



When a Sunny Day Gives You Butterflies: An Electrophysiological Investigation of Concreteness and Context Effects in Semantic Word Processing

Laura Bechtold^{ID}, Christian Bellebaum, and Marta Ghio

Abstract

■ Theories on controlled semantic cognition assume that word concreteness and linguistic context interact during semantic word processing. Methodological approaches and findings on how this interaction manifests at the electrophysiological and behavioral levels are heterogeneous. We measured ERPs and RTs applying a validated cueing paradigm with 19 healthy participants, who performed similarity judgments on concrete or abstract words (e.g., “butterfly” or “tolerance”) after reading contextual and irrelevant sentential cues. Data-driven analyses showed that concreteness increased and context decreased negative-going deflections in broadly distributed bilateral clusters covering the N400 and N700/late positive component time

range, whereas both reduced RTs. Crucially, within a frontotemporal cluster in the N400 time range, contextual (vs. irrelevant) information reduced negative-going amplitudes in response to concrete but not abstract words, whereas a contextual cue reduced RTs only in response to abstract but not concrete words. The N400 amplitudes did not explain additional variance in the RT data, which showed a stronger contextual facilitation for abstract than concrete words. Our results support separate but interacting effects of concreteness and context on automatic and controlled stages of contextual semantic processing and suggest that effects on the electrophysiological versus behavioral level obtained with this paradigm are dissociated. ■

INTRODUCTION

When we process the meaning of a word such as “butterfly,” we access its conceptual representation stored in the semantic memory. Semantic memory comprises our general knowledge about the word’s referent, such as what a butterfly looks like, how it moves, and that you usually see it on a sunny day (Binder & Desai, 2011; Mahon & Caramazza, 2011). Representational content, that is, the conceptually integrated information about a word’s referent, and linguistic context are two factors that influence semantic word processing (for a review, see Hoffman, 2016). One proxy to investigate the effects of representational content is the concrete versus abstract distinction. Concrete words (e.g., “butterfly”) refer to physical entities and their representations tap into richer, sensorimotor information compared with abstract words (e.g., “tolerance”). Priming or cueing paradigms can be used to investigate how context affects word processing. Embedding words in contextual information is thought to relieve demands on semantic control mechanisms, which select one context-relevant while inhibiting other context-irrelevant (aspects of) word meaning (Chiou, Humphreys, Jung, & Lambon Ralph, 2018; Hoffman, McClelland, & Lambon Ralph, 2018; Lambon Ralph, Jefferies, Patterson, & Rogers, 2017).

Traditionally, competing theories tried to explain a processing advantage of concrete over abstract words, the so-called concreteness effect, exclusively with a richer representational content (e.g., the dual coding theory; Paivio, 1991) or higher context availability of concrete than abstract words (e.g., the context availability model; Schwanenflugel, Harnishfeger, & Stowe, 1988). Stand-alone neither theoretical approach accounted for the complex pattern of contextual semantic processing differences between concrete and abstract words: In a range of tasks, contextual embedding either resulted in no more differences between concrete and abstract word processing (Schwanenflugel & Stowe, 1989; Schwanenflugel et al., 1988; Schwanenflugel & Shoben, 1983) or in a residual concreteness effect (Bechtold, Bellebaum, Hoffman, & Ghio, 2021; Hoffman, Binney, & Lambon Ralph, 2015). Neuroscientific evidence suggested that representational content and context availability rely on interacting neural mechanisms during semantic word processing (Hoffman et al., 2015; Jessen et al., 2000; Holcomb, Kounios, Anderson, & West, 1999).

The controlled semantic cognition framework (Hoffman et al., 2018; Lambon Ralph et al., 2017) combined these empirical observations and theoretical considerations on differential semantic control effects on concrete and abstract word processing (e.g., the hub-and-spokes model; Patterson, Nestor, & Rogers, 2007) in a neurocomputational model of semantic processing. In an

extensive line of research resulting in this framework, the authors systematically applied a similarity judgment task (SJT; originally called synonym judgment task in previous studies, Bechtold et al., 2021; Hoffman, 2016), which requires participants to choose the word most similar to a concrete or abstract probe among three test words. The line of research included investigations in neurological samples (Hoffman, Jones, & Lambon Ralph, 2013; Almaghyuli, Thompson, Lambon Ralph, & Jefferies, 2012; Hoffman & Lambon Ralph, 2011; Jefferies, Patterson, Jones, & Lambon Ralph, 2009) and in healthy participants, where the SJT was embedded in a cueing paradigm with contextual or irrelevant cue sentences (Hoffman et al., 2015; Hoffman, Jefferies, & Lambon Ralph, 2010). Among the research on healthy participants, an fMRI study (Hoffman et al., 2015) identified distinct neural correlates sensitive to word concreteness and demands on semantic control, namely, the anterior temporal lobe and the inferior frontal gyrus, respectively. RT data showed that especially abstract words profited from contextual embedding, which we replicated in German in a recent behavioral study (Bechtold et al., 2021). Altogether, this research, in combination with computational simulations (Hoffman et al., 2018), substantiated the evidence for the assumed interplay of concreteness- and context-driven semantic processes and especially its neural basis.

Even though the SJT cueing paradigm is a powerful tool to investigate contextual semantic processing, based on neurological and neuroimaging data alone it remains unclear whether the effects of representational content and context emerge at stages of (early) automatic versus (late) strategic retrieval and/or integration. Therefore, it is crucial to complement the extant functional neuroimaging and neurological findings with measures of high temporal resolution (Hauk, 2016). In this study, we thus adopted the SJT cueing paradigm to measure ERPs with excellent temporal resolution during the contextual semantic processing of concrete and abstract words.

Previous ERP studies showed electrophysiological concreteness effects in the form of higher amplitudes for concrete than abstract words at distinct processing stages reflected in the N400 and N700 ERP components (Bechtold, Ghio, & Bellebaum, 2018; Barber, Otten, Kousta, & Vigliocco, 2013; Adorni & Proverbio, 2012; West & Holcomb, 2000; Kounios & Holcomb, 1994). The N400, a negativity peaking at 300–500 msec poststimulus, is considered as a marker of semantic retrieval and integration (for a review, see Kutas & Federmeier, 2011). It reflects a stronger involvement of semantic activation or integration processes driven by the relatively richer (multimodal sensorimotor) information for concrete than abstract words, which usually goes along with better behavioral performance (but see Barber et al., 2013, for a dissociation of N400 and RTs; Lau, Phillips, & Poeppel, 2008). The N700 is a late ERP component, starting around 500 msec and lasting up to 1000 msec poststimulus. A higher N700 amplitude reflects (top–down) retrieval of information

(Adorni & Proverbio, 2012) or imagery processes when the task demands it (Gullick, Mitra, & Coch, 2013). The N700 concreteness effect possibly reflects the task-dependent strategic retrieval of visual information at a later stage of semantic processing driven by concrete words' higher imageability and tasks eliciting imagery processes (Bechtold, Ghio, & Bellebaum, 2018; Barber et al., 2013; West & Holcomb, 2000).

Contextual embedding and priming, irrespective of word concreteness, have been found to reduce centroparietal N400 amplitudes (Kotchoubey & El-Khoury, 2014; Ortu, Allan, & Donaldson, 2013; Kutas & Federmeier, 2011; Lau et al., 2008; Holcomb et al., 1999). Contextually reduced N400 amplitudes have been interpreted to reflect automatic preactivation of conceptual information in the sense of spreading activation (Pulvermüller, 1999; Collins & Loftus, 1975), semantic retrieval facilitation through prediction mechanisms (Lau, Weber, Gramfort, Hamalainen, & Kuperberg, 2016), or facilitated post-lexical semantic integration (Steinhauer, Royle, Drury, & Fromont, 2017). A late positive component (LPC; sometimes referred to as P600), emerging 600–1000 msec after stimulus onset, showed larger centroparietal amplitudes in related versus unrelated priming conditions (Meade & Coch, 2017; Grieder et al., 2012; Bouaffre & Faita-Ainseba, 2007). The LPC is thought to reflect controlled post-lexical integration processes, which might be more substantiated when a word is embedded in richer (related) contexts (Hill, Strube, Roesch-Ely, & Weisbrod, 2002). In summary, the effects of concreteness and context on electrophysiological measures are partly dissociated from each other and from behavioral measures.

So far, only few ERP studies directly investigated the interplay of concreteness and context-driven processes. Three studies compared concrete and abstract word processing in single-word semantic priming with a visual (Grieder et al., 2012; Wirth et al., 2008) or acoustic presentation (Swaab, Baynes, & Knight, 2002), and one embedded the words in sentences (Holcomb et al., 1999). Two of these ERP studies focused on the N400 in fixed time windows (300–500 msec in Holcomb et al., 1999; 350–650 msec in Swaab et al., 2002), whereas two followed a data-driven approach (Grieder et al., 2012; Wirth et al., 2008). All four studies found main effects of concreteness (concrete > abstract) and context (unrelated > related) on the N400 in line with the literature reviewed above. Furthermore, two of these studies found an interaction effect of concreteness and context on ERP amplitudes. Holcomb et al. (1999) reported that the anterior N400 concreteness effect was cancelled out when words were embedded in a congruent (vs. anomalous or neutral) sentence. The data-driven approach applied by Wirth et al. (2008) revealed an interaction on the N400 latency: An amplitude reduction by congruent primes occurred later for concrete (512–524 msec) than abstract (444–568 msec) words. Even though the applied analysis allowed only an estimate of the true onset of effects (Sassenhagen & Draschkow,

2019), the finding suggests that previous studies with fixed N400 time windows might have overlooked a potential latency modulation. On the behavioral level, only Grieder et al. (2012) reported independent beneficial effects of concreteness and context on lexical decision speed. In the other studies, RT differences following the context manipulation were either confounded by the tasks, which required different responses to the related versus unrelated conditions (congruency judgment task in Holcomb et al., 1999; semantic judgment in Swaab et al., 2002) or not available (passive reading task in Wirth et al., 2008). Even though the electrophysiological interaction pattern suggests that context differentially affects concrete and abstract word processing, no study has yet directly assessed electrophysiological effects within the controlled semantic cognition framework.

Given the heterogeneity of previous methods and findings, the main aim of the current study was to investigate the temporal dynamics of the effects of representational content (in this case word concreteness) and contextual information on semantic word processing via ERP measures. A secondary aim was to relate the electrophysiological to behavioral effects. To achieve these aims, we used the original SJT cueing paradigm (as described by Hoffman et al., 2015), translated the original English stimuli to German, optimized them (as in Bechtold et al., 2021, Experiment 2), and adapted the procedure to assess ERP data. ERP waveforms were examined with data-driven, nonparametric cluster-based permutation analyses, which allowed us to detect concreteness and context effects without restrictive a priori assumptions on time windows and electrodes (Maris & Oostenveld, 2007). As behavioral measure, we analyzed SJT RTs with linear mixed-effects (LME) analyses as reported in our previous behavioral investigation (Bechtold et al., 2021).

We expected an electrophysiological concreteness effect, with higher negative-going ERP amplitudes for concrete compared with abstract words in the N400 (approximately between 300 and 500 msec) and N700 time window (~500–1000 msec). Concerning the context effect, we expected ERP effects with lower negative-going amplitudes in line with a reduced N400 (~300–500 msec) and enhanced LPC (~500–1000 msec) after contextual compared with irrelevant cues. Crucially, we also expected an interaction, that is, a differential influence of contextual information on concrete versus abstract word processing. We refrained from a priori assumptions on the corresponding spatiotemporal cluster for this effect, as concreteness and context have been shown to go along with opposing N400 and late ERP modulations, namely, enhanced (Barber et al., 2013; Lee & Federmeier, 2008; Swaab et al., 2002; West & Holcomb, 2000) and reduced (negative) amplitudes (Kotchoubey & El-Khoury, 2014; Ortu et al., 2013; Kutas & Federmeier, 2011; Lau et al., 2008; Holcomb et al., 1999), respectively. On the behavioral level, we expected to replicate the concreteness effect (with faster similarity judgments for concrete than

abstract words) and the context effect (with faster similarity judgments for contextual than irrelevant cues). Contextual facilitation (i.e., reduced RTs) should be stronger for abstract than concrete words, replicating previous behavioral findings (Bechtold et al., 2021; Hoffman et al., 2015). Finally, we aimed to explore the relationship between behavioral and ERP data by using ERP amplitudes in relevant time windows as predictors for SJT RTs. Here, we expected the respective ERP amplitudes to be significant predictors of RTs.

METHODS

Participants

Twenty-two healthy young adults voluntarily participated in this study. Criteria for participation were German as mother language, no history of psychiatric or neurological disease or dyslexia, a normal or corrected-to-normal visual acuity, and right-handedness. The Edinburgh Handedness Inventory (Oldfield, 1971) indicated right handedness for all participants but one, who scored -0.55 (scores for the other participants ranged from 0.57 to 1, $M = 0.87$, $SD = 0.15$). This participant was excluded from data analysis, as well as two others, one due to technical artifacts in the EEG data (less than 70% of artifact-free trials in the averaged ERP waveforms) and one due to defective electrodes during recording. The final sample for all analyses thus consisted of 19 adults (5 men and 14 women) aged 18–31 years ($M = 22.84$ years, $SD = 3.75$ years). Seventeen participants were students with at least a university entrance qualification; two had finished secondary school and completed vocational training. The size of the final sample was in line with sample sizes of 16–20 participants in previous EEG studies, which found a significant interaction of concreteness and context (Wirth et al., 2008; Holcomb et al., 1999). For the ERP analyses, we did not conduct an a priori power analysis as the power of the dependent-samples t tests per time-electrode sample serving, as test statistic for the applied nonparametric cluster-based permutation analyses (see below for details) does not correspond to the power at the reported cluster level (Maris, 2012). For RT analysis, we did not conduct a priori power analysis as it aimed to replicate previous findings obtained in a larger sample (Bechtold et al., 2021). The study was in line with the ethical standards defined in the Declaration of Helsinki and was approved by the ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf. We collected written informed consent from all volunteers before participation, for which they received course credit or a monetary compensation.

Stimuli and Material

We used the modified stimuli of the SJT by Hoffman et al. (2015) translated from English into German in our

previous behavioral study (Experiment 2 in Bechtold et al., 2021). Modifications with respect to the original English stimuli included substituting adjectives by nouns or verbs to reduce variability in grammatical class of the probe and target words, adjustments to balance concrete and abstract probes for letter length and arousal, and reformulation of the second cue sentence to avoid repetition of the probe as had been the case in the original study (for more details on modifications, see Bechtold et al., 2021). The thereby carefully constructed German stimuli were divided into the abstract versus concrete category based on a median split applied to the preexperimental concreteness ratings (see below). Two words that were originally categorized as concrete and two words that were categorized as abstract by Hoffman et al. (and in our replication study, Bechtold et al., 2021) were now assigned to the other category, respectively. Furthermore, the preexperimental ratings of our German stimuli by native German speakers (for details, see Bechtold et al., 2021) ensured that concrete probes received significantly higher imageability, context availability, and concreteness ratings than abstract probes (see Table 1A). According to normative studies, these measures are considered discriminatory dimensions of concrete versus abstract words across languages (Yao, Wu, Zhang, & Wang, 2017; Della Rosa, Catricala, Vigliocco, & Cappa, 2010; Altarriba, Bauer, & Benvenuto, 1999). Concrete and abstract probes did not differ in length (number of letters), frequency of occurrence in written and spoken language, arousal, valence, and association with emotional experience (see Table 1A). The full set of stimuli, including the psycholinguistic ratings, is available in the open access repository (<https://osf.io/nvufe/>).

For each probe, we had three test words, from which the participants had to choose the word that was most similar to the probe. One test word was thus a semantically similar target word, and two test words were semantically unrelated foils. Because a differential reliance on similarity-based and associative relations between concrete and abstract words has been postulated (Crutch & Warrington, 2005), the relation of the probes and their target words was quantified in additional ratings of their similarity (i.e., “How similar are the two words? How well can you put them in a common category?”) and association strength (i.e., “How strongly associated are the two words? How well do they form a common context?”). Concrete probe–target pairs received higher ratings of association and similarity-based relatedness than abstract probe–target pairs (see Table 1B). We ruled out the potential influence of these differences on RTs by conducting covariate analyses (see below). Please note that electrophysiological results were independent of the probe–target relation, as we measured ERPs in response to the probe, which appeared 1 sec before the target/foil presentation (see the Design and Data Analysis section).

The contextual cues, each consisting of two short, positively formulated sentences, put each probe in a meaningful context, without containing the probe or a direct

antonym/synonym. The cues often described situations, in which the probe could occur (e.g., “It was a sunny day. Many insects fluttered around outside.”—for the concrete probe “butterfly”) or paraphrased a definition of the probe’s meaning (e.g., “I am against racism. I try to be open-minded.”—for the abstract probe “tolerance”). Cues for concrete and abstract probes did not differ in letter and word count, arousal, and valence, but concrete cues received significantly higher imageability and context availability ratings (see Table 1C; for details on the rating procedure, see Bechtold et al., 2021). Probes were divided into two sets, A and B, with 50 concrete and 50 abstract probes each. To create irrelevant cues, the contextual cues were randomly reassigned to probes within each set. One half of the participants saw probe set A with contextual cues and set B with irrelevant cues, for the other half vice versa (procedure as in Bechtold et al., 2021; Hoffman et al., 2015).

We conducted an additional rating on how well the probe fit into the context provided by the cue on Likert scales from 1 (*not at all*) to 7 (*strongly fitting*) with 30 German native speakers ($n = 15$ for each of the two parallel versions of the experimental stimuli). A 2×2 mixed ANOVA on these context strength ratings with the factors Cue (contextual, irrelevant) and Concreteness (concrete, abstract) revealed a significant main effect of Cue, $F(1, 198) = 6538.33, p < .001, \eta_p^2 = .971$, with higher context strength ratings for contextual ($M = 6.10, SE = 0.04$) than irrelevant ($M = 1.65, SE = 0.04$) cues. The analysis revealed no significant main effect of Concreteness, $F(1, 198) = 3.72, p < .055, \eta_p^2 = .018$, but a significant Cue \times Concreteness interaction effect, $F(1, 198) = 17.73, p < .001, \eta_p^2 = .082$ (for descriptive statistics, see Table 1C). Post hoc pairwise comparisons showed that there was no difference in context strength for contextual cues between concrete and abstract words, $p = .171$, whereas for irrelevant cues, concrete words received lower ratings than abstract words. The higher context strength of irrelevant cues for abstract words can likely be driven by the abstract words’ higher semantic ambiguity (Hoffman et al., 2013). Notably, for both concrete and abstract words, contextual cues received a significantly higher rating than irrelevant cues, both $p < .001, d > 5.04$.

Procedure

Data acquisition took place in single-subject testing sessions in an electrically shielded EEG laboratory at Heinrich Heine University. After giving written, informed consent, participants filled in a demographic questionnaire and the Edinburgh Handedness Inventory. In the meantime, EEG electrodes were attached (for details, see the EEG Recording and Preprocessing section). Stimulus timing and response recording during the EEG task were controlled by the software Presentation (Version 17.0, Neurobehavioral Systems, Inc.) on a Windows 10 Dell Intel Premium PC with a 22-in. LED Dell monitor with 1680 \times

Table 1. Descriptive and Inferential Statistics of the Psycholinguistic Variables for (A) Probes, (B) Relation Strength between Probes and Targets, and (C) Cue Sentences

| Psycholinguistic Variable | Concrete | Abstract | Inferential Statistics | | | |
|------------------------------------|---------------|---------------|------------------------|-----------|----------|----------|
| | <i>M (SE)</i> | <i>M (SE)</i> | <i>t</i> | <i>df</i> | <i>p</i> | <i>d</i> |
| (A) Probes | | | | | | |
| Length (letters) | 7.70 (0.25) | 8.06 (0.25) | 1.05 | 198 | .296 | 0.15 |
| Frequency (written) ^a | 48.04 (11.24) | 47.59 (8.20) | 0.03 | 186 | .974 | 0.00 |
| Frequency (spoken) ^b | 38.51 (9.69) | 22.13 (6.48) | 1.39 | 194 | .165 | 0.20 |
| Arousal | 2.54 (0.10) | 2.59 (0.09) | -0.42 | 198 | .675 | 0.06 |
| Valence | 0.36 (0.11) | 0.28 (0.10) | 0.5 | 198 | .616 | 0.07 |
| Emotional experience | 2.96 (0.14) | 2.59 (0.13) | 1.94 | 198 | .054 | 0.27 |
| Imageability | 6.07 (0.06) | 2.98 (0.07) | 32.76 | 198 | < .001 | 4.63 |
| Context availability | 5.74 (0.04) | 4.29 (0.07) | 17.33 | 154.38 | < .001 | 2.45 |
| Concreteness | 6.18 (0.07) | 2.52 (0.06) | 38.04 | 198 | < .001 | 5.38 |
| (B) Probe–target relation | | | | | | |
| Association | 5.84 (0.06) | 5.61 (0.06) | 2.77 | 198 | .006 | 0.39 |
| Similarity | 5.53 (0.07) | 5.18 (0.07) | 3.49 | 198 | < .001 | 0.49 |
| (C) Cue sentences | | | | | | |
| Length (letters) | 58.89 (0.90) | 59.06 (1.07) | -0.12 | 192.34 | .903 | 0.02 |
| Length (words) | 9.74 (0.17) | 9.86 (0.17) | -0.49 | 198 | .623 | 0.07 |
| Arousal | 2.55 (0.13) | 2.39 (0.11) | 0.93 | 198 | .352 | 0.13 |
| Valence | 4.06 (0.10) | 3.84 (0.09) | 1.58 | 198 | .117 | 0.22 |
| Imageability | 5.18 (0.12) | 3.30 (0.15) | 10.09 | 188.69 | < .001 | 1.43 |
| Context availability | 5.09 (0.09) | 4.13 (0.11) | 6.81 | 186.96 | < .001 | 0.96 |
| Context strength (contextual cues) | 6.16 (0.05) | 6.04 (0.07) | 1.37 | 198 | .171 | 0.19 |
| Context strength (irrelevant cues) | 1.47 (0.05) | 1.82 (0.06) | -4.46 | 195 | < .001 | 0.63 |

Imageability, context availability, concreteness, emotional experience, arousal, context strength, and the strength of the probe–target relation based on association and similarity were rated on 1–7 Likert scales, and valence was rated on a scale from -3 to +3. Independent-samples *t* tests compared the psycholinguistic variables for concrete and abstract words. *n* = 100 per condition, except for frequency (written: *n*_{concrete} = 95, *n*_{abstract} = 93; spoken: *n*_{concrete} = 100, *n*_{abstract} = 96).

^a CELEX database (Baayen, Piepenbrock, & Gulikers, 1996) frequency of occurrence of Mannheim lemmas in 1 Mio. words.

^b SUBTLEX-DE database (Brysbaert et al., 2011) frequency of occurrence of case-insensitive lemmas.

1050 pixel resolution and a refresh rate of 60 Hz. All stimuli were presented in white letters (font: Arial, font size: 30 pt) on black background.

The experimenter read the instructions of the experimental SJT presented on the computer screen to the participants. Figure 1 depicts the timing of the experimental trials. In the SJT, after a fixation cross centrally presented for 0.5 sec, either a contextual or an irrelevant cue appeared in the center of the screen for 5 sec. Then, the probe appeared alone on the screen for 1 sec, which included the 100–1000 msec time interval considered in

the cluster analysis (see below). Please note that this period was introduced to avoid that eye movements and other motor artifacts that affected the ERP in the critical period after probe presentation (different from Bechtold et al., 2021; Hoffman et al., 2015, in which probe and target words appeared simultaneously). Finally, the target word and foils appeared below the probe until a response was given or for a maximum of 4 sec. Participants were instructed to read the cue sentences carefully and were informed that the cue sentences could or could not be semantically related to the probe. Participants were

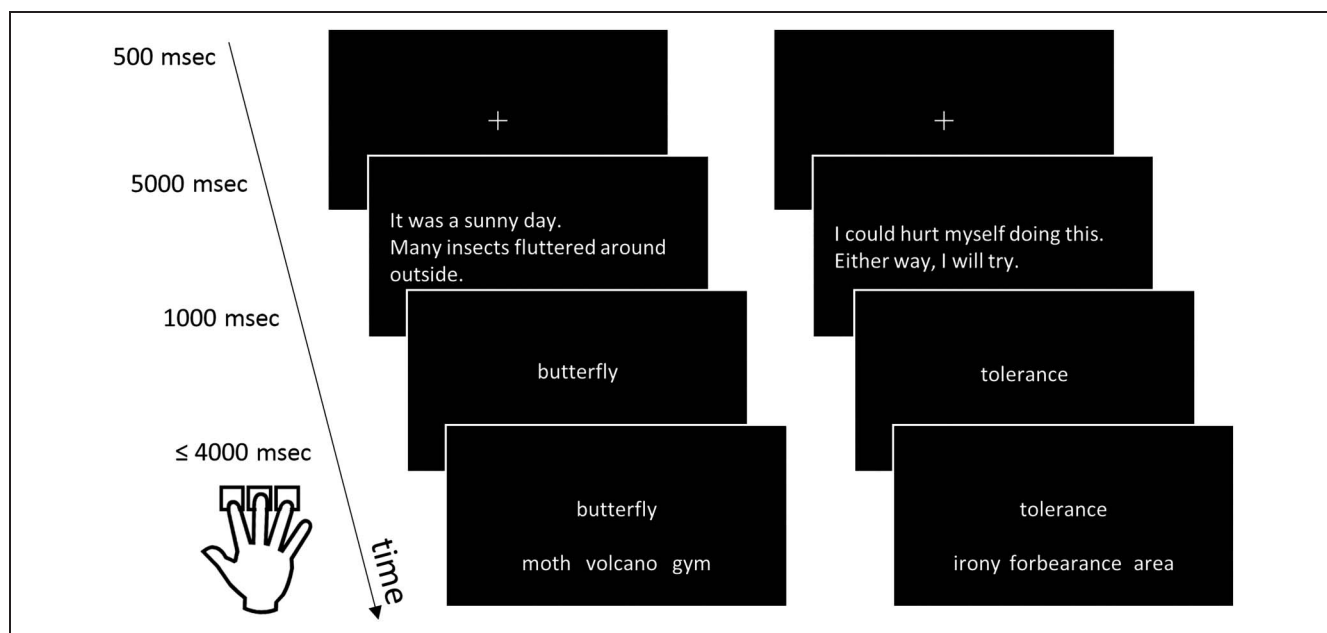


Figure 1. Timing and exemplary stimuli of an experimental trial of the SJT with (left) a concrete probe presented with a contextual cue and (right) an abstract probe presented with an irrelevant cue.

further instructed to choose the word that was most similar to the probe as fast and accurately as possible via button press. The positions of the target word and the two foils were counterbalanced across the three possible positions and randomized over all trials.

Participants used Buttons 3, 4, and 5 of an RB-740 Response Pad (Cedrus Corporation), corresponding to the three positions of the test words on screen. Participants were instructed to position the index, middle, and ring fingers of their (dominant) right hand on the three buttons to reduce movement artifacts and RT variability. Furthermore, participants were asked to look at the fixation cross as soon as it appeared on screen and to keep movements during trials to a minimum. Participants saw two practice trials and could ask questions before starting the experiment. The 200 experimental trials (50 for each condition: concrete contextual, concrete irrelevant, abstract contextual, and abstract irrelevant) were presented in randomized order. Participants could take a self-paced break every 20 trials. Overall, the SJT took about 35 min to be completed.

EEG Recording and Preprocessing

We recorded EEG data with 28 Ag/AgCl passive ring electrodes mounted on a textile BrainCap (Brain Products GmbH) from positions according to the international 10–20 system (electrode sites: F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO9, O1, Oz, O2, and PO10; Chatrian, Lettich, & Nelson, 1985). Four additional electrodes were placed at the outer canthi of the eyes and above and below the left eye to record the horizontal and vertical EOG, respectively. The ground electrode was positioned at

AFz, and the reference electrodes were positioned on the right and left mastoid. We used a BrainAmp DC amplifier (Brain Products GmbH) with a sampling rate of 1000 Hz and an amplitude resolution of 0.1 μV with BrainVision Recorder software (Version 1.20.0506) on a Windows 10 Dell Intel Premium PC. Careful scalp preparation kept impedances below 5 k Ω .

A standard EEG preprocessing procedure in Brain Vision Analyzer (Version 2.1) was conducted. We applied Butterworth zero phase filters for frequencies below 0.1 Hz (time constant: 1.59) and above 30 Hz, with a slope of 24 dB/oct as well as a 50-Hz notch filter. To remove eye blink artifacts from the signal, a fast, restricted independent component analysis with classical sphering was applied to 120 sec of the continuous signal starting after 60 sec of the recording. Visual inspection identified one or two components with a frontally pronounced positivity temporally bound to eye blink artifacts recorded by the vertical EOG. These components were excluded before an inverse independent component analysis back transformation. Subsequently, we segmented the continuous signal into epochs starting 200 msec before and ending 1000 msec after probe onset. The 200 msec before the probe onset were used for a baseline correction. Subsequently, an automatic artifact detection marked rejected epochs based on the following criteria: voltage steps of more than 50 $\mu\text{V}/\text{ms}$, signal changes below 0.1 μV or above 100 μV within 100 msec, or absolute amplitudes above or below $\pm 100 \mu\text{V}$. Electrodes T7 and T8 were excluded from this artifact rejection and the subsequent cluster analyses due to excessive muscle artifacts. We finally averaged all artifact-free trials separately for the four experimental conditions with a minimum of 40 and a maximum of all 50 trials per participant per condition (concrete contextual,

concrete irrelevant, abstract contextual, and abstract irrelevant) entering the averaged ERP waveforms ($M = 48.9$ trials, $SD = 1.8$ trials). We down-sampled the data to 100 Hz in order not to suggest millisecond precision regarding the onset and offset of effects, which the cluster-based permutation analysis cannot provide (Sassenhagen & Draschkow, 2019). Average ERP data are available in the open access repository (<https://osf.io/nvufe/>).

Design and Data Analysis

ERP Data

A nonparametric cluster-based permutation analysis was applied to the averaged ERP waveforms as implemented in the FieldTrip toolbox (Version 20210629; Oostenveld, Fries, Maris, & Schoffelen, 2011) in the MATLAB environment (Version 2020b; code and average ERP data are available in the open access repository). We conducted three separate analyses for testing: (1) the Concreteness main effect (concrete vs. abstract), (2) the Cue main effect (contextual vs. irrelevant), and (3) the Concreteness \times Cue interaction effect (tested by comparing the difference contextual–irrelevant [i.e., the context effect] for concrete vs. abstract words; Bechtold, Ghio, Lange, & Bellebaum, 2018; Maris & Oostenveld, 2007). On the first level, to detect spatiotemporal clusters corresponding to significant differences between conditions, we calculated dependent-samples t tests separately for each time-electrode sample of interest, that is, the above-mentioned 26 electrodes in the time window from 100 to 1000 msec after the probe onset (resulting in 26×91 included samples). The broad time window was chosen based upon visual inspection (see Figure 2) to not only cover the time intervals corresponding to the N400 and LPC/N700 ERP components, on which we based our hypotheses, but also earlier components, so that unexpected effects could have been detected as well (Mensen & Khatami, 2013). Based on spatiotemporal adjacency, we combined samples with at least two neighboring electrodes that reached a cluster-defining threshold of $p < .05$, to clusters. Successive time points were considered adjacent, as well as electrodes with a maximum distance of 0.225 arbitrary units (corresponding to approximately 22.5% of the nasion–inion distance according to the 10–20 system based on the customized acticap-64ch-standard2.mat 2D template layout).

The weighted cluster mass (Hayasaka & Nichols, 2004) served as test distribution for the second-level cluster statistic as it has been shown to deliver the best balance between precision and sensitivity based on a predefined cluster-forming threshold (Mensen & Khatami, 2013). On the second level, we created a reference distribution in form of 1000 random permutations of the data of the two conditions and applying the cluster detection method described above. By comparing the weighted cluster mass of the detected clusters in the ERP data with those in the random distributions, we estimated the p value of each

cluster. This approach controls for multiple comparisons and is less conservative than Bonferroni correction (Maris & Oostenveld, 2007). Most importantly, it allows detecting effects without restrictive a priori assumptions on time windows and electrode sites and avoids compromising validity by reducing the data in a possibly biased way (Mensen & Khatami, 2013). Additionally, it allows exploring the spatiotemporal development of effects, even though the permutation-based cluster analysis does not allow strong inferences on the temporal onset and offset of effects (Sassenhagen & Draschkow, 2019). Please note that the reported cluster extents are descriptive in nature and only approximate the real extent of effects (Sassenhagen & Draschkow, 2019). Furthermore, descriptive statistics for the mean amplitudes in the significant clusters were calculated by extracting and averaging the amplitude values of the time-electrode samples constituting the cluster for the two conditions involved in the respective comparison. Code and results of the ERP analyses are available in the open access repository (<https://osf.io/nvufe/>).

Finally, to further explore the Concreteness \times Cue interaction pattern, we extracted the single-trial ERP amplitudes, averaged across the time-electrode samples involved in the cluster that corresponded to a significant interaction effect in ERP Analysis 3 (see below). The overall 3720 single-trial ERP amplitude data points of all participants were entered into an LME analysis (see Bechtold et al., 2021) conducted with the lme4 package (Version 1.1-26; Bates, Mächler, Bolker, & Walker, 2015) in R (Version 3.6.3). Following recommendations of Baayen and Milin (2010), we refitted the model after excluding trials, whose standardized residuals exceeded 2.5 SD s from the residual mean ($<1.6\%$ of the data). This resulted in 3663 data points, with a mean of 48.2 single-trial RT data points ($SD = 2.2$ data points, ranging from 39 to 50) per participant per condition going into the following model.

Fixed-effects factors included in the model were Concreteness (concrete [+0.5], Abstract [−0.5]) and Cue (contextual [+0.5], irrelevant [−0.5]), as well as their interaction. We included random intercepts for the random-effects factors Participants and Items to control for interindividual variance (Baayen, Davidson, & Bates, 2008). As the inclusion of random slopes would have led to a singular fit, we did not include random slopes. We applied a restricted maximum likelihood approach (Luke, 2017) and estimated degrees of freedom and p values following the Satterthwaite method with the lmerTest package (Version 3.1-3; Kuznetsova, Brockhoff, & Christensen, 2017). We applied simple slope analyses as implemented in the R package jtool (Version 2.0.3) to resolve the interaction. We calculated Cook's distance with the package influence.ME (Version 0.9-9) to verify that no participants exerted a disproportionately strong influence on the model. No participant reached the suggested cutoff of 0.24 (Bollen & Jackman, 1985) calculated as $4/n - p$ ($n =$ sample size = 19; $p =$ number of factors = 2), with values ranging from <0.01 to 0.13 ($M = 0.04$, $SD = 0.03$).

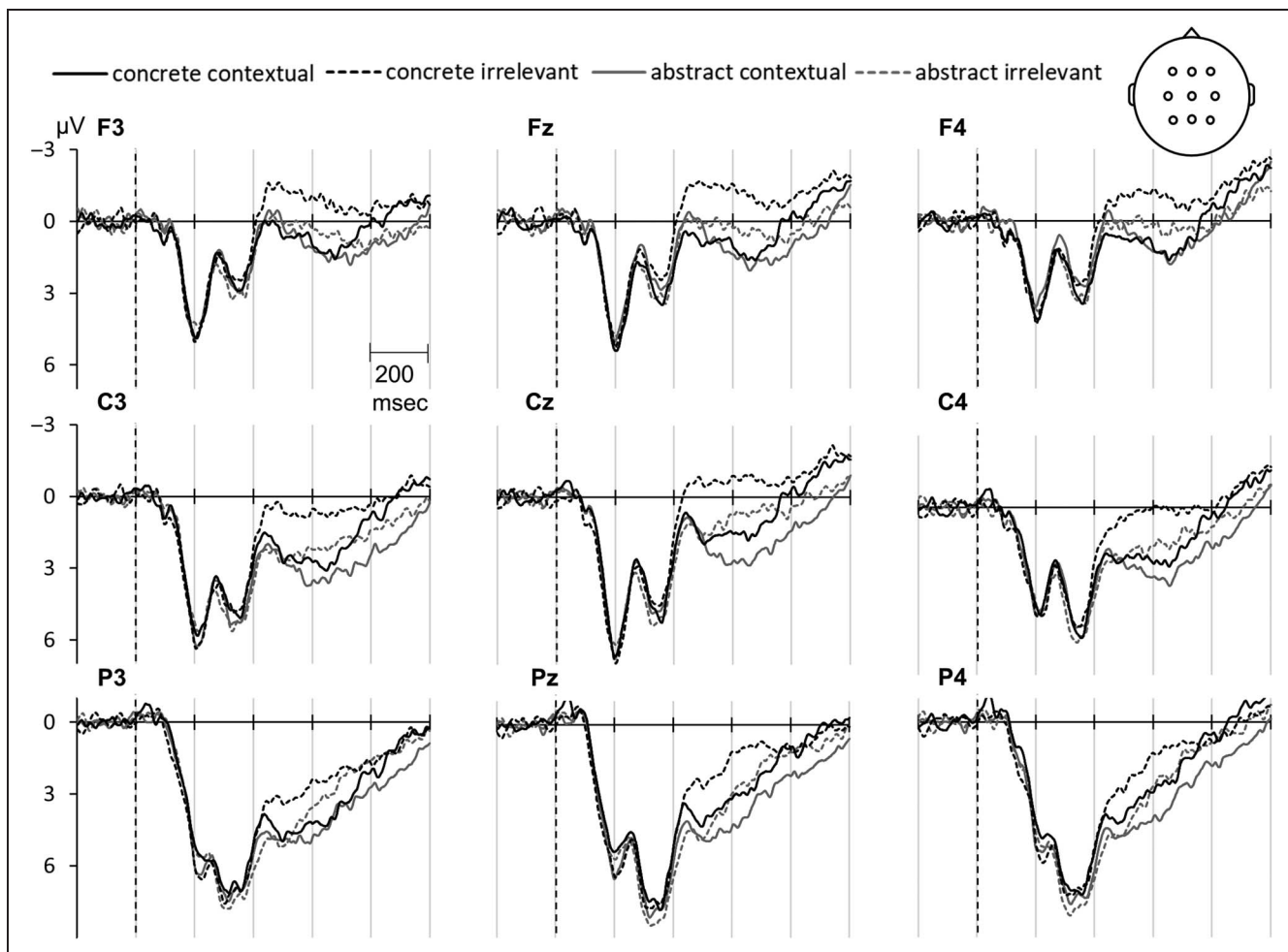


Figure 2. Grand average ($n = 19$) ERP curves from 200 msec before to 1000 msec after the probe onset (dashed vertical line) for the four experimental conditions (concrete contextual, concrete irrelevant, abstract contextual, abstract irrelevant) at nine exemplary electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4; see upper right corner for electrode positions).

To control the effects of the possibly confounding significant differences between concrete and abstract words in cue imageability, cue context availability, and context strength (please note that neither probe–target association nor similarity potentially affected ERP amplitudes as ERPs were measured before target onset), we included the respective rating measures as continuous, normalized predictors into separate LME analyses on the single-trial ERP amplitudes described above. We refitted these covariate models after excluding trials, whose absolute standardized residuals exceeded 2.5 *SDs* from the residual mean (always <1.6% of the data). This resulted in 3663 data points included in each of the reported models. Code and results of the LME ERP analyses are available in the open access repository (<https://osf.io/nvufe/>).

RT Data

We measured RTs from the onset of the three test words to the button press. Raw RT data are available in the open access repository (<https://osf.io/nvufe/>). We refrained

from a priori data trimming, as all RTs laid in a sensible range of 466–3940 msec (Baayen & Milin, 2010). We included only trials with correct responses. Responses were considered correct when the participant chose the similar target word among the three test words (participants reached a mean of 94.7% accuracy, $SD = 22.3\%$). The overall 3573 raw RT data points of all participants were entered into an LME analysis (see Bechtold et al., 2021) as described above. We refitted the model after excluding trials for which standardized residuals exceeded 2.5 *SD* from the residual mean (<2.5% of the data). This resulted in 3484 data points, with a mean of 45.8 single-trial RT data points ($SD = 3.6$ data points, ranging from 32 to 50) per participant per condition going into the following model. We included the same random effect factors as described above for the ERP amplitude model. For participants, we additionally included random slopes for Concreteteness and Cue. No participants exerted a disproportionately strong influence on the model (Cook's distance) with values ranging from <0.01 to 0.16 ($M = 0.06$, $SD = 0.05$).

We further used the single-trial ERP amplitudes from the cluster that corresponded to a significant interaction effect in ERP Analysis 3 as an additional normalized continuous predictor into the original single-trial RT LME model described above. We excluded single-trial RT values from trials containing ERP artifacts, as they did not have corresponding amplitude values, resulting in 3502 data points. We again refitted the model after excluding trials, whose absolute standardized residuals exceeded 2.5 *SDs* from the residual mean (<2.6% of the data). This resulted in 3412 data points, with a mean of 44.9 single-trial RT data points (*SD* = 4.3 data points, ranging from 25 to 50) per participant per condition. To investigate whether this additional predictor explained a significant amount of variance in the RT data, we compared the models with and without the ERP amplitude as predictor via a χ^2 test conducted with the *anova* function implemented in the *car* package (Version 3.0-10). Please note that the model comparison was conducted upon the models without model-specific residual outlier exclusion, as it can only be conducted upon models fitted to the same amount of data points.

To control the effects of the possibly confounding significant differences between concrete and abstract words in cue imageability, cue context availability (see Table 1B), context strength, as well as probe–target association and similarity (see Table 1C), we included the respective rating measures as continuous, normalized predictors into LME analyses on the single-trial ERP amplitudes described above. We refitted these covariate models after excluding trials, whose absolute standardized residuals exceeded 2.5 *SDs* from the residual mean (always <2.6% of the data). This resulted in 3481 (probe–target association covariate), 3482 (probe–target similarity covariate), and 3483 (context strength, cue context availability, and cue imageability

covariates) data points included in the reported models, respectively. Code and results of the LME RT analyses are available in the open access repository (<https://osf.io/nvufe/>).

RESULTS

ERP Data

Figure 2 depicts the grand average ERP waveforms relative to probe onset at nine exemplary electrode sites. Amplitude differences between the experimental conditions were broadly distributed over the scalp and began around 400 msec after probe onset and lasted up to 1000 msec (probe offset). The N400 peaked at ~400–500 msec and was followed by a sustained relative negativity visible in the irrelevant conditions and a less pronounced negativity or even relative positivity in the contextual conditions in the N700/LPC time range (>600 msec).

Analysis 1: Concreteness Main Effect

The first cluster-based permutation analysis indicated a significant Concreteness effect, $p = .003$. Descriptively, the effect corresponded to a large cluster with more negative (i.e., less positive) amplitudes for concrete ($M = 1.09 \mu\text{V}$, $SD = 1.70 \mu\text{V}$) than abstract probes ($M = 2.20 \mu\text{V}$, $SD = 1.68 \mu\text{V}$), extending from approximately 370–1000 msec after probe onset. Within this time interval, a centroparietally pronounced cluster roughly covering the N400, transitioned into a frontocentrally pronounced cluster covering the N700/LPC time range (for the spatiotemporal development, see Figure 3).

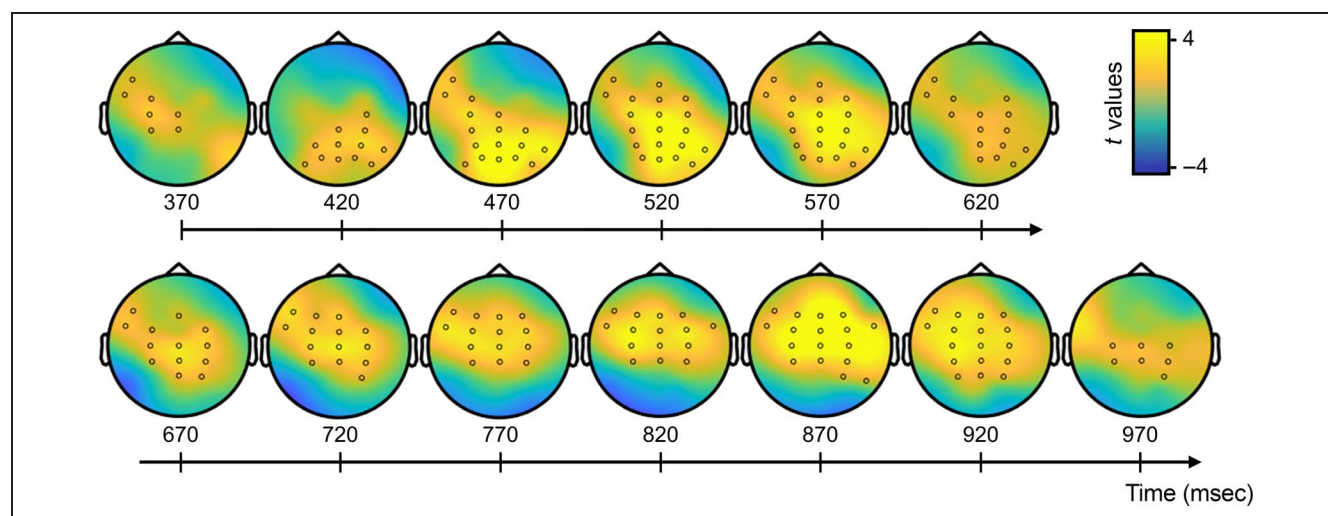
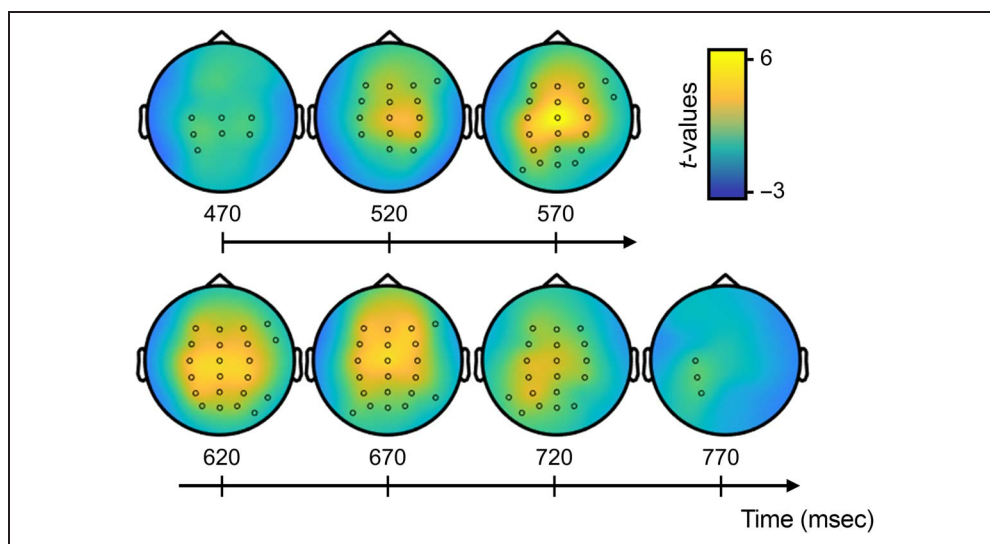


Figure 3. Spatiotemporal development (in 50-msec steps) of the cluster corresponding to a significant amplitude difference between concrete and abstract probes. Black circles mark time-electrode samples for which the concrete versus abstract comparison reached the cluster-defining threshold at the first level ($p < .05$). The color grading represents the distribution of t values. Positive t values indicate a more negative amplitude for concrete than abstract probes.

Figure 4. Spatiotemporal development (in 50-msec steps) of the cluster corresponding to a significant amplitude difference between probes after contextual and irrelevant cues. Black circles mark time-electrode samples for which the contextual versus irrelevant comparison reached the cluster-defining threshold at the first level ($p < .05$). The color grading represents the distribution of t values. Positive t values indicate a more negative amplitude for irrelevant than contextual cues.



Analysis 2: Cue Main Effect

The second analysis indicated a significant context effect, $p < .001$. Descriptively, the effect corresponded to a cluster with more negative (less positive) amplitudes for irrelevant ($M = 1.10 \mu\text{V}$, $SD = 1.99 \mu\text{V}$) than contextual cues ($M = 2.44 \mu\text{V}$, $SD = 1.79 \mu\text{V}$), extending from approximately 470–770 msec after probe onset. Within this time interval, the cluster covering the late N400 and in the N700/LPC time range was broadly distributed over

frontocentroparietal electrode sites (for the spatiotemporal development, see Figure 4).

Analysis 3: Concreteness \times Cue Interaction Effect

The third analysis indicated a significant interaction effect of Concreteness and Cue, that is, a significant difference in the context effects on concrete versus abstract word processing, $p = .031$. Descriptively, the effect corresponded to a cluster

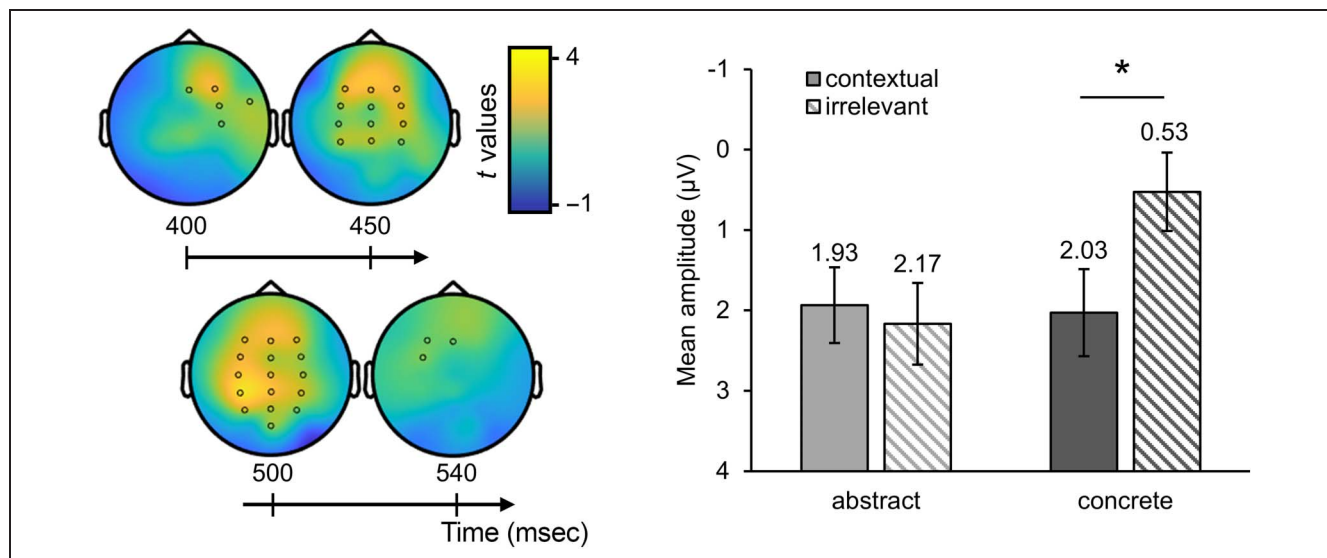


Figure 5. Left: Spatiotemporal development (in ~ 50 msec steps) of the cluster corresponding to a significant interaction, that is, a significant difference between the context effects for concrete versus abstract probes. Black circles mark time-electrode samples for which the comparison of the context effect for concrete and abstract probes reached the cluster-defining threshold at the first level ($p < .05$). The color grading represents the distribution of t values. Positive t values indicate a larger amplitude difference between irrelevant and contextual cues for concrete than abstract probes. Right: Mean amplitude of the negative-going deflection in the four experimental conditions (concrete contextual, concrete irrelevant, abstract contextual, and abstract irrelevant) in the detected cluster. Error bars depict ± 1 standard error.

with a stronger context effect (i.e., the amplitude difference of contextual minus irrelevant cues) for concrete than abstract probes in a time window extending from approximately 400–540 msec after probe onset and thus consistent with the N400. The cluster was broadly distributed over the scalp in the early phase of the time window, whereas in the later phase, it was most pronounced over frontotemporal electrode sites, mostly over the left hemisphere (for the spatiotemporal development, see Figure 5, left). The descriptive pattern of the mean amplitudes in this cluster (Figure 5, right) suggests that a contextual cue reduced the negative-going amplitudes in response to concrete probes but not in response to abstract probes. The post hoc LME analysis (for inferential statistics, see Table 2A) confirmed that participants' mean amplitudes

in this cluster differed significantly between the contextual versus irrelevant condition for concrete, $\beta = 1.24$ ($SE = 0.33$), $p < .001$, but not abstract probe processing, $\beta = -0.24$ ($SE = 0.33$), $p = .476$.

Additional covariate analyses showed that the reported inferential pattern of the interaction effect on the ERP amplitude in the detected cluster was robust (all $ps \leq .006$) against potential confounds by the cues' imageability, context availability ratings, and the context strength (see Table 2B, C, D, respectively). Notably, only in the context strength covariate analysis, the main effect of cue was not significant, $p = .333$, which was expected, as the context strength differentiates between contextual and irrelevant cues. All other main effects remained significant throughout the covariate analyses, all $ps \leq .040$.

Table 2. Inferential Statistics of the Single-trial ERP Amplitude LME Analyses

| Predictor | β | SE | df | t | p |
|---|---------|------|---------|-------|------|
| (A) Post hoc analysis | | | | | |
| Cue | 0.50 | 0.24 | 3466.56 | 2.14 | .033 |
| Concreteness | -0.79 | 0.26 | 197.00 | -3.04 | .003 |
| Concreteness \times Cue | 1.48 | 0.47 | 3466.50 | 3.14 | .002 |
| (B) Cue imageability covariate analysis | | | | | |
| Cue imageability | -0.13 | 0.15 | 965.00 | -0.91 | .364 |
| Cue | 0.48 | 0.24 | 3465.86 | 2.05 | .040 |
| Concreteness | -0.64 | 0.31 | 280.74 | -2.09 | .037 |
| Concreteness \times Cue | 1.49 | 0.47 | 3469.51 | 3.16 | .002 |
| (C) Cue context availability covariate analysis | | | | | |
| Cue context availability | -0.04 | 0.13 | 1163.42 | -0.29 | .775 |
| Cue | 0.50 | 0.24 | 3466.28 | 2.13 | .033 |
| Concreteness | -0.76 | 0.28 | 234.33 | -2.72 | .007 |
| Concreteness \times Cue | 1.48 | 0.47 | 3467.70 | 3.15 | .002 |
| (D) Context strength covariate analysis | | | | | |
| Context strength | 0.76 | 0.50 | 743.96 | 1.52 | .129 |
| Cue | -0.96 | 0.99 | 824.99 | -0.97 | .333 |
| Concreteness | -0.76 | 0.26 | 198.91 | -2.89 | .004 |
| Concreteness \times Cue | 1.33 | 0.48 | 3585.28 | 2.76 | .006 |

The ERP amplitude was extracted from the cluster detected to reflect a significant interaction effect of Concreteness and Cue in Analysis 3.

RT Data

Figure 6 shows the descriptive RT data, whereas Table 3A summarizes the inferential pattern of the LME RT analysis. We found significant main effects of Concreteness, $p < .001$, and Cue, $p = .004$. RTs for concrete words ($M = 1291$ msec, $SD = 417$ msec) were faster than those for abstract words ($M = 1543$ msec, $SD = 530$ msec), and contextual cues ($M = 1386$ msec, $SD = 476$ msec) led to faster RTs than irrelevant cues ($M = 1437$ msec, $SD = 504$ msec). The Concreteness \times Cue interaction effect was significant, $p = .001$. Simple slope analyses showed that a contextual cue significantly reduced the RTs for abstract probes, $\beta = -104.18$ ($SE = 23.18$), $p < .001$, but not for concrete probes, $\beta = -24.53$ ($SE = 22.54$), $p = .285$. Additional covariate analyses showed that the reported inferential pattern the Concreteness \times Cue interaction effects was robust (all $ps \leq .006$) against potential confounds by the association- and similarity-based probe–target relation as well as the cues' imageability and context availability ratings and context strength (see Table 3B, C, D, E, F, respectively). As for the post hoc ERP analysis, the main effect of cue was not significant in the context strength covariate analysis, $p = .057$, whereas other main effects remained significant throughout the covariate analyses, all $ps \leq .005$.

To explore the relationship between behavioral and ERP data, we included the single-trial ERP amplitude averaged across the time-electrode samples comprised in the cluster detected in ERP Data Analysis 3 as an additional predictor into the single-trial RT LME analysis reported above. The inferential pattern is shown in Table 3G. The main effects of Cue and Concreteness and their interaction remained significant, all $ps \leq .005$, whereas neither the main effect of ERP amplitude nor any of its interactions were significant, all $ps \geq .175$. Including the ERP amplitude did not explain any additional variance (Bayesian information criterion = 52,966) compared with the original model (Bayesian information criterion = 52,936), $\chi^2 = 2.96$, $p = .565$.

Figure 6. Mean RTs in the four experimental conditions; error bars depict the 90% confidence interval.

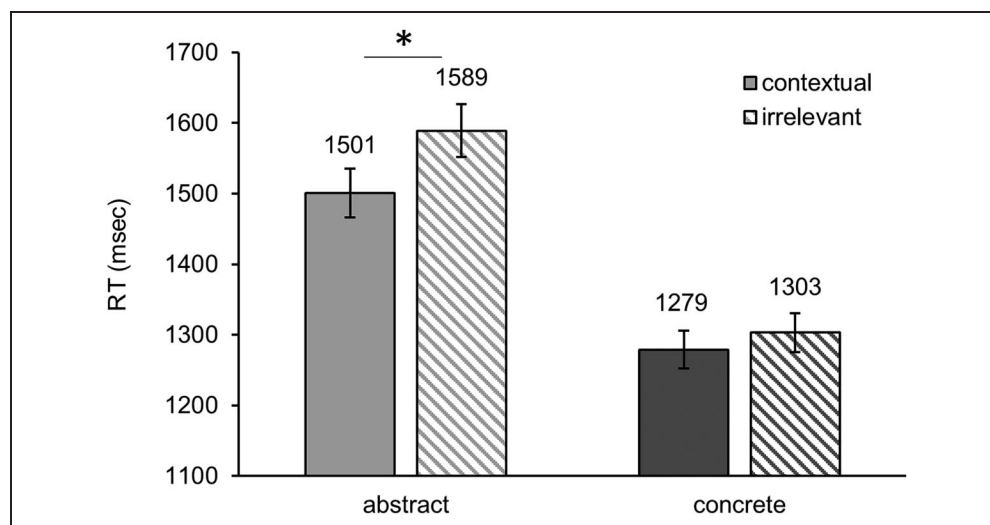


Table 3. Inferential Statistics of the Single-trial RT LME Analyses

| Predictor | β | SE | df | t | p |
|---|---------|-------|---------|-------|--------|
| (A) General analysis | | | | | |
| Cue | -64.35 | 19.33 | 17.03 | -3.33 | .004 |
| Concreteness | -272.05 | 35.10 | 107.76 | -7.75 | < .001 |
| Concreteness \times Cue | 79.65 | 24.40 | 3238.47 | 3.26 | .001 |
| (B) Probe-target similarity covariate analysis | | | | | |
| Probe-target similarity | -129.20 | 13.07 | 196.25 | -9.89 | < .001 |
| Cue | -64.22 | 18.58 | 17.57 | -3.46 | .003 |
| Concreteness | -214.26 | 30.63 | 79.21 | -7.00 | < .001 |
| Concreteness \times Cue | 72.59 | 24.32 | 3241.48 | 2.99 | .003 |
| (C) Probe-target association covariate analysis | | | | | |
| Probe-target association | -117.15 | 13.39 | 197.04 | -8.75 | < .001 |
| Cue | -63.13 | 18.38 | 17.62 | -3.44 | .003 |
| Concreteness | -232.36 | 30.93 | 87.36 | -7.51 | < .001 |
| Concreteness \times Cue | 70.88 | 24.30 | 3239.84 | 2.92 | .004 |
| (D) Cue imageability covariate analysis | | | | | |
| Cue imageability | -24.78 | 9.65 | 3133.52 | -2.57 | .010 |
| Cue | -63.25 | 19.18 | 16.93 | -3.30 | .004 |
| Concreteness | -244.78 | 36.93 | 127.62 | -6.63 | < .001 |
| Concreteness \times Cue | 79.33 | 24.36 | 3236.09 | 3.26 | .001 |

Table 3. (continued)

| Predictor | β | SE | df | t | p |
|---|---------|-------|---------|-------|--------|
| (E) Cue context availability covariate analysis | | | | | |
| Cue context availability | -10.68 | 8.42 | 3257.64 | -1.27 | .205 |
| Cue | -63.91 | 19.14 | 17.01 | -3.34 | .004 |
| Concreteness | -264.30 | 35.97 | 114.88 | -7.35 | < .001 |
| Concreteness \times Cue | 79.26 | 24.38 | 3236.15 | 3.25 | .001 |
| (F) Context strength covariate analysis | | | | | |
| Context strength | 34.51 | 33.59 | 2915.32 | 1.03 | .304 |
| Cue | -128.93 | 67.56 | 1482.61 | -1.91 | .057 |
| Concreteness | -269.86 | 35.40 | 109.85 | -7.62 | < .001 |
| Concreteness \times Cue | 68.97 | 25.20 | 3302.51 | 2.73 | .006 |
| (G) ERP predictor | | | | | |
| Cue | -62.43 | 19.12 | 17.12 | -3.26 | .005 |
| Concreteness | -275.52 | 35.03 | 108.56 | -7.87 | < .001 |
| ERP | 1.33 | 6.58 | 3249.33 | 0.20 | .839 |
| Concreteness \times Cue | 73.53 | 24.74 | 3165.70 | 2.97 | .003 |
| Cue \times ERP | 10.10 | 12.87 | 2697.03 | 0.79 | .433 |
| Concreteness \times ERP | -6.04 | 12.86 | 2361.84 | -0.47 | .639 |
| Concreteness \times Cue \times ERP | 34.38 | 25.34 | 3230.65 | 1.36 | .175 |

DISCUSSION

In this electrophysiological study, we investigated the interplay of concreteness and context effects on semantic word processing assumed by theoretical approaches on contextual semantic processing (Hoffman et al., 2018; Lambon Ralph et al., 2017; Holcomb et al., 1999). We explored its spatiotemporal dynamics by means of non-parametric cluster-based permutation analyses of ERP data acquired in a well-validated cueing paradigm with an SJT (see Bechtold et al., 2021; Hoffman et al., 2010) to complement insights from previous functional neuroimaging data (Hoffman et al., 2015). Electrophysiologically, we found a concreteness effect with higher negative-going amplitudes for concrete than abstract words covering the N400 and N700/LPC time range, as expected. Further in line with our hypotheses, contextual compared with irrelevant cues reduced negative-going amplitudes also in the N400 and N700/LPC time range. Crucially, we also found an interaction of concreteness and cue, with contextual cues reducing negative-going amplitudes in response to concrete words, erasing the (concrete > abstract) concreteness effect in the N400 time range. Unexpectedly, however, no cluster showed a significantly stronger modulation of abstract than

concrete probe processing. Analyses on the behavioral level revealed concreteness and context effects, with faster RTs for concrete (vs. abstract) probes and contextual (vs. irrelevant) cues, in line with our hypotheses. We also replicated the finding of an interaction of concreteness and context, with stronger contextual facilitation for abstract (than concrete) probes, namely, a reduced concreteness effect in the contextual versus irrelevant cue condition (see Bechtold et al., 2021; Hoffman et al., 2015). Unexpectedly, including the ERP amplitudes obtained from the cluster corresponding to the interaction effect as predictor did not explain additional variance in the RT data.

Concreteness Effects

The expected pronounced electrophysiological concreteness effect in the N400 and N700/LPC time range is in line with previous findings showing higher negative amplitudes for concrete than abstract words in both time windows (Bechtold, Ghio, & Bellebaum, 2018; Barber et al., 2013; Gullick et al., 2013; West & Holcomb, 2000; Holcomb et al., 1999; Kounios & Holcomb, 1994). In these studies, the N400 concreteness effect has been

interpreted to reflect a stronger involvement of semantic activation or integration processes driven by the relatively richer (multimodal sensorimotor) information for concrete than abstract words. The visual inspection of the extent of the detected cluster does not suggest a lateralization of the observed N400 (nor N700) concreteness effect, which would have supported the (context-extended) dual coding theory (Welcome, Paivio, McRae, & Joanisse, 2011; Kounios & Holcomb, 1994; Paivio, 1986). Furthermore, in previous studies, which focused on single-word processing, the N400 concreteness effect was most pronounced over frontal areas (Barber et al., 2013; Lee & Federmeier, 2008; Swaab et al., 2002; West & Holcomb, 2000), whereas for our cued probes, the cluster covering the N400 time window was centroparietally pronounced. Because the N400 as marker for semantic integration often shows a rather bilateral parietal topography (Kutas & Federmeier, 2011; Lau et al., 2008), the topographical discrepancy between the concreteness effect of the present study and previous single-word processing studies might be due to the high task relevance of integrating contextual information with the probe's representational content in the cued SJT.

Taken together with visual inspection of the ERP waveforms (see Figure 2), the effect corresponding to a frontally pronounced cluster in the N700/LPC time range can be interpreted as an N700 concreteness effect, which has previously been suggested to reflect strategic mental imagery processes (Bechtold, Ghio, & Bellebaum, 2018; Malhi & Buchanan, 2018; Gullick et al., 2013). The N700 effect can also be interpreted in the sense of more general controlled memory retrieval processes (i.e., not restricted to visual information; Adorni & Proverbio, 2012).

To sum up, the electrophysiological concreteness effects possibly reflect a stronger involvement of multimodal integration (in the N400; Barber et al., 2013) and mental imagery processes (in the N700; Bechtold, Ghio, & Bellebaum, 2018; Gullick et al., 2013) of the richer representational content of concrete than abstract words. In line with this interpretation, concrete probes received higher context availability, imageability, and concreteness ratings, all of which are indicators of semantic richness (Muraki, Sidhu, & Pexman, 2019; Hoffman, 2016). Taken together with the behavioral processing advantage for concrete over abstract words, our results suggest that the assumed richer representational content facilitated their processing (Hoffman, 2016; Grieder et al., 2012).

Context Effects

The significant main effect of context in our findings corresponded to a spatially broadly distributed cluster covering the N400 and N700/LPC time range, in which contextual versus irrelevant cues reduced relatively negative and enhanced relatively positive ERP deflections. Reduced N400 amplitudes in priming paradigms have been interpreted to reflect the facilitation of semantic

processing through preactivation of semantic information (Brouwer, Fitz, & Hoeks, 2012; Lau et al., 2008; Pulvermuller, 1999) or through facilitated prediction of upcoming words (Lau, Holcomb, & Kuperberg, 2013; Lau et al., 2008). After visual inspection of the grand averages in Figure 2, we interpret the later context effect in terms of an enhanced LPC amplitude after semantic priming, which has been frequently reported in the literature (Meade & Coch, 2017; Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Yao & Wang, 2014; Grieder et al., 2012; Bouaffre & Fata-Ainseba, 2007; Hill et al., 2002). The LPC priming effect has been interpreted to reflect post-lexical integration and memory retrieval in the service of strategic prime target-matching processes (Brouwer et al., 2012). Our electrophysiological context effects went along with reduced RTs in the SJT, supporting the interpretation that the reduced N400 and enhanced LPC amplitudes reflect mechanisms in favor of semantic processing. Based on our stimulus design, in which the probes were not syntactically embedded in the cue sentences, we assume to have minimized any lexicosyntactic influence on the processing of the probe, thereby maximizing the influence of semantic priming effects.

An important limitation of the interpretation of the late electrophysiological concreteness and context effects is the temporal overlap of the N700/LPC components in the late time window. Even though based on the reviewed literature and visual inspection of the grand averages we proposed an interpretation in the sense of an N700 concreteness effect and an LPC context effect, the two ERP components with opposing polarity might have affected each other. Future research could apply principal component analyses based on high-density EEG (Pourtois, Delplanque, Michel, & Vuilleumier, 2008) and time-frequency information (Bernat, Nelson, Holroyd, Gehring, & Patrick, 2008) to disentangle the underlying components and thus cognitive processes in the late ERP time window.

Interaction of Concreteness and Context

The core finding of this study is the interaction of concreteness and context, which, however, became evident in a descriptively slightly left-lateralized N400 cluster with a significantly stronger contextual modulation for concrete than abstract probes and, unexpectedly, no cluster with the opposite effect. Specifically, post hoc comparisons of the mean amplitude in the detected cluster showed that contextual cues reduced the N400 in response to concrete words in a way that cancelled out the (concrete > abstract) concreteness effect, which has been reported previously (Holcomb et al., 1999). One explanation for this interaction pattern might be that the N400 reflects a modulation of anterior temporal semantic integration processes (Matsumoto, Iidaka, Haneda, Okada, & Sadato, 2005; Rossell, Price, & Nobre, 2003) rather than context-driven inferior frontal semantic control processes (Van Petten

& Luka, 2006). Context might have especially preactivated the (richer) representational content (Lau et al., 2008) of concrete words, thereby leading to the N400 interaction pattern. Picking up the example used before, we assume that much of the representational content of the “butterfly” had already been retrieved and integrated when preceded by in a “meadow full of insects on a sunny day.”

In the grand averages (see Figure 2; e.g., electrodes C3 and Cz), it seems like the amplitude reduction by the contextual versus irrelevant cues occurred later for abstract compared with concrete words, yielding the interaction time window of reduced amplitudes for concrete but not abstract words. Explorative cluster analyses (one-sample *t* tests of the contextual–irrelevant amplitude difference for abstract and concrete words against zero, investigating the context effect separately for concrete and abstract words) showed that the context effect for abstract words involved fewer frontal electrodes, started later and ended earlier (580–670 msec) than for concrete words (440–750 msec). Our findings regarding the interaction effect thus directly oppose findings by Wirth et al. (2008), who found a stronger contextual N400 reduction for abstract than concrete words in single-word priming (abstract words: 444–568 msec and additionally 492–568 msec, concrete words: 512–524 msec). In their reasoning, Wirth et al. pointed out the possibility that attention directed to semantics in active tasks (referring to Holcomb et al., 1999) might have led to the discrepant findings based in their passive task, which might have also affected the onset latency. Please note that the applied analyses only allow an estimate of the true onset of effects (Sassenhagen & Draschkow, 2019). In how far such attentional mechanisms differentially modulate the strength as well as the latency of N400 context effects for concrete and abstract words has thus to be investigated in future research.

It was unexpected that no cluster showed a significantly stronger modulation of abstract than concrete word processing, which would have been in line with empirical findings of a stronger contextual modulation of behavioral responses to abstract than concrete words (Bechtold et al., 2021; Hoffman, 2016; Hoffman et al., 2015; Schwanenflugel & Shoben, 1983) linked to semantic control processes involving the inferior frontal cortex (Hoffman et al., 2015; Hoffman et al., 2010). Wirth et al. (2008) found such a significantly stronger contextual modulation for abstract than concrete words not only in the N400 time range as described above but also in the early N1–P1 complex (116–140 msec). They traced this early effect back to activity in the left inferior pFC; thus, their results might hint at an early involvement of semantic control mechanisms. In our study, an early (190–230 msec) bilateral centroparietal cluster showed the expected stronger modulation for abstract words, which, however, failed to reach significance ($p = .069$). Therefore, we can neither support nor rule out the possibility of early semantic control effects in our paradigm, especially when acknowledging the risk of false negatives in cluster-based permutation analyses (Sassenhagen &

Draschkow, 2019). From this pattern of results, we would like to hypothesize that context might lead to semantic control effects in early processing stages (e.g., Wirth et al., 2008) as well as task demands-specific effects on semantic retrieval and integration in later stages, which has to be tested in future research.

On the behavioral level, the interaction of concreteness and context was dissociated from the ERP pattern and showed a stronger processing facilitation by contextual information for abstract words. The behavioral pattern was in line with previous studies and has been interpreted to reflect the higher semantic diversity and thus more context-dependent processing of abstract words (Bechtold et al., 2021; Hoffman et al., 2015; Schwanenflugel & Shoben, 1983). Notably, we found the same interaction pattern in our smaller sample in this study ($n = 19$) as in two previous behavioral experiments with larger sample sizes ($n = 55$ and $n = 83$; see Bechtold et al., 2021), and in both studies, this behavioral effect persisted in additional covariate analyses, underlining its robustness. The finding that the amplitudes extracted from the detected N400 interaction cluster did not explain a significant amount of variance in SJT RTs further substantiates the assumption of a dissociation of behavioral from electrophysiological modulations. We can, however, not rule out that our design might have been too underpowered to detect a potential effect of the ERP amplitude predictor in interaction with the other fixed effect factors. For concreteness effects, a dissociation of RTs and N400 amplitudes has previously been highly dependent on the task, timing, and stimulus material (Barber et al., 2013; Grieder et al., 2012) so that we cannot assume our findings to generalize across other paradigms.

Conclusion

Our findings support separate as well as interacting effects of representational content (i.e., concreteness) and linguistic context as postulated, for example, in the controlled semantic cognition framework, with a dissociation of electrophysiological and behavioral results. By adopting the well-validated SJT cueing paradigm for this electrophysiological investigation, our findings complement previous insights from neuropsychology (Hoffman et al., 2010) and neuroimaging (Hoffman et al., 2015). Crucially, the temporally highly resolved ERP measures allowed us to disentangle effects of representational content and context on processing stages of (earlier) automatic and (later) controlled semantic retrieval and integration, reflected in the N400 and N700/LPC, respectively.

Reprint requests should be sent to Laura Bechtold, Department of Biological Psychology, Institute for Experimental Psychology, Heinrich Heine University Düsseldorf, Building 23.03, Universitätsstraße 1, Düsseldorf 40225 Germany, or via e-mail: Laura.Bechtold@hhu.de.

Data Availability Statement

Stimuli, data, ratings, and code and results are available here: <https://osf.io/nvufe/>.

Author Contributions

Laura Bechtold: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Visualization; Writing—Original draft. Christian Bellebaum: Conceptualization; Project administration; Resources; Validation; Writing—Review & editing. Marta Ghio: Conceptualization; Methodology; Validation; Writing—Review & editing.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were $M(\text{an})/M = .407$, $W(\text{oman})/M = .32$, $M/W = .115$, and $W/W = .159$, the comparable proportions for the articles that these authorship teams cited were $M/M = .549$, $W/M = .257$, $M/W = .109$, and $W/W = .085$ (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be as follows: $M/M = .448$; $W/M = .259$; $M/W = .121$; $W/W = .172$.

REFERENCES

- Adorni, R., & Proverbio, A. M. (2012). The neural manifestation of the word concreteness effect: An electrical neuroimaging study. *Neuropsychologia*, 50, 880–891. <https://doi.org/10.1016/j.neuropsychologia.2012.01.028>, PubMed: 22313624
- Almaghyuli, A., Thompson, H., Lambon Ralph, M. A., & Jefferies, E. (2012). Deficits of semantic control produce absent or reverse frequency effects in comprehension: Evidence from neuropsychology and dual task methodology. *Neuropsychologia*, 50, 1968–1979. <https://doi.org/10.1016/j.neuropsychologia.2012.04.022>, PubMed: 22579705
- Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers*, 31, 578–602. <https://doi.org/10.3758/BF03200738>, PubMed: 10633977
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3, 12–28. <https://doi.org/10.21500/20112084.807>
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). *The CELEX lexical database (CD-ROM)*. Philadelphia, PA: University of Pennsylvania.
- Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, 79, 33–41. <https://doi.org/10.1016/j.neuropsychologia.2015.10.020>, PubMed: 26476370
- Barber, H. A., Otten, L. J., Kousta, S. T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, 125, 47–53. <https://doi.org/10.1016/j.bandl.2013.01.005>, PubMed: 23454073
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bechtold, L., Bellebaum, C., Hoffman, P., & Ghio, M. (2021). Corroborating behavioral evidence for the interplay of representational richness and semantic control in semantic word processing. *Scientific Reports*, 11, 6184. <https://doi.org/10.1038/s41598-021-85711-7>, PubMed: 33731839
- Bechtold, L., Ghio, M., & Bellebaum, C. (2018). The effect of training-induced visual imageability on electrophysiological correlates of novel word processing. *Biomedicine*, 6, 75. <https://doi.org/10.3390/biomed6030075>, PubMed: 29966391
- Bechtold, L., Ghio, M., Lange, J., & Bellebaum, C. (2018). Event-related desynchronization of mu and beta oscillations during the processing of novel tool names. *Brain and Language*, 177–178, 44–55. <https://doi.org/10.1016/j.bandl.2018.01.004>, PubMed: 29421271
- Bernat, E., Nelson, L., Holroyd, C., Gehring, W., & Patrick, C. (2008). *Separating cognitive processes with principal components analysis of EEG time-frequency distributions* (Vol. 7074). SPIE.
- Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, 15, 527–536. <https://doi.org/10.1016/j.tics.2011.10.001>, PubMed: 22001867
- Bollen, K. A., & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods & Research*, 13, 510–542. <https://doi.org/10.1177/0049124185013004004>
- Bouaffre, S., & Fata-Ainseba, F. (2007). Hemispheric differences in the time-course of semantic priming processes: Evidence from event-related potentials (ERPs). *Brain and Cognition*, 63, 123–135. <https://doi.org/10.1016/j.bandc.2006.10.006>, PubMed: 17207563
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143. <https://doi.org/10.1016/j.brainres.2012.01.055>, PubMed: 22361114
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bolte, J., & Bohl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58, 412–424. <https://doi.org/10.1027/1618-3169/a000123>, PubMed: 21768069
- Chatrjian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology*, 25, 83–92. <https://doi.org/10.1080/00029238.1985.11080163>
- Chiou, R., Humphreys, G. F., Jung, J., & Lambon Ralph, M. A. (2018). Controlled semantic cognition relies upon dynamic and flexible interactions between the executive “semantic control” and hub-and-spoke “semantic representation” systems. *Cortex*, 103, 100–116. <https://doi.org/10.1016/j.cortex.2018.02.018>, PubMed: 29604611
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational

- frameworks. *Brain*, *128*, 615–627. <https://doi.org/10.1093/brain/awh349>, PubMed: 15548554
- Della Rosa, P. A., Catricala, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract-concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, *42*, 1042–1048. <https://doi.org/10.3758/BRM.42.4.1042>, PubMed: 21139171
- Grieder, M., Crinelli, R. M., Koenig, T., Wahlund, L. O., Dierks, T., & Wirth, M. (2012). Electrophysiological and behavioral correlates of stable automatic semantic retrieval in aging. *Neuropsychologia*, *50*, 160–171. <https://doi.org/10.1016/j.neuropsychologia.2011.11.014>, PubMed: 22133494
- Gullick, M. M., Mitra, P., & Coch, D. (2013). Imagining the truth and the moon: An electrophysiological study of abstract and concrete word processing. *Psychophysiology*, *50*, 431–440. <https://doi.org/10.1111/psyp.12033>, PubMed: 23445520
- Hauk, O. (2016). Only time will tell—Why temporal information is essential for our neuroscientific understanding of semantics. *Psychonomic Bulletin & Review*, *23*, 1072–1079. <https://doi.org/10.3758/s13423-015-0873-9>, PubMed: 27294424
- Hayasaka, S., & Nichols, T. E. (2004). Combining voxel intensity and cluster extent with permutation test framework. *Neuroimage*, *23*, 54–63. <https://doi.org/10.1016/j.neuroimage.2004.04.035>, PubMed: 15325352
- Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming—Differentiation by event-related potentials. *International Journal of Psychophysiology*, *44*, 197–218. [https://doi.org/10.1016/s0167-8760\(01\)00202-1](https://doi.org/10.1016/s0167-8760(01)00202-1), PubMed: 12031295
- Hoffman, P. (2016). The meaning of ‘life’ and other abstract words: Insights from neuropsychology. *Journal of Neuropsychology*, *10*, 317–343. <https://doi.org/10.1111/jnp.12065>, PubMed: 25708527
- Hoffman, P., Binney, R. J., & Lambon Ralph, M. A. (2015). Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge. *Cortex*, *63*, 250–266. <https://doi.org/10.1016/j.cortex.2014.09.001>, PubMed: 25303272
- Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2010). Ventrolateral prefrontal cortex plays an executive regulation role in comprehension of abstract words: Convergent neuropsychological and repetitive TMS evidence. *Journal of Neuroscience*, *30*, 15450–15456. <https://doi.org/10.1523/JNEUROSCI.3783-10.2010>, PubMed: 21084601
- Hoffman, P., Jones, R. W., & Lambon Ralph, M. A. (2013). Be concrete to be comprehended: Consistent imageability effects in semantic dementia for nouns, verbs, synonyms and associates. *Cortex*, *49*, 1206–1218. <https://doi.org/10.1016/j.cortex.2012.05.007>, PubMed: 22721956
- Hoffman, P., & Lambon Ralph, M. A. (2011). Reverse concreteness effects are not a typical feature of semantic dementia: Evidence for the hub-and-spoke model of conceptual representation. *Cerebral Cortex*, *21*, 2103–2112. <https://doi.org/10.1093/cercor/bhq288>, PubMed: 21285258
- Hoffman, P., McClelland, J. L., & Lambon Ralph, M. A. (2018). Concepts, control, and context: A connectionist account of normal and disordered semantic cognition. *Psychological Review*, *125*, 293–328. <https://doi.org/10.1037/rev0000094>, PubMed: 29733663
- Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 721–742. <https://doi.org/10.1037/0278-7393.25.3.721>, PubMed: 10368929
- Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology*, *23*, 492–499. <https://doi.org/10.1037/a0015452>, PubMed: 19586212
- Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., et al. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language*, *74*, 103–112. <https://doi.org/10.1006/brln.2000.2340>, PubMed: 10924219
- Kotchoubey, B., & El-Khoury, S. (2014). Event-related potentials indicate context effect in reading ambiguous words. *Brain and Cognition*, *92C*, 48–60. <https://doi.org/10.1016/j.bandc.2014.10.007>, PubMed: 25463139
- Kounios, J., & Holcomb, P. J. (1994). Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 804–823. <https://doi.org/10.1037/0278-7393.20.4.804>, PubMed: 8064248
- 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. <https://doi.org/10.1146/annurev.psych.093008.131123>, PubMed: 20809790
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*, 42–55. <https://doi.org/10.1038/nrn.2016.150>, PubMed: 27881854
- Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of Cognitive Neuroscience*, *25*, 484–502. https://doi.org/10.1162/jocn_a_00328, PubMed: 23163410
- Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920–933. <https://doi.org/10.1038/nrn2532>, PubMed: 19020511
- Lau, E. F., Weber, K., Gramfort, A., Hamalainen, M. S., & Kuperberg, G. R. (2016). Spatiotemporal signatures of lexical-semantic prediction. *Cerebral Cortex*, *26*, 1377–1387. <https://doi.org/10.1093/cercor/bhu219>, PubMed: 25316341
- Lee, C. L., & Federmeier, K. D. (2008). To watch, to see, and to differ: An event-related potential study of concreteness effects as a function of word class and lexical ambiguity. *Brain and Language*, *104*, 145–158. <https://doi.org/10.1016/j.bandl.2007.06.002>, PubMed: 17659768
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>, PubMed: 27620283
- Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, *15*, 97–103. <https://doi.org/10.1016/j.tics.2011.01.004>, PubMed: 21317022
- Malhi, S. K., & Buchanan, L. (2018). A test of the symbol interdependency hypothesis with both concrete and abstract stimuli. *PLoS One*, *13*, e0192719. <https://doi.org/10.1371/journal.pone.0192719>, PubMed: 29590121
- Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*, 549–565. <https://doi.org/10.1111/j.1469-8986.2011.01320.x>, PubMed: 22176204
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*, 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>, PubMed: 17517438

- Matsumoto, A., Iidaka, T., Haneda, K., Okada, T., & Sadato, N. (2005). Linking semantic priming effect in functional MRI and event-related potentials. *Neuroimage*, *24*, 624–634. <https://doi.org/10.1016/j.neuroimage.2004.09.008>, PubMed: 15652298
- Meade, G., & Coch, D. (2017). Word-pair priming with biased homonyms: N400 and LPC effects. *Journal of Neurolinguistics*, *41*, 24–37. <https://doi.org/10.1016/j.jneuroling.2016.09.002>
- Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage*, *67*, 111–118. <https://doi.org/10.1016/j.neuroimage.2012.10.027>, PubMed: 23123297
- Muraki, E. J., Sidhu, D. M., & Pexman, P. M. (2019). Mapping semantic space: Property norms and semantic richness. *Cognitive Processing*, *21*, 637–649. <https://doi.org/10.1007/s10339-019-00933-y>, PubMed: 31552508
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4), PubMed: 5146491
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. <https://doi.org/10.1155/2011/156869>, PubMed: 21253357
- Ortu, D., Allan, K., & Donaldson, D. I. (2013). Is the N400 effect a neurophysiological index of associative relationships? *Neuropsychologia*, *51*, 1742–1748. <https://doi.org/10.1016/j.neuropsychologia.2013.05.003>, PubMed: 23707682
- Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford, United Kingdom: Oxford University Press.
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*, 255–287. <https://doi.org/10.1037/h0084295>
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–987. <https://doi.org/10.1038/nrn2277>, PubMed: 18026167
- Pourtois, G., Delplanque, S., Michel, C., & Vuilleumier, P. (2008). Beyond conventional event-related brain potential (ERP): Exploring the time-course of visual emotion processing using topographic and principal component analyses. *Brain Topography*, *20*, 265–277. <https://doi.org/10.1007/s10548-008-0053-6>, PubMed: 18338243
- Pulvermuller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences*, *22*, 253–279; discussion 280–336. <https://doi.org/10.1017/S0140525X9900182X>, PubMed: 11301524
- Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia*, *41*, 550–564. [https://doi.org/10.1016/S0028-3932\(02\)00181-1](https://doi.org/10.1016/S0028-3932(02)00181-1), PubMed: 12638581
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*, e13335. <https://doi.org/10.1111/psyp.13335>, PubMed: 30657176
- Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*, 499–520. [https://doi.org/10.1016/0749-596X\(88\)90022-8](https://doi.org/10.1016/0749-596X(88)90022-8)
- Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 82–102. <https://doi.org/10.1037/0278-7393.9.1.82>
- Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, *24*, 114–126. <https://doi.org/10.2307/748013>
- Steinhauer, K., Royle, P., Drury, J. E., & Fromont, L. A. (2017). The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience Letters*, *651*, 192–197. <https://doi.org/10.1016/j.neulet.2017.05.007>, PubMed: 28483650
- Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: An ERP study. *Cognitive Brain Research*, *15*, 99–103. [https://doi.org/10.1016/s0926-6410\(02\)00219-7](https://doi.org/10.1016/s0926-6410(02)00219-7), PubMed: 12433385
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*, 279–293. <https://doi.org/10.1016/j.bandl.2005.11.003>, PubMed: 16343606
- Welcome, S. E., Paivio, A., McRae, K., & Joanisse, M. F. (2011). An electrophysiological study of task demands on concreteness effects: Evidence for dual coding theory. *Experimental Brain Research*, *212*, 347–358. <https://doi.org/10.1007/s00221-011-2734-8>, PubMed: 21656220
- West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*, 1024–1037. <https://doi.org/10.1162/08989290051137558>, PubMed: 11177422
- Wirth, M., Horn, H., Koenig, T., Razafimandimby, A., Stein, M., Mueller, T., et al. (2008). The early context effect reflects activity in the temporo-prefrontal semantic system: Evidence from electrical neuroimaging of abstract and concrete word reading. *Neuroimage*, *42*, 423–436. <https://doi.org/10.1016/j.neuroimage.2008.03.045>, PubMed: 18511302
- Yao, Z., & Wang, Z. (2014). Concreteness of positive word contributions to affective priming: An ERP study. *International Journal of Psychophysiology*, *93*, 275–282. <https://doi.org/10.1016/j.ijpsycho.2014.06.005>, PubMed: 24937349
- Yao, Z., Wu, J., Zhang, Y., & Wang, Z. (2017). Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behavior Research Methods*, *49*, 1374–1385. <https://doi.org/10.3758/s13428-016-0793-2>, PubMed: 27553483