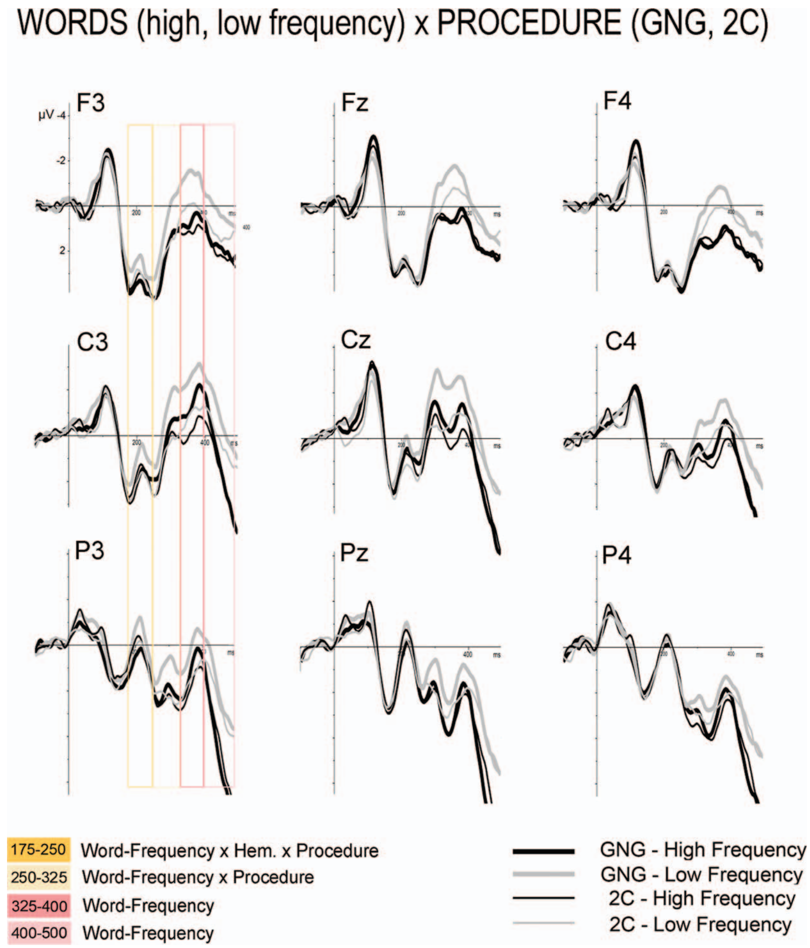

Figure 4. Grand average event-related potentials to high- and low-frequency words in the two task procedures (go/no-go [GNG] and two-choice [2C]) in nine representative electrodes from the nine areas of interest. Negativity is plotted upward, and each tick mark on the horizontal axis represents 100 ms. Significant word frequency effects (pink, light pink) or word frequency by task procedure interactions (orange, light orange) are represented for the four time-windows of interest (175–250 ms, 250–325 ms, 325–400 ms, and 400–500 ms). See the online article for the color version of this figure.

325–400 ms epoch. The ANOVA showed a main effect of WF ($F(1, 15) = 14.37$, $p = .002$, $\eta_p^2 = .489$) that was modulated by an interaction with AP distribution, $F(1, 15) = 6.84$, $p = .015$, $\eta_p^2 = .313$: anterior: $F(1, 15) = 20.15$, $p < .001$; central: $F(1, 15) = 13.78$, $p < .001$; posterior: $F(1, 15) = 3.04$, $p = .102$, reflecting larger negative values for low- than for high-frequency words across both procedures over frontal and central scalp areas. The ANOVA also showed a significant interaction between response procedure, laterality, and AP distribution, $F(1, 15) = 4.34$, $p = .017$, $\eta_p^2 = .225$. The effect of response procedure was significant over the following scalp areas: left-medial ($F(1, 15) = 11.04$, $p = .005$), left-posterior ($F(1, 15) = 8.77$, $p = .01$), and central-media, ($F(1, 15) = 4.97$, $p = .04$), where the GNG procedure elicited larger negativities than the 2C procedure. The factor Response procedure did not approach significance in the other areas of interest: left-anterior ($F(1, 15) = 2.55$, $p = .13$), central-anterior ($F(1, 15) = 2.14$, $p = .16$), central-posterior and right-anterior (both Fs < 1), right-medial ($F(1, 15) = 1.37$, $p = .26$), and right-posterior ($F < 1$).

400–500 ms epoch. The ANOVA showed a main effect of WF ($F(1, 15) = 17.17$, $p = .001$, $\eta_p^2 = .534$): larger negative values were elicited by low- than by high-frequency words. This effect was modulated by an interaction with laterality and AP distribution ($F(1, 15) = 6.60$, $p = .002$, $\eta_p^2 = .306$). The effect of WF was significant over the following scalp areas: left-medial ($F(1, 15) = 26.65$, $p < .001$), left-posterior ($F(1, 15) = 36.59$, $p < .001$), central-anterior ($F(1, 15) = 7.54$, $p = .015$), central-medial ($F(1, 15) = 17.09$, $p = .001$), central-posterior ($F(1, 15) = 18.08$, $p = .001$), right-medial ($F(1, 15) = 10.96$, $p = .005$), and right-posterior ($F(1, 15) = 19.51$, $p < .001$). The effect of WF did not approach significance in the other areas of interest: left-anterior ($F(1, 15) = 2.89$, $p = .11$) and right-anterior ($F(1, 15) = 1.68$, $p = .21$). That is, unlike the previous time epoch, the effect of WF was greater over posterior than anterior scalp areas (posterior: 1.67 mV; central: 1.47 mV; anterior: .78 mV). Finally, the main effect of response procedure did not approach significance ($F < 1$).

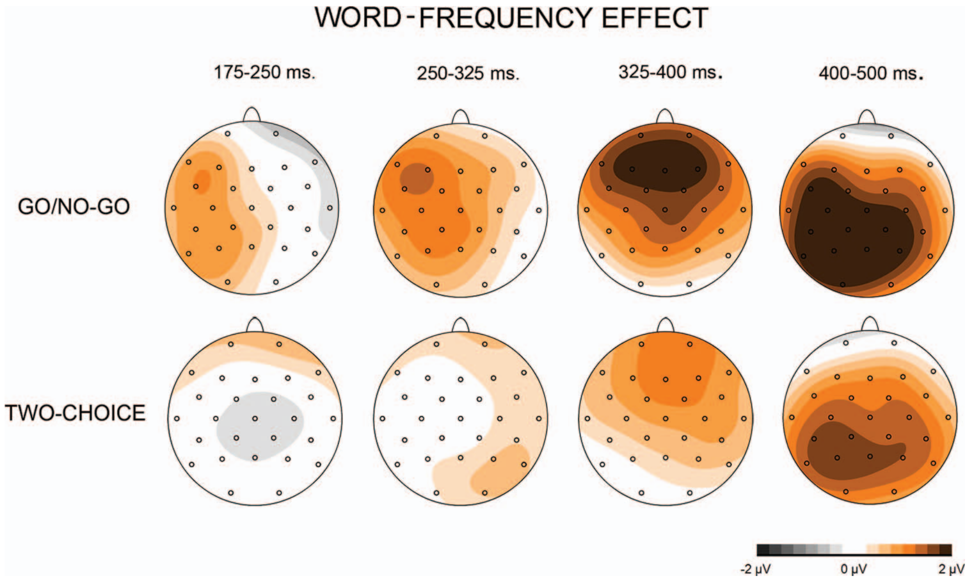

Figure 5. Topographic distribution of the word-frequency effect (calculated as the difference in voltage amplitude between the event-related potential responses to high- minus low-word-frequency words) in the four time windows of analysis plotted for the go/no-go and the two-choice procedures, respectively. See the online article for the color version of this figure.

Analysis on the Readiness Potential

The ANOVA performed on the averaged voltage values across FP1 and FP2 in the −500–0-ms time window included the factors WF and response procedure (Figure 6). The results revealed a main effect of procedure: $F(1, 15) = 11.58$, $p = .004$, $\eta_p^2 = .436$, with larger negative values for 2C than the GNG procedure.

In summary, the ERP data showed significant effects of WF in the GNG procedure starting as early as 200 ms (approximately) poststimuli. For the 2C procedure, the WF effect emerged later, at about 300 ms. To quantitatively assess the robustness of this critical ERP finding, we computed Cronbach's alpha on the interaction between WF and response procedure over left scalp areas in the 175–250-ms time epoch (Thigpen et al., 2017). For this analysis, the between-condition averaged ERP effects (high- minus low-frequency words across each response procedure) across each level of laterality (left, central and right) served as "items." Whereas high Cronbach's alpha values would reflect high consistency of items (i.e., the effects are similar across scalp areas), a decrease in Cronbach's alpha would reveal that the WF effects are different across scalp areas. Furthermore, as this interaction was confined to the 175–250-ms epoch, we used other time epochs of the waveform (0–100 ms, 100–175 ms, 250–325 ms, 325–400 ms, and 400–500 ms) to address the consistency of the overall waveforms. Results showed that the lowest Cronbach's alpha occurred in the 175–250-ms epoch (.73), while the other time epochs showed higher Cronbach's alpha consistency values (0–100 ms: $\alpha = .90$; 100–175 ms: $\alpha = .87$; 250–325 ms: $\alpha = .86$; 325–400 ms: $\alpha = .85$; 400–500 ms: $\alpha = .90$). Thus, the relatively lower consistency between WF effects (across different scalp areas) in the time epoch of interest (175–250 ms), compared to the

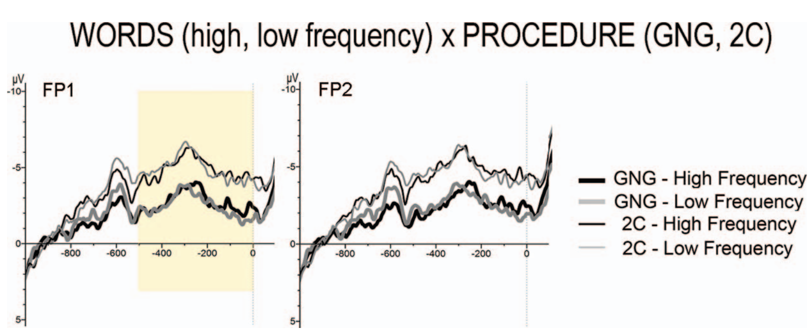

Figure 6. Grand average response preparation event-related potentials calculated in the −1,000–0-ms epoch preceding stimulus presentation and baselined from −1,000 to −800 ms. Significant effects of response procedure were obtained in the −500–0-ms time window. See the online article for the color version of this figure.

high consistency obtained in the other time-windows, adds robustness to the present pattern of data and converges with the analysis of the experimental variables in the ANOVAs. Note that in line with the consistency analysis reported in the EEG Recording and Analysis section, the robustness analysis of ERP effects by means of Cronbach's alpha computation, allows to characterize the reliability of the present ERP findings.

The implications of this dissociation between GNG and 2C procedures regarding the WF effect in early ERP components will be examined in the next section.

Discussion

Understanding the underpinnings of the procedures that we employ in laboratory tasks is a core issue in cognitive neuroscience with important theoretical implications. Here we examined the extent to which response procedure (GNG vs. 2C) interacts with central components of processing (i.e., lexical processes). Prior behavioral (latency-based) experiments have shown that response procedure may affect core processes in lexical decision and other tasks. For instance, Hino and Lupker (1998, 2000) found a greater WF effect in the GNG than in the 2C version of the lexical-decision task (see Gomez et al., 2007, for a similar pattern). Here, we have replicated this behavioral finding. More importantly, the electrophysiological signature of WF was modulated by response procedure as early as 200 ms poststimuli. In the GNG procedure, the WF effect emerged in early time epochs (around 200 ms). In contrast, in the 2C procedure, the onset of the WF effect started around 300 ms. Furthermore, at this time window the magnitude of the WF effect was slightly larger in the GNG than in the 2C procedure (see Figure 3). Taken together, this pattern reveals that: (a) response procedure (GNG vs. 2C) affects early processes in lexical decision, and (b) the GNG lexical decision may be more sensitive to lexical effects than the 2C lexical decision.

At the earliest time epochs of analysis (175–250 ms and 250–325 ms), we found a robust left-lateralized WF effect (with larger negative values for low- than for high-frequency words) only in the GNG procedure. As stated in the introduction, WF effects have been reported throughout this time window and scalp distribution in previous studies (Hauk et al., 2006; Hauk & Pulvermüller, 2004). In the present experiment, the early negativity (175–250 ms) matches well with the N2 component. This ERP component is mainly obtained in the context of the GNG paradigm and is maximal at around 200 ms poststimuli over most frontal and central scalp areas. Larger N2 amplitudes are obtained under conditions where there is a tendency to make an a priori likely but incorrect response (e.g., low rate stimuli; Nieuwenhuis, Yeung, & Cohen, 2004; see Folstein & Van Petten, 2008, for a review). Furthermore, the N2 amplitude varies with the degree of perceptual overlap between the go and no-go stimuli: there is a larger N2 amplitude elicited by stimuli that are hard to discriminate in no-go trails relative to easy-to-discriminate stimuli (Jodo & Kayama, 1992; Maguire et al., 2009; Nieuwenhuis et al., 2004). While this component has been related to a cognitive top-down inhibition mechanism needed to suppress the incorrect tendency to respond (Eimer, 1993; van Boxtel, van der Molen, Jennings, & Brunia, 2001), more recent evidence suggests that it reflects conflict-monitoring (i.e., the competition between two alternative response representations; Donkers & van Boxtel, 2004; Enriquez-Geppert, Konrad, Pantev, & Huster, 2010; Nieuwenhuis et al., 2003).

The larger N2 amplitudes for the low- than the high-frequency words in the GNG procedure may be accounted for in terms of larger response conflict for low-frequency words. In the GNG procedure, and for each stimulus, two alternative response representations (trigger vs. withhold a response) compete until enough evidence is mapped onto its corresponding lexical representation. This competition process may be more apparent at early stages when processing relatively unfamiliar words, as the amount of information on lexicality would be still ambiguous (i.e., the evidence has not reached the threshold for a response)—note that in all models of word recognition this process is completed earlier for high-frequency than for low-frequency words. Consistent with this interpretation, this conflict monitoring (response conflict) would be larger for more ambiguous stimuli such as low-frequency words, which would lead to larger N2 amplitudes than for high-frequency words in the GNG procedure.

In order to understand the dissociation between procedures regarding the N2 ERP frequency effect, one might argue that the participants' decision on whether to trigger a response follows the same path in the GNG and 2C procedures: in both scenarios, it is the result of the interaction between accumulation of evidence, the ability to suppress an incorrect tendency to respond (response conflict), and response preparation. Therefore, we might have expected larger response conflict for low versus high frequency words not only in the GNG but in the 2C procedure as well. In fact, models on executive control of conflict assume that the nature of conflict between response representations is similar in both 2C and GNG procedures (Jones et al., 2002; Nieuwenhuis et al., 2003).

However, we have to consider the subtle differences between the procedures in relation to response preparation and the impact this may have onto stimulus processing. As outlined in the introduction, premature response preparation does not run free of risk in the 2C procedure, especially in low-frequency words, as participants might be wrongly responding with a no response. In the GNG procedure instead, the incorrect activation of a no-go representation for a yes trial (e.g., in the case of a low-frequency word) does not necessarily fall into a manifest error: the correct response can still be attained if the deadline (end of the trial) has not been reached yet. Hence, the 2C requirement of a larger control over response preparation, as compared to the GNG, may have imposed larger processing demands, leading to a readjustment of attentional resources onto response preparation at the expense of reducing resources on stimulus processing. In fact, Smid, Fiedler, and Heinze (2000) argued that "the difficulty of response selection/preparation in the 2C may have drawn resources allocated to feature integration to be reallocated to response selection/preparation" (p. 1069). This might explain why during visual word processing in the two task procedures, an early ERP component (namely, N2) was sensitive to WF in the GNG but not in the 2C procedure. In fact, the largest values in the N2 were obtained for the low-frequency words in the GNG condition, while the N2 response for the high-frequency words was similar to the N2 response for both word types in the 2C procedure. This finding reveals that, by this time window (200 ms) in the 2C procedure, the processing of words did not seem to proceed as thoroughly as in the GNG procedure.

Notably, in the present experiment, the subtle particularities regarding response preparation seemed to affect very early stages of lexical-semantic information processing while later stages of visual word recognition remained intact. The analysis on the readiness potential allowed us to check for the differences regarding response preparation between the two tasks. As shown in Figure 6, statistically significant larger negative amplitudes were obtained for the time period preceding the stimulus in the 2C compared to the GNG procedure. Increases in RP amplitude reflect differences in psychological factors including attention, motivation or response strategy (Shibasaki & Hallett, 2006). More specifically, the larger negative amplitude of the RP would reflect increased attentional demands and preprogrammed control (Masaki, Takasawa, & Yamazaki, 1998). Although the present scenario was not designed to disentangle the unique source of this RP effect, it adds support to the interpretation of the dissociation of WF effects across the two response procedures in terms of differences in response preparation.

In the following epochs of analysis (325–400 ms, and 400–500 ms), low-frequency words, regardless of response procedure, elicited larger negativities than high-frequency words, with the largest difference peaking at around 400 ms (see Figures 4 and 5)—note that the WF effect was numerically larger in the GNG than in the 2C procedure.[4] This ERP deflection matches the classic N400 component, a psychophysiological reflection of lexical-semantic access (Kutas & Federmeier, 2011; Vergara-Martínez & Swaab, 2012). Thus, by this time epoch, the electrophysiological markers of word processing largely converge in the two procedures. In other words, the processing cost when retrieving lexical-semantic word properties from long-term memory (N400), as indicated by the WF effect, proceeds independently of the subtle differences in the GNG and 2C procedures. An effect of procedure was also observed in the early N400 with larger amplitudes for words in the GNG than in the 2C procedure. According to our present theoretical framework and in studies, taxing attentional resources (as occurs in the 2C compared to the GNG procedures) results in a decrease of N400 amplitudes (Lien, Ruthruff, Cornett, Goodin, & Allen, 2008; see Van Petten, 2014, for a review).

Taken together, the present findings qualify previous claims by Gomez et al. (2007) within the context of evidence accumulator models, who argued that the main difference between the two procedures was mostly in the response execution phase. Gomez et al. (2007) used the DDM to examine whether 2C and GNG procedures merely vary in ancillary processes such as the time of encoding and response (T_{er} parameter in the model) or whether they differed in a core decision parameter such as the quality of information (drift-rate parameter). Fits of the DDM to the data revealed that the similarities between the two response methods were more important than the differences (see also Ratcliff, Huang-Pollock, & McKoon, 2018, for a similar conclusion). Leaving aside that the GNG requires an implicit no-go boundary (i.e., the model without a no-go boundary produced poor fits), the differences across procedures did not occur in the rate of evidence accumulation, which led them to conclude that the core components were unaffected. Gomez et al. (2007) did find a difference in the T_{er} parameter (shorter T_{er} for the GNG procedure), which they interpreted as a difference in the response execution state. Note, however, that the T_{er} parameter in the diffusion model cannot distinguish between encoding and response processes, and that the use of behavioral data (RTs and error rates) only provides one final outcome for the processes that underlie a word/nonword response in lexical decision. After that article was published, further research has indicated that lexical processes do affect the encoding part of the T_{er} parameter (e.g., see Gomez & Perea, 2014; Gomez, Perea, & Ratcliff, 2013). Thanks to the exquisite temporal resolution of the ERP technique, we were able to track the electrophysiological counterparts of the similarities/differences across procedures. Altogether, the early latency of the differences between procedures point to differences in early processes such as the encoding processes—this would be consistent with the presence of a difference in T_{er} (time of encoding and response) between GNG and 2C lexical decision (Gomez et al., 2007), but is inconsistent with their original interpretation.

Our findings are a new demonstration of the impact of different attentional demands on early stages of word processing. Previous research had addressed this issue by changing the stimulus criteria on which participants' judgments were required (Chen et al., 2015; Norris, Kinoshita, Hall, & Henson, 2018; Strijkers et al., 2015). Those studies tested the degree to which different aspects of stimulus processing (e.g., WF) remained impervious to top-down modulation of attention on specific features (from more superficial to semantic ones). For example, Strijkers et al. (2015) found very early WF ERP effects (120 ms) in a semantic GNG categorization task than in a color GNG categorization task, indicating that top-down modulation already affects early information retrieval in visual word recognition. Chen et al. (2015) also obtained early WF effects (starting at around 150 ms) in the occipitotemporal cortex in the lexical-decision task, compared to two different psycholinguistic tasks (semantic categorization and silent reading) that differed in the specific response selection demands.

The above results revealed that visual word recognition does not unfold in a straightforward manner, but instead it is sensitive to different methodological factors from very early during stimulus processing. These factors shall include not only the final goals of the recognition process (as assessed by Chen et al., 2015, or Strijkers et al., 2015), but others related to specific response displays. As shown in the present experiment, even the subtle change in the response pattern (GNG vs. 2C) may affect early processing stages in visual word recognition. In binary classification tasks, participants are explicitly directed to pay attention to specific aspects of the stimuli, and developing expectances to perform the task successfully is normal. The lexical-decision task (i.e., a widely employed binary task) directs attention to the "wordness" of stimuli, and it is likely that the processing of relevant information (from low-level perceptual to more abstract word features) is amplified, as in Chen et al. (2015). The earlier WF ERP effects observed in the GNG compared to the 2C procedure suggest that wordness information may be available earlier (and to a larger degree) in the GNG procedure, contrary to what is observed in the 2C. This might result from the differences between proce-

[4] Although the WF effect was numerically larger in the GNG than in the 2C procedure in the N400 time windows (see Figures 4 and 5), the interaction was not statistically significant: Frequency × Procedure: N400a: $F(1,15) = 2.7$, $p = .116$, $\eta_p^2 = .156$; N400b: $F(1,15) = 1.4$, $p = .248$, $\eta_p^2 = .088$. We acknowledge that a more powerful design could have detected this subtle interaction—note however that the critical dissociation regarding WF and procedure was captured in the early time windows.

dures regarding response preparation, which limits the available resources to a larger extent in the 2C than in the GNG procedure. In a nutshell, while a word/nonword discrimination is the task to be made in both GNG and 2C lexical decision, the strength to which the system allocates attention to wordness properties seems to be counteracted by cognitive control on response preparation. The neuronal mechanisms that mediate this modulation will need further detailed analyses of brain dynamics, which stablishes new challenges for future investigation.

We acknowledge that although we have considered the WF effect as an upper bound of lexical access, this premise is not free from criticism. Due to the pervasive intercorrelation of lexical variables (Laszlo & Federmeier, 2014), it is difficult to disentangle the extent to which WF effects may reflect access to orthographic, sublexical, or semantic features during lexical retrieval. Hence, we cannot assert whether the early GNG effects might also be sensitive to other factors (e.g., low-level orthographic features, sublexical units or lexical-semantic representations)—this would be beyond the scope of the current experiment.[5] Instead, the central outcome of the present study was that, in spite of using the exact same set of high- and low-frequency words, and despite the fact that N400 effects were stable across both response procedures, we found an earlier marker of lexical frequency in the GNG procedure than in the 2C procedure.

From the present results, one may conclude that due to the dynamics of response preparation and response deadline, the 2C is therefore more resource demanding than the GNG. Hence one might wonder whether changing any aspect of the task that increases attentional resources might impact the nature of word processing in the same way as we have found here. There are many potential avenues for future research on the interplay between attentional resources and lexical processing, as measured by the N2 component. As suggested by two reviewers, one might design experiments manipulating elements such as (a) the presence/absence of a response deadline, (b) the different type of nonwords, (c) the same/different mapping of the responses and hands (e.g., 2 fingers of the same hand vs. 1 finger of each hand), or (d) the proportion of words/nonwords. First, having a response deadline may recruit all available sources in order to perform the task accurately including the fast processing of very early perceptual information. As "taxing" the attentional system by including a response deadline may be counteracted by performance optimization, we would expect (even) early effects of WF (in line with Hauk et al., 2006, who employ fast rate stimulus presentation: 100ms). Second, regarding the effect of the type of nonwords, Ziegler, Besson, Jacobs, Nazir, and Carr (1997), found a large N2 for pseudowords compared to both words and nonwords in a GNG semantic categorization task. This result was interpreted as an early successful categorization of words and nonwords compared to difficulties (response conflict) encountered with pseudowords: pseudowords were not as easily categorized as the two other stimulus types (words and nonwords) due to being orthographically and phonologically similar to words (Vergara-Martínez, Perea, Gómez, & Swaab, 2013). Third, given that responses are slower when yes and no responses involve the same hand than when they involve different hands (Gilmour Reeve & Proctor, 1988), one might hypothesize—on the basis of an attentional explanation of the phenomenon—that the N2 WF would be attenuated in a same-hand block compared to a different-hand block. Fourth, the manipulation of the proportion of nonwords across blocks in an ERP experiment measuring the N2 component may be used to shed some light on the flexibility of the word recognition process and how this flexibility dynamically changes along the experiment itself (see Wagenmakers, Ratcliff, Gomez, & McKoon, 2008, for behavioral evidence). The idea is that participants may need to rearrange their mental setting regarding response preparation and response deadline according to the difficulty of the experimental context.

In sum, our results reflect an online performance adjustment in the way words are processed following a subtle change in the response procedure (GNG vs. 2C). This manipulation had an impact on the nature of stimulus processing during word recognition as revealed by the earlier effect of WF in the GNG versus the 2C procedure. This pattern strongly suggests that there is a substantial attentional component in the word recognition process—including the WF effect (e.g., see also Lachter, Forster, & Ruthruff, 2004, for a null effect of masked repetition priming when the primes were displayed in an unattended location). Future implementations of models of visual word recognition should incorporate attentional mechanisms that are sensitive to the specific interrelation between accumulation of evidence, response conflict, and response preparation.

[5] We also considered the potential role of age of acquisition (AoA) in the current experiment. Unfortunately, unlike English, there are no available databases for objective AoA in Spanish, and 22% of the words did not appear in the largest subjective AoA database in Spanish (Alonso, Fernandez, & Díez, 2015). For the existing values, WF and AoA were only moderately correlated (−.4). Importantly, the two factors appear to play a different role during visual word recognition. In Spanish, Cuetos, Barbón, Urrutia, and Domínguez (2009) found early WF effects (175–250 ms) but relatively late AoA effects (400–610 ms). This late effect was replicated by Räling, Holzgrefe-Lang, Schröder, and Wartenburger (2015). Furthermore, the N400 AoA effect in the Räling et al. (2015)experiment was opposite to the WF N400 effect, with larger N400 amplitudes for the early than the late acquired words. As the electrophysiological marker of the AoA in the above-cited experiments differs from our findings, it is unlikely that AoA influenced the WF effects in the current experiment.

References

Alonso, M. A., Fernandez, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods, 47,* 268–274. http://dx.doi.org/10.3758/s13428-014-0454-2

Assadollahi, R., & Pulvermüller, F. (2003). Early influences of word length and frequency: A group study using MEG. *NeuroReport: For Rapid Communication of Neuroscience Research, 14,* 1183–1187. http://dx.doi.org/10.1097/00001756-200306110-00016

Bacon-Macé, N., Kirchner, H., Fabre-Thorpe, M., & Thorpe, S. J. (2007). Effects of task requirements on rapid natural scene processing: From common sensory encoding to distinct decisional mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1013–1026. http://dx.doi.org/10.1037/0096-1523.33.5.1013

Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *NeuroReport: For Rapid Communication of Neuroscience Research, 15,* 545–548. http://dx.doi.org/10.1097/00001756-200403010-00032

Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., & Jacobs, A. M. (2009). Pseudohomophone effects provide evidence of early lexico-phonological processing in visual word recognition. *Human Brain Mapping, 30,* 1977–1989. http://dx.doi.org/10.1002/hbm.20643

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex, 11,* 825–836. http://dx.doi.org/10.1093/cercor/11.9.825

Broadbent, D. E., & Gregory, M. (1962). Donders' B- and C-reactions and S-R compatibility. *Journal of Experimental Psychology, 63,* 575–578. http://dx.doi.org/10.1037/h0044674

Callan, J., Klisz, D., & Parsons, O. A. (1974). Strength of auditory stimulus-response compatibility as a function of task complexity. *Journal of Experimental Psychology, 102,* 1039–1045. http://dx.doi.org/10.1037/h0036329

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what, when, where, and how of visual word recognition. *Trends in Cognitive Sciences, 18,* 90–98. http://dx.doi.org/10.1016/j.tics.2013.11.005

Carreiras, M., Vergara, M., & Barber, H. (2005). Early event-related potential effects of syllabic processing during visual word recognition. *Journal of Cognitive Neuroscience, 17,* 1803–1817. http://dx.doi.org/10.1162/089892905774589217

Chen, Y., Davis, M. H., Pulvermüller, F., & Hauk, O. (2015). Early visual word processing is flexible: Evidence from spatiotemporal brain dynamics. *Journal of Cognitive Neuroscience, 27,* 1738–1751. http://dx.doi.org/10.1162/jocn_a_00815

Cuetos, F., Barbón, A., Urrutia, M., & Domínguez, A. (2009). Determining the time course of lexical frequency and age of acquisition using ERP. *Clinical Neurophysiology, 120,* 285–294. http://dx.doi.org/10.1016/j.clinph.2008.11.003

Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia, 50,* 1852–1870. http://dx.doi.org/10.1016/j.neuropsychologia.2012.04.011

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research, 1155,* 147–162. http://dx.doi.org/10.1016/j.brainres.2007.04.027

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods, 37,* 665–671. http://dx.doi.org/10.3758/BF03192738

Deecke, L., Scheid, P., & Kornhuber, H. H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research, 7,* 158–168. http://dx.doi.org/10.1007/BF00235441

Di Russo, F., Berchicci, M., Bozzacchi, C., Perri, R. L., Pitzalis, S., & Spinelli, D. (2017). Beyond the "Bereitschaftspotential": Action preparation behind cognitive functions. *Neuroscience and Biobehavioral Reviews, 78,* 57–81. http://dx.doi.org/10.1016/j.neubiorev.2017.04.019

Donders, F. C. (1969). Over de snelheid van psychische processen [On the speed of psychological processes]. *Acta Psychologica, 30,* 412–431. (Original work published 1868)

Donkers, F. C., & van Boxtel, G. J. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56,* 165–176. http://dx.doi.org/10.1016/j.bandc.2004.04.005

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a go/nogo task. *Biological Psychology, 35,* 123–138. http://dx.doi.org/10.1016/0301-0511(93)90009-W

Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage, 51,* 877–887. http://dx.doi.org/10.1016/j.neuroimage.2010.02.043

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

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior, 12,* 627–635. http://dx.doi.org/10.1016/S0022-5371(73)80042-8

Gilmour Reeve, T., & Proctor, R. W. (1988). Determinants of two-choice reaction-time patterns for same-hand and different-hand finger pairings. *Journal of Motor Behavior, 20,* 317–340. http://dx.doi.org/10.1080/00222895.1988.10735448

Gomez, P., & Perea, M. (2014). Decomposing encoding and decisional components in visual-word recognition: A diffusion model analysis. *The Quarterly Journal of Experimental Psychology, 67,* 2455–2466. http://dx.doi.org/10.1080/17470218.2014.937447

Gomez, P., Perea, M., & Ratcliff, R. (2013). A diffusion model account of masked versus unmasked priming: Are they qualitatively different? *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1731–1740. http://dx.doi.org/10.1037/a0032333

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136,* 389–413. http://dx.doi.org/10.1037/0096-3445.136.3.389

Gordon, B. (1983). Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning & Verbal Behavior, 22,* 24–44. http://dx.doi.org/10.1016/S0022-5371(83)80004-8

Gottsdanker, R., & Shragg, G. P. (1985). Verification of Donders' subtraction method. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 765–776. http://dx.doi.org/10.1037/0096-1523.11.6.765

Grice, G. R., & Reed, J. M. (1992). What makes targets redundant? *Perception & Psychophysics, 51,* 437–442. http://dx.doi.org/10.3758/BF03211639

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28,* 240–244. http://dx.doi.org/10.1111/j.1469-8986.1991.tb00417.x

Hackley, S. A., Schäffer, R., & Miller, J. (1990). Preparation for Donders' type B and C reaction tasks. *Acta Psychologica, 74,* 15–33. http://dx.doi.org/10.1016/0001-6918(90)90032-B

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage, 30,* 1383–1400. http://dx.doi.org/10.1016/j.neuroimage.2005.11.048

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology, 115,* 1090–1103. http://dx.doi.org/10.1016/j.clinph.2003.12.020

Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology, 80,* 64–74. http://dx.doi.org/10.1016/j.biopsycho.2008.04.015

Hino, Y., & Lupker, S. J. (1998). The effects of word frequency for Japanese Kana and Kanji words in naming and lexical decision: Can the dual-route model save the lexical-selection account? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1431–1453. http://dx.doi.org/10.1037/0096-1523.24.5.1431

Hino, Y., & Lupker, S. J. (2000). The effects of word frequency and spelling-to- sound regularity in naming with and without lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 166–183. http://dx.doi.org/10.1037/0096-1523.26.1.166

Hinton, P., McMurray, I., & Brownlow, C. (2014). *SPSS explained*. London, UK: Routledge. http://dx.doi.org/10.4324/9781315797298

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a go/no-go task. *Electroencephalography & Clinical Neurophysiology, 82,* 477–482. http://dx.doi.org/10.1016/0013-4694(92)90054-L

Jones, A. D., Cho, R. Y., Nystrom, L. E., Cohen, J. D., & Braver, T. S. (2002). A computational model of anterior cingulate function in speeded response tasks: Effects of frequency, sequence, and conflict. *Cognitive,* *Affective & Behavioral Neuroscience, 2,* 300–317. http://dx.doi.org/10.3758/CABN.2.4.300

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37,* 163–178. http://dx.doi.org/10.1111/1469-8986.3720163

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42,* 627–633. http://dx.doi.org/10.3758/BRM.42.3.627

King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters, 244,* 61–64. http://dx.doi.org/10.1016/S0304-3940(98)00140-2

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647. http://dx.doi.org/10.1146/annurev.psych.093008.131123

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111,* 880–913. http://dx.doi.org/10.1037/0033-295X.111.4.880

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single item ERPs. *Language, Cognition and Neuroscience, 29,* 642–661. http://dx.doi.org/10.1080/01690965.2013.866259

Lien, M.-C., Ruthruff, E., Cornett, L., Goodin, Z., & Allen, P. A. (2008). On the nonautomaticity of visual word processing: Electrophysiological evidence that word processing requires central attention. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 751–773. http://dx.doi.org/10.1037/0096-1523.34.3.751

Maguire, M. J., Brier, M. R., Moore, P. S., Ferree, T. C., Ray, D., Mostofsky, S., . . . Kraut, M. A. (2009). The influence of perceptual and semantic categorization on inhibitory processing as measured by the N2–P3 response. *Brain and Cognition, 71,* 196–203. http://dx.doi.org/10.1016/j.bandc.2009.08.018

Masaki, H., Takasawa, N., & Yamazaki, K. (1998). Enhanced negative slope of the readiness potential preceding a target force production task. *Electroencephalography and Clinical Neurophysiology, 108,* 390–397. http://dx.doi.org/10.1016/S0168-5597(98)00019-7

Münte, T. F., Wieringa, B. M., Weyerts, H., Szentkuti, A., Matzke, M., & Johannes, S. (2001). Differences in brain potentials to open and closed class words: Class and frequency effects. *Neuropsychologia, 39,* 91–102. http://dx.doi.org/10.1016/S0028-3932(00)00095-6

Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology, 41,* 157–160. http://dx.doi.org/10.1046/j.1469-8986.2003.00128.x

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective & Behavioral Neuroscience, 3,* 17–26. http://dx.doi.org/10.3758/CABN.3.1.17

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General, 137,* 434–455. http://dx.doi.org/10.1037/a0012799

Norris, D., Kinoshita, S., Hall, J., & Henson, R. (2018). Is reading automatic? Are the ERP correlates of masked priming really lexical? *Language, Cognition and Neuroscience, 33,* 1152–1167. http://dx.doi.org/10.1080/23273798.2018.1493516

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

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology, 46,* 143–168. http://dx.doi.org/10.1016/S0301-0511(97)05250-2

Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology, 74,* 374–388. http://dx.doi.org/10.1016/j.biopsycho.2006.09.008

Perea, M., Mallouh, R. A., & Carreiras, M. (2014). Are root letters compulsory for lexical access in Semitic languages? The case of masked form-priming in Arabic. *Cognition, 132,* 491–500. http://dx.doi.org/10.1016/j.cognition.2014.05.008

Perea, M., & Rosa, E. (2003). Los efectos de facilitación semántica con las tareas de decisión léxica sí-no y sólo-sí [The effects of semantic priming with the yes/no and the go/no-go lexical decision task]. *Psicothema, 15,* 114–119.

Perea, M., Rosa, E., & Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition, 30,* 34–45. http://dx.doi.org/10.3758/BF03195263

Perea, M., Soares, A. P., & Comesaña, M. (2013). Contextual diversity is a main determinant of word identification times in young readers. *Journal of Experimental Child Psychology, 116,* 37–44. http://dx.doi.org/10.1016/j.jecp.2012.10.014

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 785–794. http://dx.doi.org/10.1037/0278-7393.21.3.785

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *Journal of General Psychology, 13,* 199–203. http://dx.doi.org/10.1080/00221309.1935.9917878

Proverbio, A. M., Vecchi, L., & Zani, A. (2004). From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *Journal of Cognitive Neuroscience, 16,* 301–317. http://dx.doi.org/10.1162/089892904322984580

Räling, R., Holzgrefe-Lang, J., Schröder, A., & Wartenburger, I. (2015). On the influence of typicality and age of acquisition on semantic processing: Diverging evidence from behavioural and ERP responses. *Neuropsychologia, 75,* 186–200. http://dx.doi.org/10.1016/j.neuropsychologia.2015.05.031

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182. http://dx.doi.org/10.1037/0033-295X.111.1.159

Ratcliff, R., Huang-Pollock, C., & McKoon, G. (2018). Modeling individual differences in the go/no-go task with a diffusion model. *Decision, 5,* 42–62. http://dx.doi.org/10.1037/dec0000065

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning & Verbal Behavior, 9,* 487–494. http://dx.doi.org/10.1016/S0022-5371(70)80091-3

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition, 18,* 367–379. http://dx.doi.org/10.3758/BF03197126

Rugg, M. D., & Doyle, M. C. (1992). Event-related potentials and recognition memory for low- and high-frequency words. *Journal of Cognitive Neuroscience, 4,* 69–79. http://dx.doi.org/10.1162/jocn.1992.4.1.69

Sebastián-Gallés, N., Martí, M. A., Cuetos, F., & Carreiras, M. (2000). *LEXESP: Léxico informatizado del español* [LEXESP: Computerized lexicon of Spanish]. Barcelona, Spain: Edicions de la Universitat.

Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition. *NeuroReport: For Rapid Communication of Neuroscience Research, 9,* 2195–2200. http://dx.doi.org/10.1097/00001756-199807130-00009

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology, 117,* 2341–2356. http://dx.doi.org/10.1016/j.clinph.2006.04.025

Siakaluk, P. D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition, 31,* 100–113. http://dx.doi.org/10.3758/BF03196086

Smid, H. G., Fiedler, R., & Heinze, H. J. (2000). An electrophysiological study of the insertion of overt response choice. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1053–1071. http://dx.doi.org/10.1037/0096-1523.26.3.1053

Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43,* 195–201. http://dx.doi.org/10.1037/h0054636

Strijkers, K., Bertrand, D., & Grainger, J. (2015). Seeing the same words differently: The time course of automaticity and top-down intention in reading. *Journal of Cognitive Neuroscience, 27,* 1542–1551. http://dx.doi.org/10.1162/jocn_a_00797

Thigpen, N. N., Kappenman, E. S., & Keil, A. (2017). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology, 54,* 123–138. http://dx.doi.org/10.1111/psyp.12629

van Boxtel, G. J., van der Molen, M. W., Jennings, J. R., & Brunia, C. H. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology, 58,* 229–262. http://dx.doi.org/10.1016/S0301-0511(01)00117-X

Van Petten, C. (2014). Selective attention, processing load, and semantics. In G. R. R. Mangun (Ed.), *Cognitive electrophysiology of attention* (pp. 236–253). New York, NY: Elsevier. http://dx.doi.org/10.1016/b978-0-12-398451-7.00019-1

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18,* 380–393. http://dx.doi.org/10.3758/BF03197127

Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective & Behavioral Neuroscience, 17,* 461–474. http://dx.doi.org/10.3758/s13415-016-0491-7

Vergara-Martínez, M., & Gomez, P. (2020). *Should I stay or should I go/nogo*. Retrieved from osf.io/gpc46

Vergara-Martínez, M., Perea, M., Gómez, P., & Swaab, T. Y. (2013). ERP correlates of letter identity and letter position are modulated by lexical frequency. *Brain and Language, 125,* 11–27. http://dx.doi.org/10.1016/j.bandl.2012.12.009

Vergara-Martínez, M., & Swaab, T. Y. (2012). Orthographic neighborhood effects as a function of word frequency: An event-related potential study. *Psychophysiology, 49,* 1277–1289. http://dx.doi.org/10.1111/j.1469-8986.2012.01410.x

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. http://dx.doi.org/10.1016/j.jml.2007.04.006

Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience, 9,* 758–775. http://dx.doi.org/10.1162/jocn.1997.9.6.758

Appendix

List of Words and Nonwords in the Experiment

High-frequency words: abajo; ajeno; árbol; beber; bomba; breve; cifra; color; débil; droga; dulce; etapa; fácil; feliz; frase; fuego; grave; héroe; hotel; ideal; juego; lejos; líder; local; miedo; nariz; nieve; pedir; pieza; plazo; poeta; radio; reina; rumbo; señal; suave; tesis; unión; venir; virus; actor; ánimo; avión; bolsa; brazo; carne; civil; clase; culpa; dosis; dueño; error; ética; final; fruto; golpe; grito; hielo; humor; joven; jugar; letra; línea; metro; mitad; negro; paseo; piano; plato; poema; poner; regla; ritmo; salir; sitio; sudor; traje; varón; viaje; vuelo.

Low-frequency words: acero; andén; asilo; balsa; bicho; boina; clavo; coser; crudo; cutis; denso; diván; élite; enano; flaco; funda; globo; gripe; himno; ídolo; lápiz; licor; lindo; lucir; miope; momia; morbo; mutuo; oasis; opaco; óvulo; polen; rezar; rubor; soplo; tenaz; tibia; trigo; veloz; yegua; acoso; araña; ataúd; belga; blusa; burla; cloro; colmo; crema; cuota; danés; diosa; dogma; faena; freno; furor; golfo; hábil; honra; indio; lavar; limón; logro; lunar; mixto; monja; multa; navío; oliva; oveja; ozono; prosa; robot; sesgo; tedio; tenso; tinte; tripa; verbo; zurdo.

Nonwords: avigo; abera; árnel; rebir; bumpa; creje; cegra; codir; gébel; cloma; durle; eciva; láril; leriz; plade; viego; braje; necue; horol; imeon; zaigo; fepos; fícer; fodal; paido; galiz; huive; mesir; mauza; claño; teita; banio; riama; runzo; refal; sulbe; secis; usial; fesir; diris; eltor; asago; abial; bulma; trafo; canve; cijel; glane; cempa; dodus; gauño; ebrir; écito; vimal; drito; hulpe; treco; hauro; huder; jobun; zunar; lirra; niseo; peclo; cucaz; geclo; sarea; peiso; clado; soego; cejor; becla; ricno; racir; ritia; bumor; plave; galal; vaipe; vieno; adina; ardin; acaco; basma; lirro; biara; flajo; conir; gruco; munis; dorgo; hivén; écare; evino; drado; gunca; llopo; brige; huano; ícoma; bábiz; legor; lisco; fusir; miabe; mogio; merso; mutid; oisel; ocica; ózula; ponin; lebar; sujor; todro; selaz; sibio; chego; delaz; yeged; agoca; aciga; atiez; relma; druna; bunsa; drolo; cosgo; clega; cilta; vanís; deuma; decma; naeto; plelo; fuler; gulvo; nápil; horpa; intia; fazar; fisón; focro; burar; mecto; morza; micta; hadao; oraza; ocego; ojoca; froma; robay; tesmo; tesia; sengo; ronte; brefa; derpo; jundo.

Received July 28, 2019
Revision received June 12, 2020
Accepted June 15, 2020 ■