The reliability of the N400 in single subjects: Implications for patients with disorders of consciousness

Damian Cruse [*,a], Steve Beukema[a], Srivas Chennu[b], Jeffrey G. Malins[a], Adrian M. Owen[a], Ken McRae[a,c]

[a]Brain and Mind Institute, University of Western Ontario, London, Ontario N6A 5B7, Canada
[b]Department of Clinical Neurosciences, University of Cambridge, Cambridge CB2 0QQ, UK
[c]Department of Psychology, University of Western Ontario, London, Ontario N6A 5B7, Canada

ARTICLE INFO

ABSTRACT

Functional neuroimaging assessments of residual cognitive capacities, including those that support language, can improve diagnostic and prognostic accuracy in patients with disorders of consciousness. Due to the portability and relative inexpensiveness of electroencephalography, the N400 event-related potential component has been proposed as a clinically valid means to identify preserved linguistic function in non-communicative patients. Across three experiments, we show that changes in both stimuli and task demands significantly influence the probability of detecting statistically significant N400 effects — that is, the difference in N400 amplitudes caused by the experimental manipulation. In terms of task demands, passively heard linguistic stimuli were significantly less likely to elicit N400 effects than task-relevant stimuli. Due to the inability of the majority of patients with disorders of consciousness to follow task commands, the insensitivity of passive listening would impede the identification of residual language abilities even when such abilities exist. In terms of stimuli, passively heard normatively associated word pairs produced the highest detection rate of N400 effects (50% of the participants), compared with semantically-similar word pairs (0%) and high-cloze sentences (17%). This result is consistent with a prediction error account of N400 magnitude, with highly predictable targets leading to smaller N400 waves, and therefore larger N400 effects. Overall, our data indicate that non-repeating normatively associated word pairs provide the highest probability of detecting single-subject N400s during passive listening, and may thereby provide a clinically viable means of assessing residual linguistic function. We also show that more liberal analyses may further increase the detection-rate, but at the potential cost of increased false alarms.

1. Introduction

In recent years it has become increasingly evident that functional neuroimaging assessments of residual cognitive capacities can improve both diagnostic and prognostic accuracy following a severe brain-injury (Cruse and Owen, 2010; Owen, 2013). To this end, it is often desirable to determine the extent to which the neural networks that support language may be preserved in a non-communicative patient (Duncan et al., 2009). Indeed, in patients with chronic disorders of consciousness – that is, the vegetative and minimally conscious states – the potential for recovery may be predicted from the relative preservation of neural responses to speech, as detected with functional magnetic resonance imaging (fMRI; Coleman et al., 2009, 2007). However, many patients are precluded from an fMRI assessment due to its cost and issues of scanner availability. Electroencephalography

(EEG) has the potential to reach a greater number of patients than fMRI because it is considerably less expensive and can be performed at the bedside. Moreover, EEG is known to provide an index of linguistic processing in healthy individuals, with one well-studied component being the N400 event-related potential (ERP) that is sensitive to semantic (meaning) processing.

Following the presentation of a variety of meaningful stimuli, a negative-going ERP deflection is observed typically over centroparietal scalp locations that peaks around 400 ms post-stimulus (Kutas and Federmeier, 2011). The amplitude of this so-called N400 is primarily sensitive to the context in which an item occurs. For example, when words are presented in pairs, the second word of the pair (the target) elicits a larger N400 when the words in the pair are unrelated (e.g., cat–*chair*) than when they are related (e.g., couch–*chair* or table–*chair*; Bentin et al., 1985). Similarly, in sentences, a word elicits a larger N400 when it is incongruent with the meaning of the sentence relative to when it is congruent. Congruency is often measured in terms of cloze probability, which is the proportion of participants who complete a fragment with a specific word. For example, following

"I take my coffee with cream and", *sugar* has a high cloze probability whereas *mud* is incongruent/unexpected. For consistency, here we refer to the difference in N400 amplitude that is produced by these types of semantic manipulations as the 'N400 effect' (Kutas and Federmeier, 2011). While the precise cognitive processes that are reflected by this component are still a matter of debate, it is clear from the extensive body of research conducted over the last 30 years that, at its simplest, the N400 reflects the "brain's normal response to words" (Kutas and Federmeier, 2000), and may therefore provide a clinically viable marker of residual linguistic processes (Connolly et al., 1999).

The small number of studies that have investigated the presence of N400 effects in patients with disorders of consciousness has typically relied on 'visual inspection' to draw their conclusions — i.e., subjective judgments of the presence or absence of an N400-like waveform in the averaged ERP (Connolly et al., 1999; Duncan et al., 2009; Schoenle and Witzke, 2004; Steppacher et al., 2013). According to this method, 12–15% of patients in the vegetative state and 21% of patients in the minimally conscious state have been considered to elicit N400 effects (Balconi et al., 2013; Schoenle and Witzke, 2004; Steppacher et al., 2013). However, while it is a standard clinical approach to ERP interpretation, it is unclear what criteria are employed when these visual judgments are performed. Indeed, visual inspection of an averaged ERP waveform does not provide any information about the cross-trial variance, or the contribution of outlying data-points to the average. Furthermore, the key aspect of analyses of N400 data concerns the difference between conditions, rather than a single assumed N400 wave in and of itself. Due to the noise inherent in the EEG signal, inferential statistics are vital in order to draw reliable conclusions regarding the presence of any experimental effect.

Furthermore, little is known about the sensitivity of the N400 effect. Indeed, the clinical utility of the N400 effect is predicated on the assumption that it reliably reflects the presence of normal language processing. Nevertheless, in none of the group studies described above was a healthy control group employed, precluding an estimation of the likelihood of detecting single-subject N400 effects in the only participant group for whom the presence of linguistic capacity can be verified. While the vast majority of the N400 literature involves group-level analyses – i.e., average effects observed across a group of participants – in a clinical setting, it is necessary to detect a reliable N400 in the ERPs of a single subject, a situation that suffers from a relatively lower signal to noise ratio than group-level analyses. To accurately interpret patient data, therefore, it is crucial to estimate the single-subject sensitivity of the N400 effect.

While N400 effects have been observed under reduced levels of attention in healthy populations, during passive listening, and during some stages of sleep, they are consistently smaller than the N400 effects elicited by attended stimuli that are task-relevant (Ibáñez et al., 2006, 2009, 2008). Nevertheless, in none of the studies described above were the patients instructed to engage in a task. Therefore, if the magnitudes of group-level N400 effects are reduced during passive listening – i.e., when no explicit task instructions are to be followed – then it is possible that the probability of detecting a single-subject N400 effect is equally diminished under these circumstances. To compound this problem, a clinical diagnosis of 'unconscious' follows from an inability to reliably follow task instructions (Kalmar and Giacino, 2005). As a result, many patients with disorders of consciousness are by definition unable to make task-relevant responses to stimuli (Giacino et al., 2002; Jennett and Plum, 1972). While it is known that a minority of patients with disorders of consciousness have been misdiagnosed and are able to covertly follow task instructions (Owen, 2013), the majority who do not possess this ability may thereby be precluded from exhibiting reliable N400 effects — even if they retain function in those brain networks which are responsible for its generation.

To determine the utility of the N400 approach to detecting preserved speech processing, we tested the sensitivity of three paradigms

that have consistently elicited N400 effects: (1) priming between semantically-similar word pairs in Experiment 1, (2) priming between normatively-associated word pairs in Experiment 2, and (3) high-cloze sentence completions versus anomalous completions in Experiment 3. First, in Experiment 1, we investigated the effects of task-relevance on the probabilities of detecting N400 effects with inferential statistics at a single-subject level. Three groups of healthy participants completed a word-pair semantic-priming paradigm in which they were instructed to either: (1) indicate the semantic relatedness of each word-pair with a button press (Overt condition), (2) make a mental judgment of the semantic relatedness of the word pair without a behavioral response (Covert condition), or (3) passively attend to the stimuli (Passive condition). We investigated whether the probability of detecting a significant N400 effect at a single-subject level would decrease when the participants were not engaged in an active task.

2. Experiment 1: semantically similar word-pairs

2.1. Methods

2.1.1. Participants

Fifty-one participants were recruited from the University of Western Ontario Psychology Participant, Pool, and were compensated with course credit. Data from two participants were excluded due to excessive movement artifacts in their EEG recordings (>50% bad trials), and data from one participant was excluded due to an equipment fault. Of the remaining 48 participants (mean age: 21.50, SD: 4.95), 24 were male. The first 12 participants that were recruited were assigned to the Validation task. Each of the subsequent 36 participants was randomly assigned to one of the three experimental conditions (Overt, Covert, or Passive). Age did not significantly differ across groups ($F(3,44) = 1.95$, $p = .14$). All participants were right-handed, native English speakers. The Psychology Research Ethics Board of the University of Western Ontario, Canada, provided ethical approval for this study.

2.1.2. Stimuli

The goal of stimulus construction was to create as large a set as possible of word/concept pairs that were as strongly semantically similar as possible. Four hundred concrete nouns were chosen from the feature production norms described in Cree and McRae (2003) and McRae et al. (2005), of which 120 were creatures (types of animals), 40 were fruits or vegetables, and 240 were various types of nonliving things. Concepts from McRae and colleagues' feature production norms were used because of the huge number of conceptual and lexical variables that are part of the database, enabling strict stimulus control. From these 400 stimuli, 100 semantically similar word pairs were generated based on items that have produced semantic priming effects in previous behavioral studies (McRae et al., 1997; McRae and Boisvert, 1998). The 100 semantically related pairs were composed of 30 creature pairs (e.g., moth–butterfly), 10 fruit/vegetable pairs (e.g., lemon–lime), and 60 non-living object pairs (e.g., coat–jacket). The first and second words in each pair shall be referred to as the *prime* and *target*, respectively. Related primes and targets were chosen on the basis of semantic similarity. These pairs shared numerous semantic features according to McRae et al.'s (2005) norms, and/or had been rated as highly semantically similar in previous studies (McRae and Boisvert, 1998). Because the goal was to construct items that would show priming effects, we were not concerned with whether or not related primes and targets also were associated according to word association norms (Nelson et al., 1998). Many of the 100 related pairs are associated according to those norms (e.g., lemon–lime, bull–cow, lamb–sheep). Finally, note that 100 prime–target pairs is a substantially larger stimulus set than is used in priming experiments with

healthy adults. All related and unrelated pairs are presented in Appendix A of the Supplementary materials.

In a typical priming study, each participant is presented with half of the targets preceded by related primes, and half preceded by unrelated primes. Thus, there are two stimulus lists, and each participant sees only one of them so that they are presented with each word only once. Because analyses are based on single participants in this study, we chose different words to be the unrelated primes and targets so that every participant was presented with every prime–target pair, and no word was presented more than once. While repeating words is common in single-subject N400 studies (Kotchoubey et al., 2005; Kotchoubey, 2005; Rama et al., 2010), doing so in word-pair studies can reduce the magnitude of overall N400 effects (see Appendix D). Therefore, from the remaining 200 stimuli, 100 words were chosen to be unrelated targets, and they were matched with the related targets on the criteria listed in Supplementary materials Table F1. The remaining 100 words were matched with the related primes, and were then used as primes for the semantically unrelated pairs. Descriptive statistics for all stimuli can be found in Supplementary materials Table F1. There were no significant differences between related and unrelated targets (two-tailed t-tests, all $p > .12$), or between related and unrelated primes (two-tailed t-tests, all $p > .09$) on any of these variables. Signal-correlated noise stimuli were generated from all primes according to Schroeder (1968).

In total, the materials consisted of 100 related word-pairs, 100 unrelated word-pairs, and 200 signal-correlated noise stimuli. Thus, the proportion of related prime–target pairs was 0.5. Stimuli were digitally recorded by a male, native Canadian-English speaker, and their amplitudes were normalized (mean stimulus length: 613 ms, SD: 113 ms, range: 355–978 ms). There were no significant differences in the durations of the stimuli between related and unrelated targets ($t(198) = 0.51$, $p = .613$, two-tailed) or between related and unrelated primes ($t(198) = 1.16$, $p = .246$, two-tailed).

2.1.3. Stimulus validation procedure (validation condition)

Due to the single-subject nature of the analyses, it was crucial to ensure that there were no stimulus-driven differences in the ERPs elicited by the target stimuli that could confound N400 effects. Therefore, the Validation group of participants was presented with each word from the experimental task in isolation from its paired word. Specifically, each trial began with the presentation of a signal correlated noise stimulus followed 1100 ms later by the onset of the word. A random period of 1100–2100 ms (uniform sampling on every trial) separated the onset of the word and the onset of the next trial. Care was taken to ensure that no adjacent trials contained words that were semantically related.

2.1.4. Experimental task procedure (Overt, Covert, and Passive conditions)

To signal the onset of a trial and to encourage the pairing of words, a signal-correlated noise stimulus was presented 2100 ms prior to the onset of the prime word. The target word was presented 1100 ms later and was followed 4000 ms later by the onset of the next trial. Trial order was randomized for each participant. In the Overt and Covert conditions, participants were instructed to make a binary judgment of the semantic relatedness of each target to its prime (related versus unrelated). In the Overt condition, participants signaled this judgment with a button-box under their right-hand. All button presses were made with the index and middle fingers, counterbalanced across participants so that exactly half of the Overt group signaled 'related' with their index finger, and the other half with their middle finger. In the Covert condition, participants were instructed to mentally 'say' their judgment silently to themselves following each target. In the Passive condition, participants were simply instructed to pay attention to the words. All participants completed the task with their eyes closed to

reduce ocular artifacts in the EEG recording. Brief breaks were provided upon completion of every 50 trials.

2.1.5. EEG recording and pre-processing procedures

Data were acquired from a 129-channel Electrical Geodesics Inc. (EGI, OR, USA) EEG cap with a sampling rate of 250 Hz referenced to the vertex. Impedances were kept below 50 kΩ. Data from 91 channels over the scalp surface were retained for additional analysis, after excluding those on the neck, cheeks, and forehead. These data were subsequently filtered offline between 0.5 and 20 Hz and segmented into 896 ms epochs time-locked to the onset of each stimulus (100 ms pre-stimulus plus 796 ms post-stimulus). Epochs were baseline corrected, and trials containing excessive artifacts were visually identified and excluded from analyses. Across participants, a median of 86 trials (range 62–97) contributed to the related target category, and 85.5 (range 56–97) to the unrelated target category. Bad channels were visually identified, removed, and interpolated using EEGLAB. The median number of channels interpolated was 2 (range 0–26). When ocular artifacts remained in the data after these steps, they were removed with the Independent Component Analysis (ICA) procedure of EEGLAB (Delorme and Makeig, 2004). Specifically, after ICA decomposition of the EEG data (EEGLAB extended 'runica' algorithm), those components with scalp distributions, time-courses, and spectral contents indicative of eye-blinks or eye-movements were subtracted from the EEG data, and each epoch was again baseline corrected. After this step, any remaining trials containing artifacts were visually identified and removed.

A two-way ANOVA was conducted on the log proportions of trials marked as bad, with factors of condition (Covert, Overt, Passive, Validation) and target type (unrelated, related). This revealed no significant effects or interactions (all $p > .56$), indicating that there were no significant differences in the numbers of trials contributing to the analyses between conditions or target types. A one-way ANOVA conducted on the log proportion of channels interpolated, with condition as the factor (Covert, Overt, Passive, Validation), also revealed no significant differences ($F(3,44) = 1.04$, $p = .38$). All pre-processing steps were performed using MATLAB and EEGLAB (Delorme and Makeig, 2004).

2.1.6. ERP analyses

Data were analyzed using the cluster-mass procedure of FieldTrip, described fully in Maris and Oostenveld (2007). Briefly, this procedure compares spatiotemporal data-points across conditions using t-tests. For the single-subject analyses, one-tailed independent samples t-tests were performed at every spatiotemporal point within each trial across conditions. For the within-group analyses, the single-subject ERP averages elicited by each stimulus type (related and unrelated targets) were compared using one-tailed dependent samples t-tests. For the between-group analyses, the differences between the single-subject average ERPs elicited by the unrelated and related target conditions were compared using one-tailed independent samples t-tests.

Although the t-test step is parametric, FieldTrip employs a secondary nonparametric clustering method to address the multiple comparisons problem. Specifically, t-values of adjacent spatiotemporal points whose p-values were $<.05$ were clustered together by summing their t-values, and the largest such cluster was retained. A minimum of two neighboring electrodes had to pass this threshold to form a cluster, with neighborhood defined as other electrodes within a 4 cm radius. This entire procedure, that is, calculation of t-values at each spatiotemporal point followed by clustering of adjacent t-values, was then repeated 1000 times, with recombination and randomized resampling of the ERP data before each repetition. This Monte Carlo method generated a nonparametric estimate of the p-value representing the statistical significance of the originally identified cluster. This approach provides increased power relative to other corrections

for multiple comparisons such as Bonferroni correction and False-Discovery Rate.

When statistically significant differences were observed across groups of participants, follow-up analyses were performed to determine whether this difference was quantitative or qualitative — i.e., the result of the same neural processes engaged to different degrees, or the result of two distinct neural processes. To accomplish this, single-subject average ERP amplitudes were averaged across the time-window of interest and max-min normalized to remove differences in the amplitudes of effects across conditions, leaving only differences in spatial distribution. The group data were then subjected to the same clustering analysis as described above, with the exception that this analysis reveals spatial clusters rather than spatiotemporal clusters as there is only one averaged time-point under analysis. Significantly different spatial distributions in this procedure are considered to reflect the activity of neural generators that do not entirely overlap across groups (McCarthy and Wood, 1985; Wilding, 2006).

For the three experimental conditions in Experiment 1 (Overt, Covert, Passive), all analyses were one-tailed to increase power, and included only data from 200 ms post-stimulus until the end of the epoch, as this was the time period in which the N400 effect is known to be maximal (Kutas and Federmeier, 2011). For the Validation group (in Experiments 1 and 2), all analyses were two-tailed and included all post-stimulus ERP data to detect any stimulus-driven differences across conditions.

2.2. Experiment 1: results

2.2.1. Behavioral analyses

In the Overt group, participants judged the semantic relatedness of the word pairs with a mean accuracy of 97% (SD = 2%). Decision latencies were significantly shorter for related targets (M = 877 ms, SD = 176 ms) than for unrelated targets (M = 956 ms, SD = 140 ms; $t(11) = 3.62$, $p = .004$, two-tailed).

2.2.2. Stimulus validation

In the Validation condition, there were no significant differences in the ERPs elicited by those items that formed the unrelated and related targets in the experimental conditions, or the unrelated and related primes (all $p > .025$). This was true for both group-level (primes' minimum cluster $p = .71$; targets' minimum cluster $p = .09$) and single-subject analyses (primes' median minimum cluster $p = .40$, range = .11–.71; targets' median minimum cluster $p = .47$, range = .03–.95). These results confirm that any ERP differences found between unrelated and related targets in the experimental conditions are due to semantic priming and not other aspects of the stimuli.

2.2.3. Group ERP analyses

All three experimental conditions showed significant N400 effects, defined as greater negativity for unrelated as compared to related targets. For the Overt condition, this effect started at 304 ms and lasted until the end of the epoch (796 ms post-stimulus). The same effect in the Covert condition started at 392 ms and lasted until 796 ms, whereas in the Passive condition, the effect started at 544 ms and lasted until 704 ms (see Fig. 1).

The N400 effect was significantly larger in the Overt condition than in both the Covert (384 versus 688 ms, centro-frontal scalp, $p = .038$, one-tailed) and Passive conditions (308–796 ms, central scalp, $p = .005$, one-tailed). There were no significant differences in the magnitudes of the effects between the Covert and Passive conditions (minimum cluster $p = .20$, one-tailed).

2.2.4. Single-subject analyses

Significant N400 effects were evident in 75% (9/12) of participants in the Overt group, 58% (7/12) of the Covert group, and 0% (0/12) of

the Passive group (Fig. 3). These proportions were significantly different across groups (Fisher's exact test, $p < .001$, one-tailed), with the most striking result being that none of the passive subjects showed significant priming effects. As shown in Fig. 2, relative to the Overt group, the significant N400 effects in the Covert group were of significantly shorter duration ($t(14) = 2.45$, $p = .014$, one-tailed) and lower statistical significance ($t(14) = 1.87$, $p = .041$, one-tailed). All single-subject N400 effects are presented in Supplementary Fig. 1.

2.3. Experiment 1: discussion

In accordance with previous studies, reliable group-level N400 effects were observed at all levels of task demand (Kutas and Federmeier, 2011). However, Experiment 1 shows that on a single-subject level, the likelihood of detecting a statistically significant N400 effect is highly dependent on task demands, to the extent that passive listening was not sufficient to observe significant N400 effects in any of our demonstrably healthy participants with reportedly normal language comprehension skills. These data therefore indicate that the N400 effect elicited in a semantic-similarity word-pair priming task does not provide a sensitive marker of preserved linguistic function in those non-communicative patients who lack the other higher-order cognitive functions necessary to follow task instructions.

Significant N400 effects were evident at the group level across all conditions, with the largest effects elicited when participants overtly indicated whether or not the prime and target were semantically related (see Fig. 1). Furthermore, relative to this group, the Covert response condition elicited both group-level and single-subject effects that were relatively briefer and of lower statistical significance (see Figs. 1 and 2), thus emphasizing the contribution of Overt task demands to the magnitude of the N400 effect (Bentin et al., 1993). The finding that passive listening was sufficient to produce a group-level N400 effect is also consistent with previous demonstrations of the relative automaticity of the N400 under certain circumstances (Kiefer, 2002; Vogel et al., 1998). Indeed, N400 effects have been reported in some stages of sleep (Ibáñez et al., 2009), suggesting that conscious awareness is not a prerequisite to the generation of an N400 effect. However, it is evident that the magnitude of the N400 effect is considerably reduced in the absence of explicit task demands, and is thereby more difficult to detect on a single-subject basis. The hypothetical presence of a statistically significant N400 effect in a non-communicative patient, therefore, would be indicative of the relative preservation of aspects of the neural networks that support language, but would not necessarily be indicative of conscious awareness. However, according to the current data, there is only a negligible probability of detecting an N400 effect in a patient who is not able to direct their attention to the stimuli in service of task demands. Because the majority of patients with disorders of consciousness are unable to follow task instructions, they are also unlikely to exhibit N400-based evidence of residual linguistic and semantic function with a semantic-similarity task, even if those neural networks that support the N400 are preserved. Indeed, if a patient is able to behaviorally follow verbal commands, there is no longer any question of the extent to which they understand speech, thereby rendering the presence of an N400 effect inconsequential.

Experiment 1 used strongly semantically similar word pairs, and semantically similar concepts have consistently produced priming effects. However, manipulating semantic similarity is not the only method that has been shown to produce group-level N400 effects. Indeed, there is some evidence to suggest that word-pairs generated from normative associations may lead to larger effects (Ortu et al., 2013; Rhodes and Donaldson, 2008). In these cases, the target of each related pair is selected from among the words most commonly produced as word associates to a prime (stimulus). Some prime–target pairs will be semantically similar as well, but relations among concepts also drive word association responses (McRae et al., 2012). There
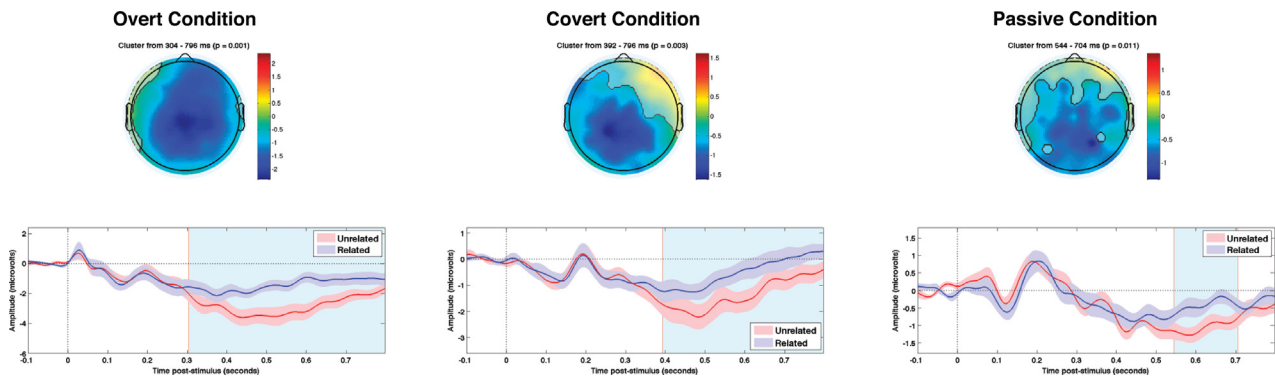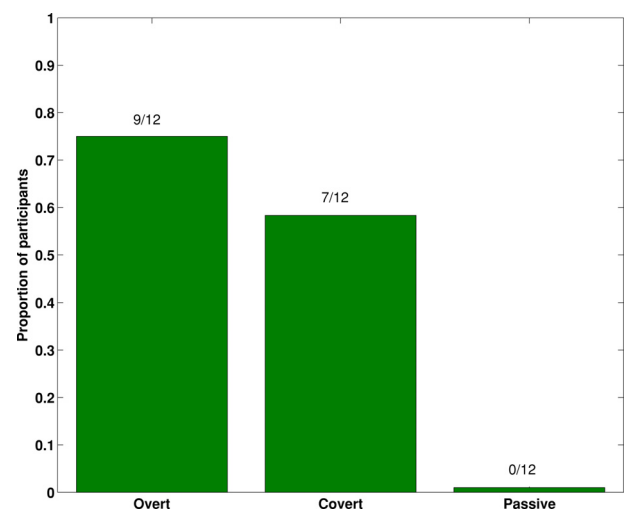
Fig. 1. Semantically similar word-pairs. Grand average N400 effects (unrelated targets < related targets) in each condition from Experiment 1. Upper panels highlight the spatial extent of the significant spatiotemporal cluster (i.e., all electrodes that contributed to the cluster). Color bars show average amplitude differences between unrelated and related targets across the temporal extent of the spatiotemporal cluster. Lower panels show the means of the ERPs within the respective spatial clusters (±1 standard error). The temporal boundaries of each cluster are shaded in light blue.
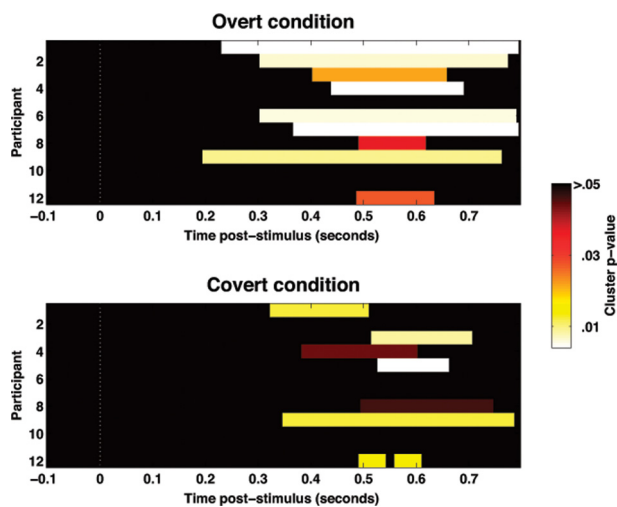
Fig. 2. Time-courses of significant single-subject N400 effects in Experiment 1. Each row of each stacked color-plot shows data from one participant. For participants eliciting significant N400 effects, the temporal extents of the significant spatiotemporal clusters are highlighted. As can be seen, significant N400 effects were on average of longer duration and greater statistical significance in the Overt group than in the Covert group. There were no significant single-subject N400 effects in the passive condition.

Fig. 3. The proportions of participants returning significant N400 effects across conditions in Experiment 1.

is evidence from computational modeling that the magnitude of the N400 reflects the extent to which prediction error occurs (Rabovsky and McRae, 2014). Under this assumption, when a target is highly likely to be produced in response to a prime – i.e., it is strongly normatively associated – it may elicit a smaller N400 waveform due to lower prediction error. Such a reduction in the magnitude of the N400 to related targets may thereby lead to a larger difference relative to the N400 elicited by unrelated targets, and thereby result in a larger overall N400 effect.

Therefore, to test the hypothesis that this type of stimuli will elicit more reliable N400 effects, in Experiment 2, a new group of 12 participants passively listened to word pairs that were taken from word association norms (Nelson et al., 1998). As in Experiment 1, to avoid order effects (see Supplementary materials: Appendix D) the stimuli were designed so that targets were never repeated. Therefore, a separate group of 12 participants also completed the stimulus validation procedure as in Experiment 1 to verify that the observed N400 effects were a reflection of priming, and not other aspects of the words themselves.

3. Experiment 2: normatively associated word-pairs

3.1. Methods

3.1.1. Participants

Twelve participants completed the validation procedure (mean age = 18.6 years, SD = 0.8 years; 6 males). Thirteen participants completed the experimental task because data from one participant were excluded due to excessive artifact (>50% bad trials). The remaining twelve participants (mean age = 18.3 years, SD = 0.5 years; 7 males) took part in the experimental task. All participants were recruited from the University of Western Ontario Psychology Participant, Pool, and were compensated with course credit. All participants were right-handed, native English speakers. The Psychology Research Ethics Board of Western University (Ontario, Canada) provided ethical approval for this study.

3.1.2. Stimuli

The goal of stimulus construction was to create a large set of the most strongly associated pairs that exist in Nelson et al.'s (1998) norms without duplicating primes or targets. One hundred of the most strongly associated related pairs from Nelson et al.'s norms were selected (e.g., left–right, keg–beer, oak–tree). The mean forward association was 0.81 (SD = 0.05), so that, on average, 81% of their participants produced the target from the prime when asked to "write the

first word that comes to mind that is meaningfully related or strongly associated to the presented word". Therefore, the items were substantially more strongly forward-associated than is the case in the vast majority of priming experiments. Furthermore, choosing word pairs based on forward association means that many of the pairs of words often directly co-occur in speech or text in the order used in Experiment 2, such as oak–tree and hound–dog. A further 100 word pairs were chosen from Nelson et al., but they were recombined to form unrelated pairs. In constructing the unrelated pairs, care was taken to ensure that there was no phonological, semantic, or associative overlap between the unrelated targets and any word that was associated to the prime in Nelson et al. (1998). Targets and primes were matched across the statistics detailed in Supplementary materials Table F2. Stimuli were digitally recorded by a male, native Canadian-English speaker, and their amplitudes were normalized (mean spoken word length = 638 ms, SD = 138 ms, range = 309–980 ms). There were no significant differences between the related and unrelated pairs in the spoken length of targets ($t(198) = 1.28$, $p = .203$) or primes ($t(198) = 0.67$, $p = .505$) between the related and unrelated pairs.

3.1.3. Experimental task procedure

The procedure was identical to the Passive condition of Experiment 1.

3.1.4. EEG recording and pre-processing procedures

All EEG recording and pre-processing procedures were identical to those used in Experiment 1.

Across participants, a median of 88 trials (range = 72–97) contributed to the related target condition, and 87 (range = 68–100) to the unrelated target condition. The median number of channels interpolated was 1 (range = 1–13).

3.1.5. ERP analyses

The analyses were identical to those used in Experiment 1, except that the only condition was passive listening, given that it is the primary condition of interest. Identical analyses were used for the Validation condition as well.

3.2. Experiment 2: results

3.2.1. Stimulus validation

In the Validation condition, there were no significant differences in the ERPs elicited by those items that formed the unrelated and related targets in the experimental condition (i.e., all $p > .025$). This was true for both group-level (primes' minimum cluster $p = .07$; targets' minimum cluster $p = .14$) and single-subject analyses (primes' median minimum cluster $p = .49$, range = .11–.77; targets' median minimum cluster $p = .40$, range = .03–.91). These results confirm that any ERP differences found between unrelated and related targets in the experimental condition are due to priming and not other features of the stimuli.

3.2.2. Group ERP analyses

The ERPs elicited by unrelated targets were significantly more negative-going than those elicited by related targets from 304 to 796 ms over centroparietal scalp ($p = .002$; see Fig. 4).

3.2.3. Single-subject analyses

Six out of twelve participants (50%) showed significant N400 effects. All single-subject N400 effects are presented in Supplementary Fig. 2.

3.2.4. Comparison with Experiment 1

At centroparietal electrodes from 348 to 540 ms, the group-level N400 effect in the current experiment was significantly larger than the N400 effect observed in Experiment 1 ($p = .008$). This effect likely reflects the later onset of the N400 effect in Experiment 1 (544 ms) relative to Experiment 2 (304 ms).

To determine whether the significant difference between experiments reflected differences in magnitude of the same effect, or qualitatively different processing occurring within the same time-window, a spatial cluster analysis was performed on the max–min normalized ERP data (McCarthy and Wood, 1985) within this time-window (348–540 ms). No significant clusters were found (minimum cluster $p = .13$) indicating that the difference in ERPs in this time-window reflects a difference in magnitude rather than in the neural processes engaged.

A comparison of the scalp distributions of the two significant N400 effects themselves (304–796 ms for Experiment 2 versus 544–704 ms for Experiment 1) revealed no significant effects either (no clusters). Together these results indicate that the stimuli in both Experiments 1 and 2 elicited the same ERP-detected processes, but with an earlier onset in Experiment 2.

3.3. Experiment 2: discussion

A reliable group-level N400 effect was again observed over centroparietal scalp electrodes, this time using word-pair stimuli generated from normative associations (Fig. 4). These stimuli, however, were considerably more successful at eliciting single-subject N400 effects than the semantically-similar pairs of Experiment 1, with 50% of participants returning significant effects during passive listening (0% in Experiment 1). The increase in sensitivity associated with these stimuli is consistent with recent evidence that the N400 waveform reflects prediction error (Rabovsky and McRae, 2014). As the related targets in this experiment had an extremely high likelihood of being produced in response to the prime during free association, the prediction error would have been minimal. When contrasted with the N400 to unrelated targets, therefore, the magnitude of the N400 effect would be increased.

As these are two separate groups of participants, it is not appropriate to directly compare the magnitudes of the ERPs to related targets between Experiments 1 and 2. However, the N400 effects were significantly different between these two experiments. Specifically, the group-level N400 effect started 240 ms earlier in Experiment 2, and was significantly greater in magnitude than the same effect in Experiment 1. This difference may reflect a greater fluency of processing highly predictable targets relative to targets that are semantically related but less predictable. While the effects onset with different latencies across the two experiments, there was no evidence that the two N400 effects were generated by non-overlapping regions of cortex. It therefore appears that both semantic similarity and normative association, as probed in Experiments 1 and 2 respectively, engage the same N400 processes. However, these processes are more rapidly engaged when targets are highly predictable. Also note that any comparison between semantically-related pairs and associatively-related pairs is not an absolute one. That is, we did not remove normatively-associated pairs from the semantically-similar items used in Experiment 1, and we did not remove semantically-similar pairs from the normatively-associated items used in Experiment 2.

A third paradigm that also draws on the predictability of target words to elicit N400 effects is one that features high-cloze words in sentences. In this paradigm, comparisons are made between words that are highly predictable – i.e. they have a high cloze probability – and those that are incongruent with the sentence context. As with the normatively-associated word-pairs, it is possible that the greater level of target predictability instantiated by the sentence contexts will lead to more reliable single-subject N400 effects. Indeed, there is some evidence that the N400 effect elicited by highly predictable versus anomalous words is larger on a group level than that elicited in a word-pair task (Kutas, 1993). In Experiment 3, we therefore investigated whether measuring N400s for words that are highly predictable
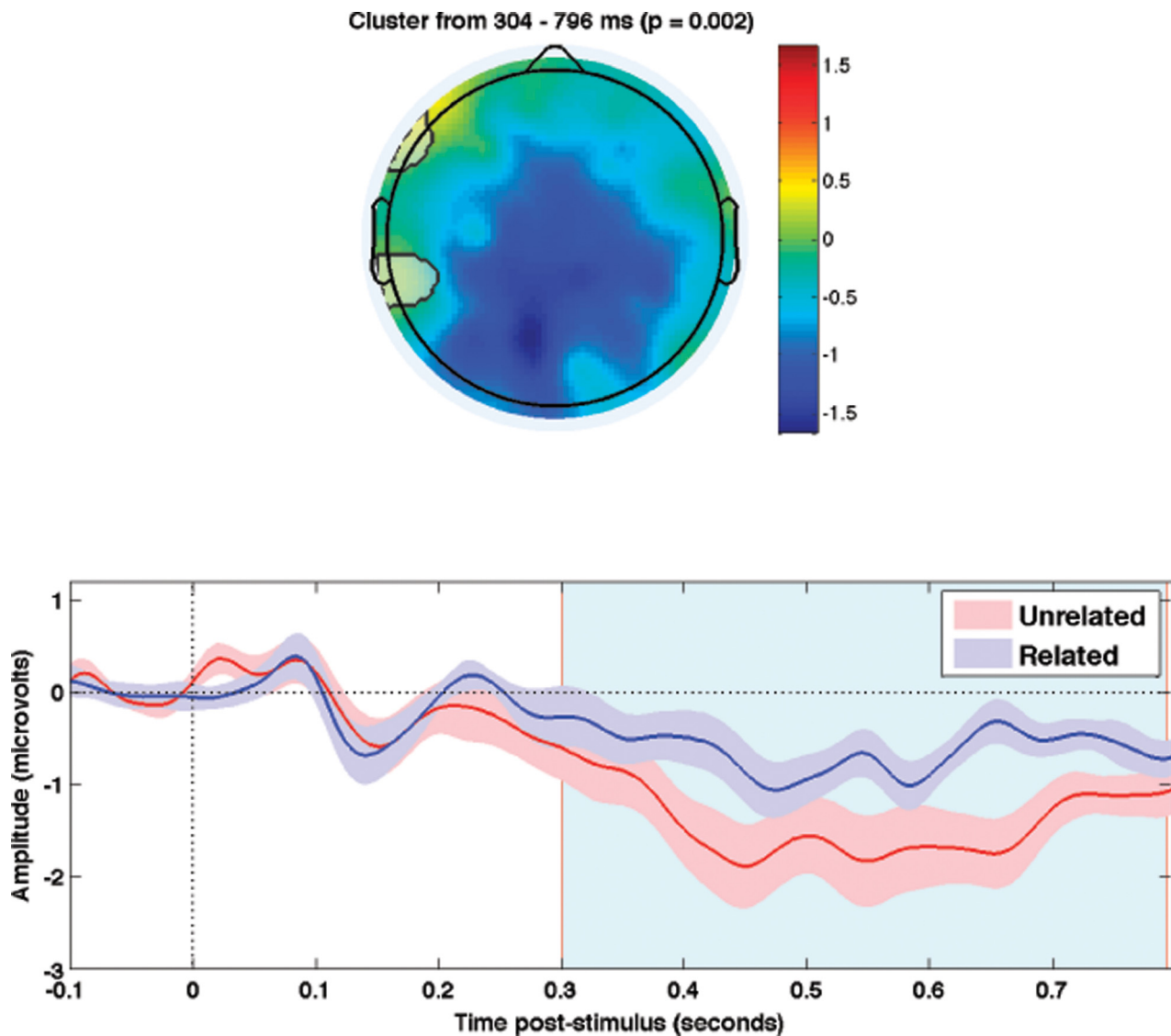
Fig. 4. Normatively associated word-pairs. Grand average N400 effects (unrelated targets < related targets) in Experiment 2. Upper panels highlight the spatial extent of the significant spatiotemporal cluster (i.e., all electrodes that contributed to the cluster). Color bars show average amplitude differences between unrelated and related targets across the temporal extent of the spatiotemporal cluster. Lower panels show the means of the ERPs within the respective spatial clusters (± 1 standard error). The temporal boundaries of each cluster are shaded in light blue.

versus anomalous in the local sentence context would further improve the sensitivity of detecting single-subject N400 effects during passive listening.

4. Experiment 3: high-cloze sentences

4.1. Material and methods

4.1.1. Participants

Twelve participants (mean age: 20.08, SD: 2.84; 6 males) were recruited from the University of Western Ontario Psychology Participant, Pool, and were compensated with course credit. All participants were right-handed, native English speakers. The Psychology Research Ethics Board of Western University (Ontario, Canada) provided ethical approval for this study.

4.1.2. Stimuli

The goal of stimulus construction was to create a large set of items in which the sentence produced a context in which a specific word was extremely highly expected. Stimuli were taken from the cloze

norms of Block and Baldwin (2010). One hundred sentences with high cloze targets were selected to form the predictable condition (mean cloze = 0.92, SD = 0.04). Therefore, the predictable words on which N400s were measured were extremely predictable in that, on average, 92% of the participants in Block and Baldwin produced that specific word as a continuation of the sentence. An additional one hundred sentences were selected from their norms to form the frames for the anomalous condition. These frames were paired with the sentence endings from the related condition to form 100 anomalous sentence–target pairs. Using high-cloze sentence frames for the anomalous target items is advantageous because specific continuations are highly expected, and therefore it is relatively straightforward to construct highly anomalous target continuations. Care was taken to ensure there was no phonological, semantic, or associative overlap between the anomalous targets and any high-cloze targets that were produced the anomalous sentence frames. For example, "class" was preceded by "She graduated at the top of her" in the predictable condition, whereas it was preceded by "Diane sank slowly into the hot" in the anomalous condition. There were a mean of 8.06 words in the predictable sentences (SD: 1.69) and 8.26 words in the

anomalous sentences (SD: 1.57). This difference was not significant ($t(198) = 0.867$, $p = .387$, two-tailed).

While repetition-priming effects can be a problem in word-pair studies (see Supplementary materials: Supplementary experiment), it is unlikely that they are detrimental in a sentence task due to the larger number of words involved. That is, rather than experiencing a set of word pairs, each participant hears many words as part of the sentence stimuli, with many of them repeated because sentences naturally overlap in their content overall. However, to investigate potential order effects, each subject heard half of the target words in a predictable context first (in the first half of the experiment), and the other half in an anomalous context first. Stimuli were digitally recorded by a male, native Canadian-English speaker, and their amplitudes were normalized. All sentences were spoken naturally, and the time-point of onset of the target word was identified in the digital recording and used to mark the stimulus onset in the EEG recording.

4.1.3. Experimental task procedure

The procedure was identical to the Passive condition of Experiment 1, except that participants heard sentences rather than word pairs.

4.1.4. EEG recording and pre-processing procedures

All EEG recording and pre-processing procedures were identical to those used in Experiment 1. Across participants, a median of 69.5 trials (range = 53–89) contributed to the predictable target condition, and 71 (range = 43–93) to the anomalous target condition. The median number of channels interpolated was 2.5 (range = 0–5).

4.1.5. ERP analyses

The analyses were identical to those used in Experiment 1, except that the only condition was passive listening, given that it is the primary condition of interest. Furthermore, there was no validation condition because the same targets were used in the predictable and anomalous conditions. N400 effects were analyzed between 200 and 800 ms post-stimulus in a one-tailed test. Because each target appeared as both a predictable and anomalous ending across the experiment, two subtraction ERPs (predictable targets minus anomalous targets) were compared between those stimuli for which the target was presented first in a predictable context, and those stimuli for which the target was presented first in an anomalous context. These effects were analyzed from stimulus onset until 800 ms post-stimulus in a two-tailed test.

4.2. Experiment 3: results

4.2.1. Order effects

At the group level, the order in which targets were heard – i.e. predictable first, or anomalous first – did not significantly affect the magnitudes of the differences between related and unrelated targets (lowest cluster $p = .198$). Therefore, repeating targets in sentence paradigms does not appear to induce order effects, and justifies the analysis of the overall N400 effect across the experiment.

4.2.2. Group ERP analyses

A significant N400 effect was observed at the group level from 360 to 756 ms post-stimulus at centroparietal sites ($p = .001$; see Fig. 5).

4.2.3. Single-subject analyses

Significant N400 effects were evident in only two of the twelve participants (17%). All single-subject N400 effects are presented in Supplementary Fig. 3.

4.2.4. Comparisons with Experiments 1 and 2

The N400 effect in Experiment 3 was significantly larger than the N400 effect found in Experiment 1 from 360 to 540 ms ($p = .006$) at

centroparietal electrodes. As in Experiment 2, this result appears to reflect the differences in N400 effect onset between the two experiments (360 versus 544 ms).

Despite the greater probability of detecting single-subject N400 effects in Experiment 2 relative to Experiment 3 (50% versus 17%), there were no significant differences between the N400 effects across these two experiments (lowest cluster $p > .21$). An average of ~16 fewer clean trials contributed to each of the conditions of interest in Experiment 3 than in Experiment 2 (predictable: $t(22) = 3.62$, $p = .002$; anomalous: $t(22) = 6.30$, $p = .002$). This is likely due to the fact that target words were spoken within a natural sentence and as such did not have a guaranteed silent baseline or a clear boundary of word onset due to co-articulation.

A Fisher's exact test confirmed that the single-subject hit-rates differed across the three experiments ($p = .014$). Subsequent pairwise Fisher's exact tests indicated that this effect was driven by the significantly higher hit-rate in Experiment 2 (normatively associated word-pairs) than in Experiment 1 ($p = 0.013$).

### 4.3. Experiment 3: discussion

Consistent with numerous previous studies, a significant N400 effect was observed between high-cloze and zero-cloze sentence endings (see Fig. 5). At the single-subject level, 17% (2/12) of participants elicited significant N400 effects with this task, indicating a lower sensitivity when compared with the normatively-associated word-pair task of Experiment 2 (50%), and a small increase relative to the semantic-relatedness word-pairs of Experiment 1 (0%). The time-course of the N400 effect was comparable with that of the normatively-associated word-pairs (Experiment 2), and onset earlier than that elicited by the semantically-related word-pairs (Experiment 1).

Despite the lower single-subject hit-rate in Experiment 3, however, there was no significant difference in the group-level N400 effect relative to that of Experiment 2. However, lower power to detect the effect may have occurred as a result of the significantly fewer clean EEG trials contributing to the analysis in Experiment 3 when compared with Experiment 2. Several factors may have led to this difference in trial numbers. First, due to the longer stimuli, the sentence task takes more time to complete than a word-pair task (~35 versus ~20 min, respectively), which may increase participant fatigue and decrease their ability to remain still during the testing session. Also note that, for patients with disorders of consciousness, due to potential fatigue and related concerns, shorter tasks are generally preferable. Second, in keeping with previous sentence studies (Holcomb and Neville, 1991; Kutas et al., 1987), the entire sentence was spoken naturally and the time-point of onset of the target word was identified in the digital recording for analysis. As a result, the baseline period of the ERP is not guaranteed to be silent in a sentence task, while in a word-pair task this can be ensured. Together these may result in fewer clean trials being available for analysis, and therefore decrease the sensitivity of the sentence task to detecting single-subject N400 effects.

5. General discussion

Across three experiments we have shown that eliciting a statistically reliable N400 effect in a single subject is not a trivial undertaking (see also Appendix D). This is true even when the individual in question is demonstrably conscious and in possession of normal linguistic processing abilities, thus highlighting the importance of rigorous task design when endeavoring to detect residual linguistic function in non-communicative patients. Specifically, the current data indicate that word-pair stimuli carefully generated to be as strongly forward-associated as possible provide the highest level of single-subject sensitivity for the N400 effect (see Fig. 6).
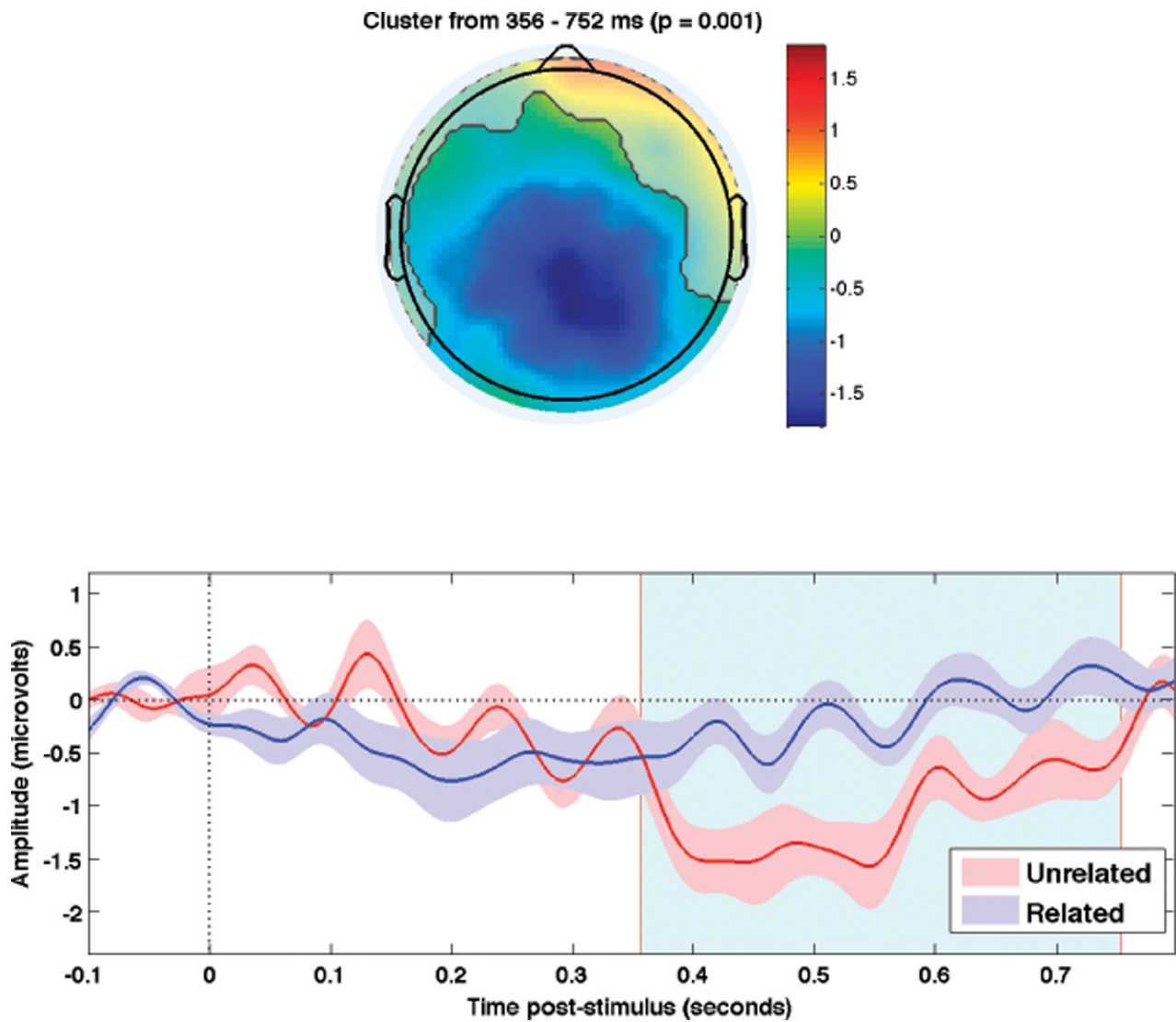
Fig. 5. High-cloze sentences. Grand average N400 effects (unrelated targets < related targets) in Experiment 3. Upper panels highlight the spatial extent of the significant spatiotemporal cluster (i.e., all electrodes that contributed to the cluster). Color bars show average amplitude differences between unrelated and related targets across the temporal extent of the spatiotemporal cluster. Lower panels show the means of the ERPs within the respective spatial clusters (±1 standard error). The temporal boundaries of each cluster are shaded in light blue.

Experiment 1 demonstrated the strong role of task demands on the detectability of the N400 effect. The same word-pair stimuli were considerably more likely to elicit significant single-subject effects when participants were engaged in a task than when they were passively listening. While it is not unexpected that the task relevance of the stimuli affects the magnitudes of the N400 effect (Bentin et al., 1993), it is of critical importance when applied to groups of patients who are unlikely to be able to follow commands, such as those with disorders of consciousness. Indeed, if a patient is capable of behaviorally following commands, there is no question regarding the presence of linguistic ability, rendering an N400 assessment purposeless. Non-communicative patients, however, are precisely those for whom markers of linguistic processing may be diagnostically and prognostically beneficial (Coleman et al., 2009, 2007). While a minority of non-communicative patients can covertly follow commands (Fernandez-Espejo and Owen, 2013), the majority are unable to do so and would thereby be precluded from eliciting evidence of residual cognitive function with the passive semantic priming task employed

in Experiment 1 — even if that function was preserved. The development of paradigms that are reliable in the absence of patient cooperation is therefore crucial to ensuring the acquisition of informative data regarding each patient's residual cognition — be it conscious or unconscious.

Across the three experiments, three classic N400 tasks were employed: namely, tasks using semantically-similar word pairs, normatively-associated word pairs, and high-cloze sentences. It is evident from the current data that a word-pair task generated from normative associations is the most sensitive approach to detecting single-subject N400 effects during passive listening. This task produced significant N400 effects in 50% of healthy participants, compared with 0% in the semantically-similar word-pair task, and 17% in the high-cloze sentence task. As targets in normatively-associated word pairs are highly likely to be produced by healthy individuals in response to the prime (Experiment 2), prediction of these targets is likely to be stronger than that of targets that are semantically-similar, but not necessarily associated to the prime (Experiment 1). This would thereby minimize the amplitude of the N400 wave elicited by related
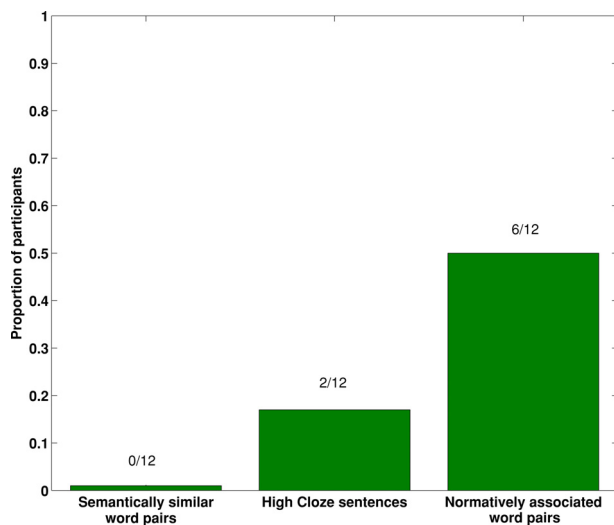
Fig. 6. The proportions of participants returning significant N400 effects during passive listening across the three N400 paradigms.

targets, and maximize the magnitude of the N400 effect — i.e., the difference between the N400s elicited by related and unrelated targets. Indeed, at the group level, the associative N400 effect started approximately 240 ms earlier than in the semantically-similar word-pair task, suggesting a greater fluency of processing of highly predictable targets. The statistically indistinguishable scalp distributions of the N400 effects across the three experiments indicate that the processes indexed by these N400 effects are functionally equivalent, albeit to differing magnitudes and latencies. The current data therefore suggest that when endeavoring to elicit a statistically reliable N400 effect in a single subject during passive listening, the optimal choice is to use pairs of words that are extremely strongly associated according to the word association task.

A further important factor to take into consideration when seeking to detect single-subject N400 effects is how the unrelated target stimuli are generated. It is common in the literature to repeat words within an experiment such that unrelated word-pairs are rearranged versions of related word-pairs — e.g. cat–dog and chair–table are recombined to form chair–dog and cat–table, (Kotchoubey et al., 2005; Kotchoubey, 2005; Rama et al., 2010). This approach has the benefit of ensuring that the target words are identical in the related and unrelated conditions (and the same primes occur across conditions as well). However, as shown in Appendix D, repeating targets in this way leads to significant order effects that reduce the overall amplitude of the N400 effect. Specifically, the N400 effect between targets heard in an unrelated pair before being heard in a related pair (*unrelated–related* order) was significantly smaller than that elicited by targets heard in a related pair before being heard in an unrelated pair (*related–unrelated* order). Indeed, in the experiment reported in Appendix D only the targets heard in the *related–unrelated* order elicited a significant N400 effect. This difference perhaps reflects a greater level of unexpectedness (and hence N400 amplitude) for unrelated targets when the appropriate context has recently been primed — as in the *related–unrelated* order. Irrespective of the mechanism, the overall N400 effect across all target stimuli would therefore be the mean of the significant *related–unrelated* effect, and the smaller and non-significant *unrelated–related* effect. The differential effects of target order would therefore reduce the overall N400 effect amplitude, and thereby its detectability.

It is therefore optimal to employ unique words throughout the experiment. This approach is more time-consuming to design, however. To be able to conclude that differences in target N400s result from the priming manipulation, it is necessary to carefully control the

stimuli along the range of factors that are known to affect N400 amplitudes, such as word frequency, familiarity, and other lexical variables (Kutas and Federmeier, 2011). For example, the stimuli employed in Experiment 2 were matched across 10 linguistic measures (see Supplementary materials Table F2). Moreover, to ensure that ERP differences between related and unrelated targets are due to priming, a separate group of participants was presented with each stimulus in isolation — i.e., in the absence of priming. This validation procedure confirmed the careful matching of related and unrelated targets, as well as related and unrelated primes, and validated their use in the word-pair context. Due to the complexities of matching stimuli in this way, we would encourage researchers to employ the stimuli used in Experiment 2 when investigating linguistic processing in single-subject native English speakers. The full stimulus list can be found in Appendix B.

Despite the markedly different levels of single-subject sensitivity between the normative-association task and the high-cloze task (17% versus 50%, respectively), there were no significant differences in the group average N400 effects. The lower single-subject hit-rate may be due to the significantly lower number of clean trials that were available for analysis in the sentence task. As the target words in this task occurred within a sentence of natural speech, they had a less controlled baseline than the word-pair tasks that have a guaranteed silent baseline prior to stimulus onset. Similarly, it is less straightforward to accurately identify target onsets in sentence designs as co-articulation causes word boundaries to be less clear than in a single-word event-related design. The ability to present a large number of stimuli in a relatively short time (20 min for Experiment 2), and to have a guaranteed silent baseline, further illustrates the benefit of the normative-association word-pair task for detecting single-subject N400s.

From the clinical perspective, it is known that detecting a range of covert cognitive capacities in patients with disorders of consciousness can impact diagnosis and prognosis (Owen, 2013). Indeed, there is evidence to suggest that some fMRI-detected responses to speech may have prognostic value (Coleman et al., 2009) and may even reflect processing that requires consciousness (Davis et al., 2007). Due to the greater clinical utility of EEG, several studies have endeavored to identify residual linguistic functioning by means of the N400 effect. However, it is a challenge to interpret the results of many of these studies as they typically have not statistically verified the presence of N400 effects in their patients, have relied on somewhat unconventional transformations of the ERP data, or have not estimated the sensitivity of the technique with a healthy control group (Hinterberger et al., 2005; Kotchoubey et al., 2005; Kotchoubey, 2005; Rama et al., 2010; Schoenle and Witzke, 2004; Steppacher et al., 2013). Furthermore, poor reporting of the method of stimulus generation is common, as is the use of inadequate stimulus validation procedures. Indeed, the current data emphasize the fact that eliciting a reliable N400 effect in a single subject is not as simple as presenting related and unrelated words. Rather, there are a number of crucial considerations that have considerable impact on the reliability of the task outcome.

While there are no guidelines for an 'acceptable' level of sensitivity for a specific test, the 50% hit-rate of the normative-association task is relatively low compared to some other neuroimaging markers of cognition (Boly et al., 2007; Chennu et al., 2013; Cruse et al., 2011; Naci et al., 2013). It is possible that alternative analysis techniques would return higher detection rates of linguistic function in single subjects — thereby making the approach more clinically viable. The analysis approach employed here has the benefit of being conventional within the N400 and ERP literature; the effects are analyzed within their native space, rather than after data transformation (Connolly et al., 1999; Kotchoubey, 2005; Steppacher et al., 2013). This ensures that any observed effects may be interpreted relative to the existing and extensive body of N400 research. The cluster-mass analyses which we implemented (Oostenveld et al., 2011) are also entirely data-driven

in the identification of effects, and as such do not rely on the selective consideration of individual electrodes or time-windows that may be inappropriate for the severely injured brain that may have undergone some level of cortical reorganization. These analyses also simultaneously control for the large number of multiple comparisons that are inherent in ERP analysis in a way that is both statistically rigorous and ensures maximal power to detect effects (Maris and Oostenveld, 2007).

However, the level of statistical conservatism required in a clinical setting is proportional to the stakes of the outcome (Cruse et al., in press). For example, if the detection of an ostensible N400 effect in a non-communicative patient will not lead to alterations in their care, then a higher level of false positives may be acceptable. However, if long-lasting changes to care were to hinge on the result of this assessment, then a low level of false positives may be more preferable. A more liberal approach to analyzing the current data would be to restrict the single-subject analysis to the time-window in which the group average N400 effect was significant. Indeed, our reported analyses were restricted to 200–800 ms post-stimulus on the basis of previous studies, while the significant group effects actually onset somewhat later – around 300 ms post-stimulus – perhaps due to the variability in the point of identification of spoken words. One possible method to achieve a higher hit-rate is to select the electrode at which the average difference between related and unrelated target ERPs is greatest, and test the significance of this difference using a single *t*-test. This approach is clearly poor statistical practice because it involves double-dipping of the data, thus creating higher false positive rates. Nevertheless, when we conducted this analysis on Experiment 2, the hit-rate increased from 50% to 75% (9/12). Ultimately the level of statistical conservatism required from an assessment of linguistic function is the decision of both clinicians and researchers, and should be balanced against the stakes of the outcome of that statistical test (see Cruse et al., in press). Furthermore, with the continuing development of sophisticated single-trial ERP analysis methods, it is possible that increasingly more clinically-viable trade-offs between hit-rates and false-alarms may be achieved (Geuze et al., 2013).

6. Conclusions

The N400 ERP effect may allow for the bedside identification of residual linguistic function in non-communicative patients, thus providing information that can impact both diagnosis and prognosis (Coleman et al., 2009, 2007; Owen, 2013). However, it is not a trivial procedure to elicit a statistically reliable N400 effect in a single subject. Rather, careful control of both stimuli and task demands is required. Specifically, the current data indicate that the most sensitive approach to eliciting significant single-subject N400 effects is with word-pair stimuli that are extremely strongly normatively associated. The optimization of assessments of residual cognition in this way will not only ensure greater reliability of findings, but may also ultimately increase diagnostic accuracy in those patients whose linguistic abilities are entirely unclear from their external behaviors.

Conflicts of interest

All authors declare no conflicts of interest.

Funding

Appendix A.   Supplementary material

References

Balconi, M., Arangio, R., Guarnerio, C., 2013. Disorders of consciousness and N400 ERP measures in response to a semantic task. Journal of Neuropsychiatry and Clinical Neurosciences 25, 237–43. http://dx.doi.org/10.1176/appi.neuropsych.12090227, 24026717.

Bentin, S., Kutas, M., Hillyard, S.A., 1993. Electrophysiological evidence for task effects on semantic priming in auditory word processing. Psychophysiology 30, 161–9. http://dx.doi.org/10.1111/j.1469-8986.1993.tb01729.x, 8434079.

Bentin, S., McCarthy, G., Wood, C.C., 1985. Event-related potentials, lexical decision and semantic priming. Electroencephalography and Clinical Neurophysiology 60, 343–55. http://dx.doi.org/10.1016/0013-4694(85)90008-2, 2579801.

Block, C., Baldwin, C., 2010. Cloze probability and completion norms for 498 sentences: behavioral and neural validation using event-related potentials. Behavior Research Methods 42, 665–70. http://dx.doi.org/10.3758/BRM.42.3.665, 20805588.

Boly, M., Coleman, M.R., Davis, M.H., Hampshire, A., Bor, D., Moonen, G. et al, 2007. When thoughts become action: an fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. Neuroimage 36, 979–92. http://dx.doi.org/10.1016/j.neuroimage.2007.02.047, 17509898.

Chennu, S., Finoia, P., Kamau, E., Monti, M.M., Allanson, J., Pickard, J.D. et al, 2013. Dissociable endogenous and exogenous attention in disorders of consciousness. NeuroImage: Clinical 3, 450–61. http://dx.doi.org/10.1016/j.nicl.2013.10.008, 24273727.

Coleman, M.R., Davis, M.H., Rodd, J.M., Robson, T., Ali, A., Owen, A.M. et al, 2009. Towards the routine use of brain imaging to aid the clinical diagnosis of disorders of consciousness. Brain: a Journal of Neurology 132, 2541–52. http://dx.doi.org/10.1093/brain/awp183, 19710182.

Coleman, M.R., Rodd, J.M., Davis, M.H., Johnsrude, I.S., Menon, D.K., Pickard, J.D. et al, 2007. Do vegetative patients retain aspects of language comprehension? Evidence from fMRI. Brain: a Journal of Neurology 130, 2494–507. http://dx.doi.org/10.1093/brain/awm170, 17827174.

Connolly, J.F., Mate-Kole, C.C., Joyce, B.M., 1999. Global aphasia: an innovative assessment approach. Archives of Physical Medicine and Rehabilitation 80, 1309–15. http://dx.doi.org/10.1016/S0003-9993(99)90035-7, 10527093.

Cree, G.S., McRae, K., 2003. Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). Journal of Experimental Psychology. General 132, 163–201. http://dx.doi.org/10.1037/0096-3445.132.2.163, 12825636.

Cruse, D., Chennu, S., Chatelle, C., Bekinschtein, T.A., Fernandez-Espejo, D., Pickard, J.D. et al, 2011. Bedside detection of awareness in the vegetative state: a cohort study. Lancet 378, 2088–94. http://dx.doi.org/10.1016/S0140-6736(11)61224-5, 22078855.

Cruse, D., Gantner, I., Soddu, A., Owen, A.M. (in press), Lies, damned lies, and diagnoses: estimating the clinical utility of assessments of covert awareness in the vegetative state. BrainInjury

Cruse, D., Owen, A.M., 2010. Consciousness revealed: new insights into the vegetative and minimally conscious states. Current Opinion in Neurology 23, 656–60. http://dx.doi.org/10.1097/WCO.0b013e32833fd4e7, 20859205.

Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F. et al, 2007. Dissociating speech perception and comprehension at reduced levels of awareness. Proceedings of the National Academy of Sciences of the United States of America 104, 16032–7. http://dx.doi.org/10.1073/pnas.0701309104, 17938125.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods 134, 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009, 15102499.

Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Näätänen, R. et al, 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 120, 1883–908. http://dx.doi.org/10.1016/j.clinph.2009.07.045, 19796989.

Fernandez-Espejo, D., Owen, A.M., 2013. Detecting awareness after severe brain injury. Nature Reviews. Neuroscience 14, 801–9. http://dx.doi.org/10.1038/nrn3608, 24088810.

Geuze, J., van Gerven, M.A.J., Farquhar, J., Desain, P., 2013. Detecting semantic priming at the single-trial level. PloS One 8, e60377. http://dx.doi.org/10.1371/journal.pone.0060377, 23565237.

Giacino, J.T., Ashwal, S., Childs, N.L., Cranford, R., Jennett, B., Katz, D.I. et al, 2002. The minimally conscious state: definition and diagnostic criteria. Neurology 58, 349–53. http://dx.doi.org/10.1212/WNL.58.3.349, 11839831.

Hinterberger, T., Wilhelm, B., Mellinger, J., Kotchoubey, B., Birbaumer, N., 2005. A device for the detection of cognitive brain functions in completely paralyzed or unresponsive patients. IEEE Transactions on Biomedical Engineering 52, 211–20. http://dx.doi.org/10.1109/TBME.2004.840190.

Holcomb, P.J., Neville, H.J., 1991. Natural speech processing: an analysis using event-related brain potentials. Psychobiology 19, 286–300.

Ibanez, A., López, V., Cornejo, C., 2006. ERPs and contextual semantic discrimination: degrees of congruence in wakefulness and sleep. Brain and Language 98, 264–75, 16782185.

Ibáñez, A.M., Martín, R.S., Hurtado, E., López, V., 2009. ERPs studies of cognitive processing during sleep. International Journal of Psychology: Journal international de Psychologie 44, 290–304. http://dx.doi.org/10.1080/00207590802194234, 22029558.

Ibáñez, A.M., San, Martín R., Hurtado, E., López, V., 2008. Methodological considerations related to sleep paradigm using event related potentials. Biological Research 41, 271–5, 19399340.

Jennett, B., Plum, F., 1972. Persistent vegetative state after brain damage. A syndrome in search of a name. Lancet 1, 734–7, 4111204.

Kalmar, K., Giacino, J.T., 2005. The JFK Coma Recovery Scale — Revised. Neuropsychological Rehabilitation 15, 454–60. http://dx.doi.org/10.1080/09602010443000425, 16350986.

Kiefer, M., 2002Brain Research. Cognitive Brain Research 13 (1), 27–39. http://dx.doi.org/10.1016/S0926-6410(01)00085-4, 11867248.

Kotchoubey, B., 2005. Apallic syndrome is not apallic: is vegetative state vegetative? Neuropsychological Rehabilitation 15, 333–56. http://dx.doi.org/10.1080/09602010443000416, 16350976.

Kotchoubey, B., Lang, S., Mezger, G., Schmalohr, D., Schneck, M., Semmler, A. et al, 2005. Information processing in severe disorders of consciousness: vegetative state and minimally conscious state. Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology 116, 2441–53. http://dx.doi.org/10.1016/j.clinph.2005.03.028, 16002333.

Kutas, M., 1993. In the company of other words: electrophysiological evidence for single-word and sentence context effects. Language and Cognitive Processes 8, 533–72. http://dx.doi.org/10.1080/01690969308407587.

Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends in Cognitive Sciences 4, 463–70. http://dx.doi.org/10.1016/S1364-6613(00)01560-6, 11115760.

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, 621–47. http://dx.doi.org/10.1146/annurev.psych.093008.131123, 20809790.

Kutas, M., Neville, H.J., Holcomb, P.J., 1987. A preliminary comparison of the N400 response to semantic anomalies during reading, listening and signing. Electroencephalography and Clinical Neurophysiology 39, 325–30.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods 164, 177–90. http://dx.doi.org/10.1016/j.jneumeth.2007.03.024, 17517438.

McCarthy, G., Wood, C.C., 1985. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalography and Clinical Neurophysiology 62, 203–8. http://dx.doi.org/10.1016/0168-5597(85)90015-2, 2581760.

McRae, K., Boisvert, S., 1998. Automatic semantic similarity priming. Journal of Experimental Psychology: Learning, Memory, and Cognition 24, 558–72.

McRae, K., Cree, G.S., Seidenberg, M.S., Mcnorgan, C., 2005. Semantic feature production norms for a large set of living and nonliving things. Behavior Research Methods 37, 547–59. http://dx.doi.org/10.3758/BF03192726, 16629288.

McRae, K., de Sa, V.R., Seidenberg, M.S., 1997. On the nature and scope of featural representations of word meaning. Journal of Experimental Psychology. General 126, 99–130. http://dx.doi.org/10.1037/0096-3445.126.2.99, 9163932.

McRae, K., Khalkhali, S., Hare, M., 2012. Semantic and associative relations: examining a tenuous dichotomy, in: Reyna, V.F., Chapman, S.B., Dougherty, M.R., Confrey, J. (Eds.), The Adolescent Brain: Learning, Reasoning, and Decision Making. Washington, DC: American Psychological Association, pp. 39–66.

Naci, L., Cusack, R., Jia, V.Z., Owen, A.M., 2013. The brain's silent messenger: using selective attention to decode human thought for brain-based communication. Journal of Neuroscience: the Official Journal of the Society for Neuroscience 33, 9385–93. http://dx.doi.org/10.1523/JNEUROSCI.5577-12.2013, 23719806.

Nelson, D.L., McEvoy, C.L., Schreiber, T.A. (1998), The University of South Florida word association, rhyme, and word fragment norms. URL ⟨5:underline ⟩http://web.usf.edu/FreeAssociation/⟨/5:underline⟩

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, 1–9, 21837235.

Ortu, D., Allan, K., Donaldson, D.I., 2013. Is the N400 effect a neurophysiological index of associative relationships? Neuropsychologia 51, 1742–8. http://dx.doi.org/10.1016/j.neuropsychologia.2013.05.003, 23707682.

Owen, A.M., 2013. Detecting consciousness: a unique role for neuroimaging. Annual Review of Psychology 64, 109–33. http://dx.doi.org/10.1146/annurev-psych-113011-143729, 23043305.

Rabovsky, M., McRae, K., 2014. Simulating the N400 ERP component as semantic network error: insights from a feature-based connectionist attractor model of word meaning. Cognition 132 (1), 68–89.

Rama, P., Relander-Syrjanen, K., Ohman, J., Laakso, A., Naatanen, R., Kujala, T., 2010. Semantic processing in comatose patients with intact temporal lobes as reflected by the N400 event-related potential. Neuroscience Letters 474, 88–92. http://dx.doi.org/10.1016/j.neulet.2010.03.012, 20226842.

Rhodes, S.M., Donaldson, D.I., 2008. Association and not semantic relationships elicit the N400 effect: electrophysiological evidence from an explicit language comprehension task. Psychophysiology 45, 50–9, 17868263.

Schoenle, P.W., Witzke, W., 2004. How vegetative is the vegetative state? Preserved semantic processing in VS patients — evidence from N 400 event-related potentials. NeuroRehabilitation 19, 329–34, 15671587.

Schroeder, M.R., 1968. Reference signal for signal quality studies. Journal of the Acoustical Society of America 44, 1735–6. http://dx.doi.org/10.1121/1.1911323.

Steppacher, I., Eickhoff, S., Jordanov, T., Kaps, M., Witzke, W., Kissler, J., 2013. N400 predicts recovery from disorders of consciousness. Annals of Neurology 73, 594–602. http://dx.doi.org/10.1002/ana.23835, 23443907.

Vogel, E.K., Luck, S.J., Shapiro, K.L., 1998. Electrophysiological evidence for a post-perceptual locus of suppression during the attentional blink. Journal of Experimental Psychology. Human Perception and Performance 24, 1656–74. http://dx.doi.org/10.1037/0096-1523.24.6.1656, 9861716.

Wilding, E.L., 2006. The practice of rescaling scalp-recorded event-related potentials. Biological Psychology 72, 325–32. http://dx.doi.org/10.1016/j.biopsycho.2005.12.002, 16442202.