Brain & Language 117 (2011) 12–22

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

EEG decoding of semantic category reveals distributed representations for single concepts

Brian Murphy [a,*], Massimo Poesio [a,b,d], Francesca Bovolo [b], Lorenzo Bruzzone [b], Michele Dalponte [b], Heba Lakany [c]

[a] Centre for Mind/Brain Sciences, University of Trento, Corso Bettini 31, 38068 Rovereto (TN), Italy
[b] Department of Information Engineering and Computer Science, University of Trento, Italy
[c] Bioengineering Unit, University of Strathclyde, United Kingdom
[d] School of Computer Science and Electronic Engineering, University of Essex, United Kingdom

ARTICLE INFO

Article history:
Accepted 18 September 2010

Keywords:
Concepts
Semantics
Categorisation
EEG
Data mining
Machine learning
Distributed representations
Exclusion of confounds

ABSTRACT

Achieving a clearer picture of categorial distinctions in the brain is essential for our understanding of the conceptual lexicon, but much more fine-grained investigations are required in order for this evidence to contribute to lexical research. Here we present a collection of advanced data-mining techniques that allows the category of individual concepts to be decoded from single trials of EEG data. Neural activity was recorded while participants silently named images of mammals and tools, and category could be detected in single trials with an accuracy well above chance, both when considering data from single participants, and when group-training across participants. By aggregating across all trials, single concepts could be correctly assigned to their category with an accuracy of 98%. The pattern of classifications made by the algorithm confirmed that the neural patterns identified are due to conceptual category, and not any of a series of processing-related confounds. The time intervals, frequency bands and scalp locations that proved most informative for prediction permit physiological interpretation: the widespread activation shortly after appearance of the stimulus (from 100 ms) is consistent both with accounts of multi-pass processing, and distributed representations of categories. These methods provide an alternative to fMRI for fine-grained, large-scale investigations of the conceptual lexicon.

1. Introduction

Achieving a clearer picture of categorial distinctions in the brain is essential for our understanding of the conceptual lexicon, but much more fine-grained investigations both on categorial distinctions and on other aspects of conceptual representation (e.g. Maguire, Brier, & Ferree, 2010; Sachs et al., 2008; Siri et al., 2007) will be required in order for this evidence to contribute to lexical research. Though semantics clearly occupies a central role in the human language capacity, as transfer of meaning is the aim of much purposeful communication, our understanding of its functional instantiation and localisation in the brain is far from complete (Pulvermüller, 2002). One reason for this is the complex, multi-modal character of meaning: conceptual representations are jointly rooted in the language, visual and other systems, as is evidenced by the influence of mother tongue on visual perception (Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009), and imaging results suggesting that classically "visual" centres, particularly the left fusiform gyrus, are sensitive to language semantics independently of their visual regularities (Simons, Koutstaal, Prince, Wagner, & Schacter, 2003; Vuilleumier, Henson, Driver, & Dolan, 2002).

Recordings of brain activity from neuroimaging and electrophysiological techniques have long established themselves as crucial sources of evidence in the study of conceptual knowledge. Imaging studies (e.g. Chao, Weisberg, & Martin, 2002; Martin & Chao, 2001) have provided ample evidence supporting hypotheses concerning the localization of conceptual knowledge derived from earlier work with patients displaying semantic category deficits (Caramazza & Mahon, 2003; Warrington & Shallice, 1984). Now computationally intensive methods are increasingly becoming the method of choice for analysing such data (Haxby et al., 2001; Shinkareva et al., 2008). The low signal-to-noise ratio of neural recordings means that many regularities remain hidden and can only be detected through large-scale pattern analysis (Norman, Polyn, Detre, & Haxby, 2006).

However, such methods have generally led to very high-level insights about conceptual representations, rather than to answers to the more fine-grained questions raised by studies in other disciplines, such as behavioural experiments (Collins & Quillian, 1970; Neely, 1991; Quillian, 1967; Rosch, 1973; Smith, Shoben, & Rips,

* Corresponding author.
E-mail address: brian.murphy@unitn.it (B. Murphy).

0093-934X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2010.09.013

1974), judgements elicited from informants (McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008), language acquisition work (Bloom, 2002; Mandler, 1992), computational simulations (Plaut, 1995; McClelland & Rogers, 2003) or models derived from corpora (Baroni, Murphy, Barbu, & Poesio, 2010; Lund, Burgess, & Atchley, 1995; Landauer & Dumais, 1997; Poesio & Almuhareb, 2005; Padó & Lapata, 2007; Rapp, 2004) – not to mention work on formal models such as Fillmore (1982), Gärdenfors (2000), Kamp and Partee (1995), and Pustejovsky (1995). Most studies examine only very coarse-grained distinctions, for example comparing abstract and concrete concepts, verbs and nouns, or natural and artefactual kinds (Pulvermüller, 2002). A further limitation of this work is that the well-established differences seen between groups of stimuli representing natural and artefactual concepts (e.g. Chao et al., 2002; Kiefer, 2001) are compatible with a number of interpretations. Since the natural concepts investigated are predominantly animals and plants, and artefactual stimuli are typically small functional objects conventionally termed 'tools', several semantic distinctions provide plausible explanations: biological/non-biological entities, larger/smaller objects, moving/non-moving entities, sentient/non-sentient entities and manipulable/non-manipulable objects. More pessimistically, such effects could be due to confounding variables (e.g. lexical frequency, image complexity) that may vary systematically across stimulus groups. Finally, the statistics applied to neural data tend to identify single or small numbers of loci in the neural activity (at a precise location and approximate timing with fMRI; at a precise time and approximate location with EEG) that are activated differentially by each group of stimuli. This is despite the fact that recent experimental work demonstrates more widely spread and overlapping activations (e.g. Haxby et al., 2001; Pulvermüller, 2005; Mitchell et al., 2008), consistent with theories that favour distributed representations (Barsalou, 2003; Spitzer, 1999; Tyler & Moss, 2001).

We believe that large-scale, systematic explorations of the mental lexicon with neural data are necessary, involving both a more careful analysis of conceptual distinctions and a greater range of categories. However, the high cost of fMRI studies may make such systematic explorations impractical. Fortunately, the theory that conceptual knowledge is distributed would predict some form of synchronization between the relevant areas, of the kind that may be detected using EEG.

In this paper we detect which of two familiar semantic classes are being processed (land mammals or work tools, which should have similar conceptual representations across languages), by applying data mining and machine learning techniques to single trials of recorded EEG signals. The behavioural task was a simple one that demands activation of conceptual representations: silent naming of image stimuli, which requires object recognition, lexical retrieval and sub-vocal language production. As is discussed in the following section, tasks that involve conceptual processing have been shown to elicit neural responses that persist across tasks in the visual and auditory modality, and for image and lexical stimuli, suggesting that they access shared representations, jointly based on linguistic knowledge and perceptual experience.

Our adaptive analyses search through the signal to identify the scalp locations, time intervals and frequency bands at which the categorial distinction is most clearly present. On this basis, we can predict the semantic category of a single stimulus presentation with an accuracy well above chance both when training and evaluation data come from the same participant (mean 72%, $p < 0.001$, binomial test with chance at 0.5), and when they come from different participants (mean 61%, $p < 0.001$). By amalgamating predictions over multiple analyses, the categorial membership of single stimuli can be determined with an accuracy of 98%. Crucially, since the analysis technique can classify single stimulus trials, a range of lexical, visual and task-related confounding variables can be excluded as possible explanations for the performance of the algorithm, allowing us to conclude that the activations that are identified correspond to conceptual distinctions rather than group differences in the cognitive load of perceptual processing, lexical search, or other task-related activity.

The paper proceeds as follows. The next section provides additional theoretical background, particularly on the relationship between the patterns identified by the algorithm and underlying brain physiology and function. Section 3 describes in detail the experimental paradigm used, and the analytical methods applied. Classification results are given in Section 4, for analyses on individual participant sessions and group-prediction across participants, together with a regression model analysis to establish that the accuracies achieved are not due to confounding variables, whether singly or in combination. Finally, in the discussion section we consider the physical and functional interpretation of the results, and what future directions could be taken with this analytical paradigm.

2. Background

Early work on the neural instantiation of conceptual knowledge suggested that different kinds of semantic information are isolated in anatomically distinct areas of the brain. Work with brain damaged patients (Warrington & Shallice, 1984) found selective impairment of conceptual knowledge on animals and tools, leading to the hypothesis that sensory (primarily visual) and functional (conventional uses and associated behaviours) properties were located in different brain areas. Brain imaging work that followed (Martin, Wiggs, Ungerleider, & Haxby, 1996) was consistent with the related proposal (Caramazza & Shelton, 1998) that animate and inanimate kinds were localised in different parts of cortex. Such studies also provided evidence that the location of these activations is largely consistent across individuals and across tasks (Martin & Chao, 2001).

More recent work however seems to indicate that the neural patterns associated with the representation of conceptual knowledge are more complex than such 'hot-spot' models would suggest. Haxby et al. (2001) demonstrated that neural activity in marginal and overlapping areas of cortex, beyond the foci of activity identified by conventional statistical tests, could be used to correctly decode the semantic category of image stimuli presented during an fMRI session. Mitchell et al. (2008) showed that the meaning of concrete nouns could be decomposed into properties that held a predictive relationship with widely distributed activities in the brain.

Much of the work just cited uses image stimuli, and so these effects could be interpreted as being due to visual regularities that are confounded with semantic class. But distributed conceptual representations have also been observed during passive linguistic tasks that did not involve visual presentation (Pulvermüller, 2005). And the dissociations in neural activity that are typically seen during object recognition have also been observed during reading of word stimuli, both with PET (Perani et al., 1999) and fMRI (Chao, Haxby, & Martin, 1999). The fact that congenitally blind participants also show these same patterns (Mahon, Anzellotti, Schwarzbach, Zampini, & Caramazza, 2009), demonstrates that this cannot be a secondary effect of visual experience evoked by word stimuli. Further evidence for the modality-independence of these representations at a finer spatial scale is provided by Connolly and Haxby (2010), whose clustering analysis shows that while patterns of neural activity in early visual areas reflect perceptual similarities, the ventral temporal cortex encodes conceptual distinctions instead.

Such empirical work is consistent with recent theories that hold that conceptual knowledge is distributed, perhaps taking the form of a 'conceptual map' (Spitzer, 1999), of 'situated simulations' (Barsalou, 2003) or of a 'word web' (Pulvermüller, 2002), according to which concepts are represented in 'neuron webs' with distinct cortical topographies, and where conceptual representations are jointly determined by perceptual experience and linguistic knowledge.

This suggests the need for some form of synchronization between the firing of these neurons, of the kind that might be detected with EEG. Indeed Tallon-Baudry and Bertand (1999) make just such a proposal, and find evidence of event-related synchronisation in the gamma band (≈40 Hz) during object recognition. Here we are interested in differential activation to semantic classes, and category specific differences in spectral power have been found at lower frequencies (in the band 1–20 Hz) with MEG (Gilbert, Shapiro, & Barnes, 2009) for animals versus tools, and with EEG for faces versus other objects (Rousselet, Husk, Bennett, & Sekuler, 2007). ERP evidence of category specificity has been found at both early (≈150 ms, Itier & Taylor, 2004; Rossion, Joyce, Cottrell, & Tarr, 2003) and late (≈400 ms, Kiefer, 2001) intervals after onset during image processing tasks. Some work interprets the first interval as being tied to perceptual differences and the second to semantic processing – a conclusion corroborated by Paz-Caballero, Cuetos, and Dobarro (2006), who found that the early ERPs were modulated by task, while the later ones were not. Other results suggest that the first burst of activity is driven by higher-level conceptual differences – Rousselet et al. (2007) compare faces to visually balanced (by image configuration, spectral frequencies, and brightness histograms) houses, meaning that low-level visual differences (form and texture) cannot account for the effects; and Rousselet, Mace, and Fabre-Thorpe (2004) see the same effects for animals versus inanimate objects, meaning that they are not specific to the interpretation of human faces (e.g. determining identity, reading expressions).

As already noted, the majority of neuroimaging studies of semantic category rely on group effects, revealed with grand averages across trials and participants. In the case of EEG, this is due to the impoverished nature of the signal that can be detected at the scalp. Here, we use a discriminative data-mining technique developed for Brain–Computer-Interaction purposes (Dalponte, Bovolo, & Bruzzone, 2007), to perform category-based classification of single trials. The algorithm first identifies the time interval, frequency band and combination of scalp locations in the EEG signals that best distinguish the categories. It is a supervised method in that a portion of the neural data, labelled for semantic category, is used for learning (the "training" data), and the remaining data is then used to test the quality of learning (the "evaluation" data) by comparing the true category labels to those predicted by the algorithm. The signal decomposition used (CSP, Koles, Lazar, & Zhou, 1990) yields measures of signal spectral power that respond selectively to the stimulus (analogous to event related synchronisation/desynchronisation, or ERS/ERD; see Pfurtscheller & Lopes da Silva, 1999)[1] and that correspond to synchronous neural assemblies, or networks of assemblies, that are functionally related to the processing or representation of semantic classes. A generic machine learning algorithm then uses these measures of spectral power to classify single stimulus trials that were excluded from the training phase.

3. Methods

3.1. Participants

Seven staff and post-graduate students at the University of Trento took part in the study, all native speakers of Italian. Five of the participants were male and two female (age range 25–33, mean 29). One identified herself as left-handed, and two as ambidextrous or of mixed-handedness. All had normal or corrected-to-normal vision. Participants received compensation of €7 per hour. The studies were conducted under the approval of the ethics committee at the University of Trento, and participants gave informed consent.

3.2. Experimental paradigm

Participants were asked to perform a silent naming task on grey-scale images of 30 land-mammals and 30 work tools. Each stimulus was presented six times, for a total 360 trials, their order being randomized on each session. [2] Participants sat in a relaxed upright position 60 cm from a computer monitor in reduced lighting conditions. Images were presented on a medium grey background and fell within a 10° viewing angle. The task duration was split into four blocks and participants were given the choice to pause between each. The cumulative task time did not exceed 45 min.

Each trial began with the presentation of a fixation cross for 0.5 s, followed by the stimulus image, a further fixation cross for 0.5 s and a blank screen for 2 s. Participants were instructed to silently name the object represented in their native tongue (Italian), using the first appropriate label that came to mind, and to press the keyboard space-bar with the left-hand to indicate they had found an appropriate word. If participants could not think of a suitable label, they were asked not to make a response. The image remained on the screen until the participant responded, or until a time-out of three seconds was reached. Participants were asked to keep still during the task, and to avoid eye-movements and facial muscle activity in particular, except during the 2 s blank period.

Mean reaction times varied from 1 s (participant C) to 1.8 s (participant G), and null-responses ranged from 1% to 7%. After the experimental session, each participant was asked to complete a questionnaire, in which they recorded the names they used during the EEG session. In particular they were asked to indicate whether they were unable to identify any of the stimuli, could not find an appropriate name, or used multiple labels on the various presentations of a single stimulus. Naming agreement in this task ranged from 87% to 90%.

3.3. Materials

As mentioned in the introduction, the semantic categories used in studies of this kind are often somewhat arbitrary and so can present problems for the interpretation of results. Here a set of 30 land mammals were chosen to be both non-domesticated and non-threatening, to avoid emotional confounds whether positive (e.g. pets) or negative (e.g. predators). Thirty hardware and garden implements were chosen as genuine work tools. Appropriate photographs were sourced from the internet, and normalised visually (see Fig. S.5 in the Supplementary materials): each image file measured 300 pixels square; the image proper was converted to grey-scale, superimposed on a homogeneous light-grey background and had maximal horizontal and vertical dimensions of 250 pixels; image contrast was normalised. [3] The concepts represented are listed below.

[1] In conventional measures of signal power (ERS/ERD, or ERSP in the terminology of Delorme & Makeig, 2003) power estimates are calculated relative to a baseline preceding each epoch. In this work, power estimates are absolute measures for each epoch, and comparisons are made across epochs. These measures will be termed "spectral power".

[2] The first three participants viewed a slightly broader set of images (42 animals, 44 tools) some of which were discarded in later experiments due to difficulty of naming. All analyses reported in this paper refer to the reduced set of 30 animals and 30 tools.

[3] Using the ImageMagick function *normalise*; see http://www.imagemagick.org/.

Land mammals ant-eater, armadillo, badger, beaver, bison, boar, camel, chamois, chimpanzee, deer, elephant, fox, giraffe, gorilla, hare, hedgehog, hippopotamus, ibex, kangaroo, koala, llama, mole, monkey, mouse, otter, panda, rhinoceros, skunk, squirrel, zebra (*Italian* formichiere, armadillo, tasso, castoro, bisonte, cinghiale, cammello, camoscio, scimpanzé, cervo, elefante, volpe, giraffa, gorilla, coniglio, riccio, ippopotamo, stambecco, canguro, koala, lama, talpa, scimmia, topo, lontra, panda, rinoceronte, puzzola, scoiattolo, zebra)
Work tools Allen key, axe, chainsaw, craft-knife, crowbar, file, garden fork, garden trowel, hacksaw, hammer, mallet, nail, paint brush, paint roller, penknife, pick-axe, plaster trowel, pliers, plunger, pneumatic drill, power-drill, rake, saw, scissors, scraper, screw, screwdriver, sickle, spanner, tape-measure (*Italian* brugola, ascia, motosega, taglierino, piede di porco, lima, forcone, paletta, seghetto, martello, mazza, chiodo, pennello, rullo, coltellino svizzero, piccone, cazzuola, pinza, stura lavandini, martello pneumatico, trapano, rastrello, sega, forbici, spatola, vite, cacciavite, falce, chiave inglese, metro)

3.4. Recording and preprocessing

The experiment was conducted at the CIMeC/DiSCoF laboratories at Trento University, using a 64-electrode Brain Vision BrainAmp system, recording at 500 Hz. A wide-coverage montage based on the 10–20 system was used, with a single right earlobe reference, and ground at location AFz. Electrode impedances were generally kept below 10 kΩ. However, sessions including electrodes that exceeded this limit were still included in subsequent analysis, as the techniques used proved robust to such noise.

Data preprocessing was conducted using the EEGLAB package (Delorme & Makeig, 2003). The data was band-pass filtered at 1–120 Hz to remove slow drifts in the signal and high-frequency noise, and then down-sampled to 300 Hz. [4] An ICA analysis was next applied using the EEGLAB implementation of the Infomax algorithm (Makeig, Bell, Jung, & Sejnowski, 1996). Artefactual ICA components were then identified and removed by hand in each data-set. In all cases, eye-artefact components were removed – usually one component for vertical movements including blinks, and another for horizontal movements. In a few cases components were found that isolated 50 Hz electrical noise, or electrode specific noise, but in general a conservative approach was taken, leaving such components in the signal to avoid the inadvertent removal of neural activity. It was not possible to remove muscle noise, as such activity was rarely isolated in a single component. However the adaptive analysis techniques used proved robust to both intermittent and continuous muscle noise.

3.5. Analysis

The analysis method (Dalponte et al., 2007) consists of a time/frequency window search to identify an information-rich band and interval for the distinction of interest; a supervised decomposition (Common Spatial Patterns, or CSP – see Model & Zibulevsky, 2006; Parra, Spence, Gerson, & Sajda, 2005; Philiastides, Ratcliff, & Sajda, 2006 for examples of other applications to cognitive neuroscience) to extract components of whole-scalp synchronous activity that are sensitive to this class distinction; and a general purpose machine learning algorithm (Support-Vector Machine or SVM) that uses the resulting measures of spectral power to predict the semantic class of each trial. Individual trial epochs are arbitrarily allocated to one of k interlaced partitions of equal size in a k-fold training/evaluation procedure (e.g. in a three-fold partition, one third of the data is held out in turn for evaluation, and training is performed on the remaining two thirds).

The time/frequency window search was developed for Brain–Computer Interface (BCI) applications, achieving state-of-the-art performance on a competition task involving imagined hand and foot movements (Dalponte et al., 2007). The goal of the technique is to define the best combination of time and frequency intervals for the separation of the analysed categories. This method can be divided into two parts: a search strategy, and a computation of resulting class separability. The search strategy used in this paper was a grid-based region search: the time range (0–500 ms, relative to the onset of the stimulus) and frequency range (0–50 Hz) of interest were divided into 15 intervals of equal length (33 ms and 3.3 Hz respectively), to yield a total of 120 time spans (0–33 ms, 0–67 ms, 0–100 ms,...,467–500 ms) and 120 frequency spans (0–3.3 Hz, 0–6.7 Hz, 0–10 Hz,...,46.7–50 Hz), and all binary combinations of these spans were explored. For each of these 14,400 time/frequency windows, the signal data was band-pass filtered, cropped in time, and the class-labelled trials were processed with CSP to extract class-sensitive measures of spectral power (as described below). The computation of class-separability used the Bhattacharyya distance, a probabilistic measure of cluster separation (Bhattacharyya, 1943; Bruzzone, Roli, & Serpico, 1995). Of all the windows examined, the combination of time and frequency spans that yields the largest separation between classes is considered the most informative window for classification purposes.

The decomposition method used, CSP (Koles et al., 1990), can be related to PCA or ICA, in that it extracts components of EEG activity that correspond to synchronous neural sub-assemblies. However it differs in that it is a supervised technique, in which the class-membership of each trial is also input to the algorithm. Thus, rather than extracting C new component signals of whole-scalp activity (where C is the number of input EEG channels) without regard to the task as PCA/ICA do, CSP extracts a series of components that are ranked by their sensitivity to the class-separation of interest, with an optimal variance for the two populations of EEG signals (i.e., high variance between classes and low variance within classes). In the considered case CSP extracts a set of C components jointly ranked by the amplitude of the signal found in trials of the first class (Mammals, $\omega_{Mammals}$) and inversely by their signal amplitude during trials of the second class (Tools, ω_{Tools}).

The CSP decomposition is calculated as follows. Let E be a single trial of EEG signal, represented as a $C \times T$ matrix, where C is the number of recording electrodes (64 in this case) and T is the number of samples per epoch. The transformation matrix that permits the extraction of signal components specific to the classification is derived with the following steps:

1. Let R be the normalized spatial covariance of a signal trial, defined as:

$$R = \frac{EE^T}{trace(EE^T)} \tag{1}$$

2. Compute the mean normalized spatial covariance matrices $\overline{R}_{Tools}$ and $\overline{R}_{Mammals}$ of trials ω_{Tools} and $\omega_{Mammals}$, respectively, by averaging over all the trials of the same class.
3. Compute the composite spatial covariance matrix as:

$$\overline{R}_C = \overline{R}_{Tools} + \overline{R}_{Mammals} = U_C \lambda_C U_C^T \tag{2}$$

where U_C is the matrix of the eigenvectors and λ_C is the diagonal matrix of the eigenvalues of $\overline{R}_C$.

[4] In some cases data was further filtered and re-sampled to lower rates (between 150 Hz and 100 Hz) to reduce the computational memory requirements of the analysis, after exploratory analyses had determined that the higher frequencies thus excluded were not informative for classification.

4. Let P be the whitening transformation matrix computed as:

$$P = \sqrt{\lambda_C^{-1}} U_C^T \tag{3}$$

5. Transform the matrices $\overline{R}_{Tools}$ and $\overline{R}_{Mammals}$ individually as:

$$S_{Tools} = P\overline{R}_{Tools}P^T, \quad S_{Mammals} = P\overline{R}_{Mammals}P^T \tag{4}$$

These two matrices share common eigenvectors, i.e. $S_{Tools} = U\lambda_{Tools}U^T$ and $S_{Mammals} = U\lambda_{Mammals}U^T$. Moreover the sum of the corresponding eigenvalues is always 1, i.e. $\lambda_{Tools} + \lambda_{Mammals} = I$. Thus, the eigenvector with the largest eigenvalue for class ω_{Tools} has the smallest eigenvalue for class $\omega_{Mammals}$, and vice-versa.

6. Compute the spatial filter SF to be used in the transformation of the signal, as:

$$SF = U^T \cdot P \tag{5}$$

7. Transform each trial EEG E into the desired specific components:

$$J = SF \cdot E \tag{6}$$

The matrix J represents the final output of the CSP algorithm. As is usually done in the literature (Ramoser, Gerking, & Pfurtscheller, 2000), we select the first and the last rows of this matrix as the components that are most representative for the classes ω_{Tools} and $\omega_{Mammals}$, respectively. [5] This procedure can be interpreted as extracting the event-related spectral activity (i.e. the relative event-related synchronisation) of two synchronous neural structures which have been found to have an optimally differential response to the stimulus categories of interest.

The final categorization step is based on a Support-Vector Machine (SVM) classifier (Boser, Guyon, & Vapnik, 1992; Vapnik, 1998), a state-of-the-art machine learning technique that has been widely adopted in recent years. Relative to other classifiers, it provides high classification accuracy and very good generalization; involves few control parameters; is not computationally intensive; and has proven effective in classification tasks where the number of training samples is limited (see Noble, 2006 for a non-technical introduction). The SVM input for each trial consisted of two numbers, derived as such from the pair of signal components produced by the CSP algorithm:

$$x_p = \log\left(\mathrm{var}(J_p)\right) \tag{7}$$

where J_p is the $1 \times T$ vector representing the pth signal component of$J$. Taking the variance of the vector produces a measure that is proportional to signal power, and the log function transforms the data to approximate a normal distribution (Ramoser et al., 2000). The resulting pair of measures represent modulations in spectral power which are sensitive to mammal and tool stimuli respectively:

$$x = [\log\left(\mathrm{var}(J_1)\right),\ \log\left(\mathrm{var}(J_M)\right)] \tag{8}$$

where J_1 and J_M represent the first and the last components of J, respectively. Features in x were scaled to a range of -1 to +1 across all trials. The SVM implementation used was LIBSVM (Chang & Lin, 2001).

[5] Preliminary analyses using more than one component per class did not provide a substantial improvement in classification accuracy. For example, over the seven individual analyses reported in Section 4.2, taking two components per class gave an average 1 percentage point improvement in classification accuracy, while taking five components per class gave no improvement.

4. Results

4.1. Time/frequency interval optimisation

As detailed in the previous section, a grid based region search was performed on the time-range 0–500 ms and the frequency range 0–50 Hz, in steps of 33 ms and 3.3 Hz respectively. In total just under 15 thousand time/frequency windows were considered for each participant. For each window, the EEG signals were band-pass filtered, divided into epochs of the relevant time interval, and decomposed with CSP to identify class specific components of neural activity. The Bhattacharyya metric of class separability was computed on the spectral power measures between the set of Mammal trials and the set of Tool trials. This process was carried out separately for each of the seven subject sessions. As mentioned in the previous section, the time/frequency cropping parameters for each participant that yield the best separation between mammal and tool epochs constitute an optimum, maximising the size of the effect between classes.

Fig. 1 (right panel) shows the time/frequency space for the EEG data from a single session (participant B). The value at each point on the graph is the mean Bhattacharyya distance for all windows that included that portion of the time/frequency space. As is apparent from the figure, the most informative interval (warm colours) runs from shortly after stimulus onset to around 350 ms, and the theta (4–8 Hz) and lower alpha (8–10 Hz) bands are the most relevant frequency range.

The optimal window selected according to the maximum separability in this case was (100–370 ms; 3–17 Hz), indicated in black. For comparison, the next three near-optimal solutions are indicated in white.

The left panel of Fig. 1 shows the corresponding plot taking a mean over the spaces of all seven participants. Here the time and frequency intervals are broadly similar, starting slightly later, and extending somewhat into the low beta band. The optimal windows for each individual participant are shown in white, and the mean optimal window for the group is shown in black: 95–360 ms and 4.1–18.3 Hz. This aggregate time/frequency window is used in all subsequent analyses, unless specified otherwise. Individual time/frequency spaces for each participant are included in the Supplementary materials (Fig. S.3).

4.2. Individual classification

For each participant, the 64-channel EEG data was band-filtered and epoched according to an average optimum window just described. Using a five-fold partition procedure, the 360 trials were split into five non-overlapping evaluation sets of equal size (72 trials, evenly split between each class). For each evaluation set, the remaining 4/5 of the data served as the training set (288 trials). Accuracy was then computed by comparing the SVM predictions of membership to the true category of the image presented in each evaluation trial, over all five folds, on the basis of the spectral power measures extracted (the distribution of features extracted for a typical participant are included in the Supplementary materials as Fig. S.1). Default SVM settings were used: a radius basis function kernel with settings of $1/n$ for the gamma value (where n is the number of training data points, 288) and 1 for the cost parameter C. [6] As Fig. 2 shows,

[6] Usually, optimisation of SVM parameters can lead to considerable increases in classification performance. In our case, several tuning strategies were investigated, but did not improve results. We believe that this is because CSP is a powerful supervised learning technique that largely succeeds in linearly separating the training data, and so a ceiling is reached in cross-validation performance at the SVM. Hence a simpler classifier, such as a linear discriminant approach, could be expected to achieve similar accuracies.

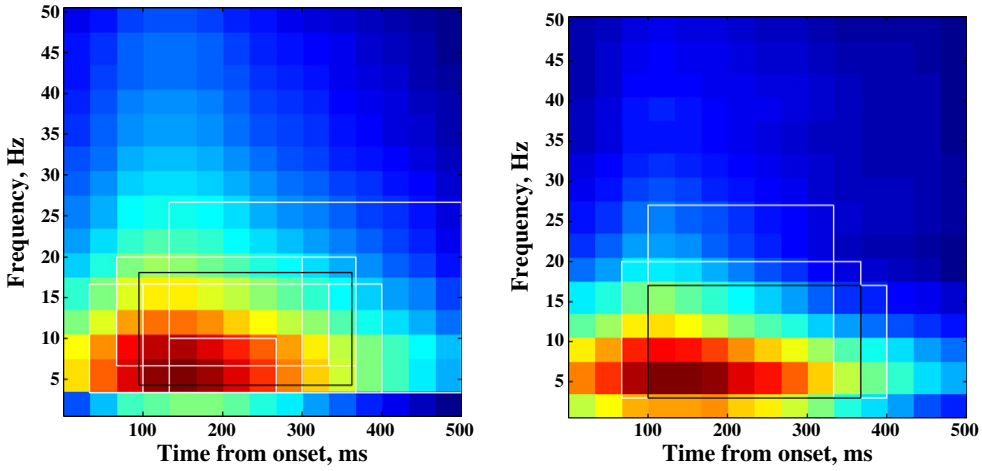

Fig. 1. Epoch category separability in time/frequency space of all participants (left panel), and a single participant (B, right panel). Blue to red scale indicates low to high informativity of a time/frequency region. The black box indicates the region identified as optimal for classification, and the white boxes several sub-optimal solutions.

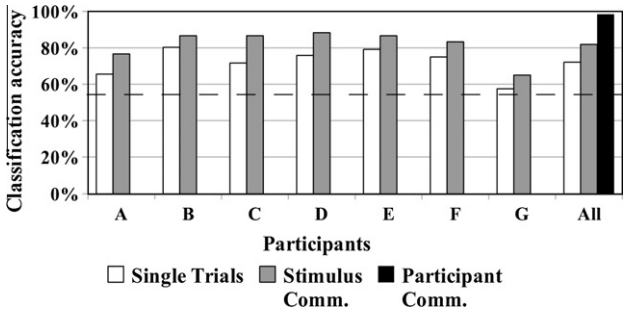

Fig. 2. Classification accuracy for mammals vs tools: on single trials; on stimulus committee classification over 6 presentations; and on committee classification over all participants (dotted line indicates performance significantly above chance on single trials).

classification accuracy over single trials was significant for all participants ($p < 0.01$) ranging from 57% to 80% (performance significantly above chance at $\alpha = 0.05$ is indicated by a dotted line), with a mean of 72% ($p < 0.001$). Significance was calculated using a binomial test ($n = 360$, $p = 0.5$), and validated with a bootstrapping simulation (based on 1000 random permutations of category labels). [7] To guard against the dangers of "double-dipping" (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Olivetti, Mognon, Greiner, & Avesani, 2010; Pereira, Mitchell, & Botvinick, 2009), the classification was repeated with the windowing parameters for each participant being selected on the basis of the optimization data from the other six participants only, ensuring that there was no overlap between test and training data. While there were minor differences in the classification accuracy for individual participants (up to 2% points), all were still significant, and the average performance remained unchanged at 72%, indicating that the use of a uniform time/frequency window for all analyses was not leading to spurious inflation of the accuracy estimates.

The CSP components extracted can be visualised by means of scalp-maps to give an indication of the distribution of the neural sub-assemblies that are informative for category. The precise distribution varied across participants, but a wide range of occipital, parietal and frontal areas played a role. The scalp maps from participant C, shown in Fig. 3 (top panel), are typical, showing activation distributed over all these areas. The mean activation maps over all participants are shown in the lower panel. Individual maps for all participants are available in the Supplementary materials (Fig. S.3).

4.3. Group classification

As discussed above, there are some differences between participants in the optimal time/frequency windows found, in the mammal and tool-specific spatial components extracted, and in the final categorisation accuracy achieved. This raises the question of to what extent category-specific neural activities are shared across participants. Cross-participant generalisation in EEG work is complicated by many factors: the signal to noise ratio at each electrode is affected by the impedance achieved on the contact to the scalp, which is influenced by local differences in skin condition; the spatial location of electrodes relative to underlying cortex will vary according to the size and shape of the head; and as is also the case with other imaging technologies, there may be individual differences in brain anatomy or functional

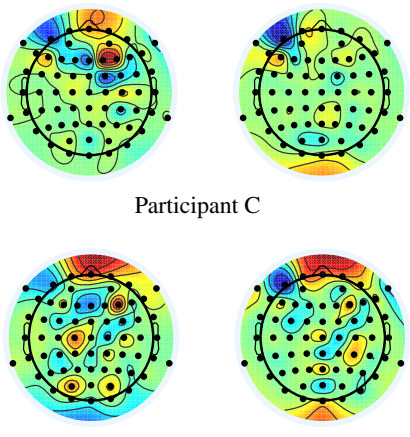

Fig. 3. Spatial components extracted for individual analyses (left scalp-map: mammal-sensitive spectral power; right scalp-map: tool-sensitive spectral power). Green indicates scalp areas that do not contribute to a component. Blue-scale colours indicate a negative weighting and red-scale colours a positive weighting. The black dots indicate the electrode positions.

[7] The distribution assumed by the binomial test had a mean and standard deviation over trials of 180 and 9.487 respectively. The permutation-based simulation yielded a mean and standard deviation of 179.9 and 9.58 trials.

localisation across participants. While Brain–Computer-Interaction tasks have demonstrated that generalisation is possible for the activity produced during motor planning and execution (see e.g. Curran & Stokes, 2003), it has not been established that activities associated with higher cognitive states generalise similarly well across participants.

The group analysis was carried out in a similar way to the individual analyses, using the same aggregate time/frequency window. Since the signal to noise ratio varies across channels and across participants, all signals were normalised to *z*-score values (i.e. each channel from each participant was individually transformed so that it had a mean value of 0 and a standard deviation of 1). As before, the data was band-filtered and epoched, before being partitioned into evaluation and training sets. Here, a seven-fold partition procedure was used, where in each case the data from six participants (2160 trials) was used for training and the data from the left-out participant (360 trials) was used for evaluation. Accuracy was then computed by comparing the SVM predictions of category membership to the category of the image presented in each evaluation trial.

The results are illustrated in Fig. 4. Classification accuracy was significant for all participant sessions ($p < 0.01$), ranging from 56% to 68%. The average accuracy was 61%. The components isolated for group classification (Fig. 5) overlap considerably with those for individual classification (Fig. 3), but the information gained from frontal areas is reduced, and that from occipital areas is increased.

4.4. Discounting nuisance variables

An obvious question about these results is whether they represent a successful detection of neural activity specific to semantic category, or they are rather driven by some lower level confound such as word length, ease of naming, or image complexity. Activity in the frequency bands identified here have been found to be modulated by a wide variety of cognitive tasks in verbal, visual and spatial processing (see reviews by Kahana, 2006; Pfurtscheller & Lopes da Silva, 1999). Our stimuli were chosen to be reasonably frequent, visually identifiable and easily nameable, while constituting plausible categories that people use in everyday life and which are commonly lexicalised in human languages. No attempt was made to balance the materials for the many conceptual, visual and lexical confounds that have been raised in the literature, as we did not consider it feasible to find a set of stimuli that were reasonably familiar, while also having comparable distributions on the values of all potential confounds. Instead, the trial-by-trial predictions made by the algorithm allow the contribution of any confound to be evaluated systematically at the analysis stage, even if that confound was not considered in the experimental.

An initial answer to the question of confounds is provided by the distribution over the scalp of the informative signal components. On average the depictions of mammals occupied a larger proportion of the screen display area and are more detailed, and so one possible alternative explanation of the results would be that we are detecting the additional perceptual processing that they demand. Cursory examination of the individual-level scalp-maps (Fig. S.3, Supplementary materials) shows that the informative activity is predominantly outside of the occipital areas where visual processing is focused, making this unlikely. But for the group analyses (Figure S.4) it remains a possible explanation.

A further test is to look for patterns of errors in the classification. If the same items are repeatedly misclassified by the algorithm, it may be some confounding characteristic of the stimulus that is responsible. For example, if the analysis had identified components of neural activity that are sensitive to lexical length, and was using this to guess category membership (on average the mammal words were somewhat shorter than the tool words — see Table 1 below), then a mammal with a longer name (such as *ippopotamo* 'hippopotamus') should be consistently grouped with the tools, and a tool with a shorter name (e.g. *sega* 'saw') should be grouped with the mammals.

If on the other hand the classification errors are simply due to random noise, we would expect classification of stimuli to be more reliable after amalgamating predictions over multiple presentations. Here committee predictions are made by taking the majority classification over several trials (e.g. if a single stimulus was classified as a mammal on two presentations and as a tool on four presentations, the committee prediction is a tool). Making committee predictions separately for each participant does increase accuracy, represented by the grey bars on Fig. 2. In all participants there is a noticeable improvement in accuracy, and the mean rate rises to 80% (grey bar on right). Similarly, taking the majority classification for each stimulus over all presentations to all participants raises classification accuracy to 98% (black bar on the far right), indicating that no single confound can account for the overall behaviour of the algorithm.

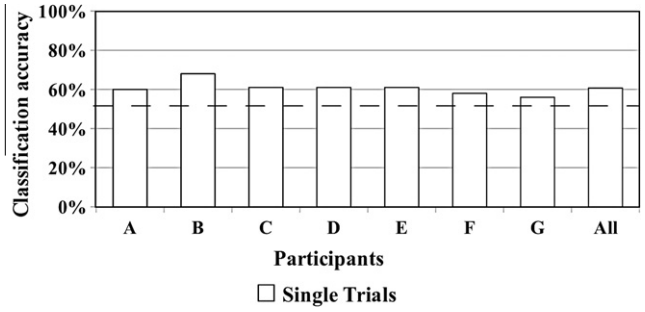

Fig. 4. Classification accuracy for mammals vs tools training across participants, single trials (dotted line indicates performance significantly above chance).

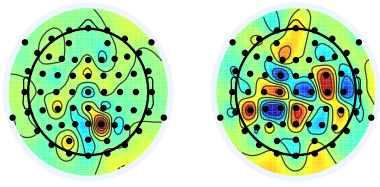

Fig. 5. Mean spatial components extracted for group analysis (left scalp-map: mammal-sensitive spectral power; right scalp-map: tool-sensitive spectral power). Green indicates scalp areas that do not contribute to a component. Blue-scale colours indicate a negative weighting and red-scale colours a positive weighting. The black dots indicate the electrode positions.

Table 1
Confounding of stimulus properties with category.

Variable	Imbalance p	True category ρ	Predicted category ρ
Category			0.90
Image brightness	<0.001***	0.73	0.72
Image detail	0.042*	0.21	0.20
Image complexity	<0.001***	−0.80	−0.78
Word length	0.105	0.12	0.10
Word frequency	0.192	−0.18	−0.20
Reaction time	0.101	−0.28	−0.24
Naming consistency	0.759	0.02	0.02
Null responses	0.055	−0.17	−0.14

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$

However, this does not exclude the possibility that some combination of confounding factors is being exploited by the algorithm. To investigate this, a regression model can be used to determine the relative contributions of a set of variables to the output of interest. In this case the output (the dependent variable, which we are trying to explain) is the committee category prediction for each stimulus, averaged over all presentations to all participants. The possible input variables (independent variables) are the true stimulus category, and a range of confounding variables related to the stimulus images, the words they represent, and task performance.

The lexical variables considered were word length in syllables, and lexical log frequency calculated using a large web-collected corpus (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). [8] The possible visual confounds examined were image brightness, mean spatial frequency (a measure of low-level detail), and visual complexity (calculated using GIF and JPEG image compression – see Forsythe, Mulhern, & Sawey, 2008; Székely & Bates, 2000). Finally the three metrics of participant performance were naming consistency, reaction time, and number of null reactions (i.e. how often a time-out was reached, because a participant could not think of an appropriate word within three seconds). Table 1 lists the variables, together with several measures of the degree of balance across the categories: the first column of numbers gives the probability that this variable is balanced across the two groups of stimuli, using a *t*-test; the second and third columns give respectively the rank correlation of each variable with the true category of the stimuli, and the linear correlation to the committee prediction produced by the algorithm. The correlations were performed with the category of tools coded as 1 and mammals as 0 – so positive correlations indicate a larger value in the confound for tools, and a negative correlation indicates a larger value in the confound for mammals (e.g. the tool words were longer, and the mammal words were more frequent).

A least-squares linear regression model was then constructed using all these variables. [9] All variables were *z*-score normed, so that the weights assigned to them by the regression model would be comparable. Care must be taken when selecting the inputs for such a model, as closely correlated variables can make the model unstable. For this reason image complexity was excluded, since it correlated very highly with image brightness ($r = 0.92$). [10]

The full model containing the true category value and all these possible confounds explained a very high amount of the variance ($r^2 = 0.81$) in the predictions made in individual analyses. The relative contributions of each variable can be seen in Table 2. The first column shows that only two variables – the true stimulus category and image brightness – reach significance in explaining the behaviour of the algorithm in individual analyses. A reduced version of the model, where superfluous variables are automatically removed step-wise to improve overall fit, settles on these same two variables (figures in brackets). [11] These weights suggest that category has over seven times the explanatory power of any of the confounds.

A second linear model to explain the group analysis gave similar results (second column in Table 2). Despite a lower overall fit ($r^2 = 0.63$), it again showed that category had many times more explanatory power than any of the confounds, which in this case reduced to image detail (based on spatial frequencies) and word length (though this variable's contribution was not significant).

Table 2
Explanatory power of category and confounds.

Variable	Individual analysis, β	Group analysis, β
Category	1.60 *** (1.51***)	1.60*** (1.50***)
Image brightness	−0.17* (−0.19**)	0.04
Image detail	−0.01	0.22* (0.22*)
Word length	−0.10	−0.14 (0.08)
Word frequency	−0.07	0.03
Reaction time	0.09	0.03
Naming consistency	0.02	0.09
Null responses	−0.01	0.06

The beta values indicate the weight given by the model to each input variable, and so its explanatory power. The figures in brackets show the weight assigned after variable pruning with step reduction.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Finally, manual variations on these models can be built to compare the variance explained by subsets of variables. A model using all seven potential confounds but excluding the category of the stimuli explains only 51% and 34% of variance (for individual and group analyses respectively). Models using significant confounds only (image brightness for the individual model; and image detail for the group model) explained 51% and 15%. A model using the true category alone accounts for 80% and 61% of the variance respectively – explaining more of the algorithm's predictive behaviour than any of the confounds, individually or in combination.

All this evidence indicates that the category of the stimuli provides the best explanation for the behaviour of the algorithm. While the confounds can also model some of the algorithm's behaviour, this can be explained by their partial correlations with semantic class. The comparison of models with ($r^2 = 0.81$ and 0.63) and without ($r^2 = 0.80$ and 0.61) confounds demonstrates that the information supplied by confounds is almost entirely superfluous.

5. Discussion

This is the first work to report the decoding of semantic categories from neural activity using EEG data, and also to take advantage of single-trial analyses to systematically exclude a range of low-level confounds as alternative explanations. Comparison with similar work using imaging data (Haxby et al., 2001, who achieves 85–100% accuracy on a larger number of categories) suggests that fMRI signals may be more information-rich in this respect, but we hope that continuing work on improving these techniques may close that gap. As EEG studies can be performed at much lower cost than fMRI, they may be a more feasible methodology for larger-scale investigations of the lexicon. We are now replicating these experiments with an expanded set of materials to establish the precise nature of the semantic distinction being detected, and with MEG to see to what extent it can improve individual and group-level prediction (see Murphy & Poesio, 2010 for a preliminary report on this work).

The methods described in this paper show clear differences in the pattern of EEG signals observed during the processing of superordinate categories. However, the trained algorithm did not prove to be sensitive to conceptual relatedness within categories – for example that *chimpanzee* is more similar to *gorilla* than either are to *camel*, and that *garden fork* and *shovel* are more closely related than they are to *chainsaw*. Ideally one could find correspondences of the semantic space described by this technique (see Fig. S.2 in the Supplementary materials for a graphical representation) with

[8] Conceptual variables were not available for the Italian words used in the study, but concreteness should not vary substantially across these exclusively nominal object stimuli, and imagery should play no role in a picture naming task. Familiarity may play a role, and is indirectly reflected in the model as it co-varies strongly with lexical frequency (e.g. in the MRC Psycholinguistic Database, Wilson, 1988, there is a 0.74 rank correlation between familiarity and frequency).

[9] Using the lm() function in R.

[10] The linear models described below were replicated using image complexity in place of image brightness, and gave very similar results, with marginally lower measures of model fit.

[11] The step() function in R derives an optimal linear model by using the Akaike information criteria to incrementally remove uninformative variables.

those derived from informants' judgements of typicality or semantic similarity (e.g. DeDeyne et al., 2008; Morrow & Duffy, 2005), or with analogous measures of relatedness derived from language corpora and lexica (e.g. Lund et al., 1995; Lin, 1998; Patwardhan, Banerjee, & Pedersen, 2003), which are of particular interest to us – but none of several comparisons of this kind that we carried out yielded significant results. This indicates that the discriminative classification techniques used are working as they are designed to, optimising the class distinction while ignoring all other dimensions of variation in the stimuli. We are currently developing other data mining techniques that are better suited to extracting intra-categorial structure also (see Murphy, Baroni, & Poesio, 2009 for an example of alternative feature extraction methods). Nevertheless, these trial-by-trial techniques make it possible for the first time to systematically exclude suspected confound variables (without exhaustive balancing of authentic materials, or the use of tailored fillers), and to potentially distinguish between the various semantic interpretations that can be attached to the group analyses reported in many neuroimaging and electrophysiological studies.

Currently we achieve highly significant classification for single participant data, and while generalising across individuals of a single language group, but we do not yet know if the neural patterns identified would generalise to other materials. Some preliminary analyses were also carried out based on earlier smaller scale experiments that used more heterogeneous materials (100 trials of mixed animals and mixed artefacts), to test if cross-language decoding (training on data from Italian speakers and evaluating on data from English speakers) and cross-modal decoding (training on data from a silent image naming task and evaluating on data from a lexical visualisation task) is possible. While encouraging, these results only approached significance, and need to be replicated on a larger scale. In particular, we intend to use visual, rather than auditory presentation of lexical stimuli, to make cross-trial synchronisation more reliable (it is not straight-forward to determine at what point during each auditory word enough information has been received to allow categorisation to proceed). Lexical stimuli will also allow us to investigate more abstract categories than those that can be easily and unambiguously represented with images. Finally, it should be noted that the analysis presented here may represent either the process of categorisation, or the resultant representations, or both – though further analysis of the time-course of categorisation together with examination of later latencies may provide insights. Similarly with the current experimental design we cannot be conclusive about what phases of the image naming process (including object recognition and lexical retrieval) our algorithms take advantage of.

In terms of the neural realisation of semantics, we have identified scalp locations, time intervals and frequency bands that are especially informative about categorial differences. There is still no well-established consensus on the functional interpretation of variations in spectral power (see Kahana, 2006; Pfurtscheller & Lopes da Silva, 1999). The localised increases in spectral power seen here are conventionally interpreted as a reflection of a decrease in neural activity in some frequency bands ('cortical idling', in the alpha-band ≈8–12 Hz), and a reflection of an increase in activity in others (theta, 4–8 Hz; gamma, > 20 Hz), while the magnitude of the change in power is seen to be modulated by task difficulty or complexity (Jensen, Kaiser, & Lachaux, 2007). More specifically, Pfurtscheller and Lopes da Silva (1999) describe upper alpha activity (10–12 Hz) as being linked to sensory and semantic processing, particularly in parietal-occipital areas, like those seen in this study. And Kahana (2006) report theta band activity (4–8 Hz, also seen here) for verbal working memory tasks. The lowest frequencies (<4 Hz) may also reflect ERP components with a period of more than 250 ms, such as N1/N170 effects.

Further, the latencies identified by our algorithms may seem rather early when compared to some of the ERP literature: the signals seen by Kiefer (2001) and Paz-Caballero et al. (2006) peak in separability after 300 ms, and any earlier differences are interpreted as being due to perceptual processing. However, the time interval we identify (≈100–350 ms) is more consistent with recent work that suggests that at least a first pass, coarse categorisation is carried out much earlier (see Thorpe, 2009, for discussion): Meeren, Hadjikhani, Ahlfors, Hämäläinen, and De Gelder (2008) found high-level category specific effects in MEG signals, both within visual processing centres and elsewhere, 70-100ms after image onset; and Liu, Agam, Madsen, and Kreiman (2009) detected view-invariant differences across categories in single trials as early as 100 ms after onset with intercranial electrodes. Further, visual input can reach the frontal lobe as early as 40–65 ms after visual onset (Kirchner, Barbeau, Thorpe, Regis, & Liegeois-Chauvel, 2009), and participants can make an eye-movement as early as 120 ms after onset in a category detection task (Kirchner & Thorpe, 2006), meaning that information on category is indeed available to higher cognition at these early latencies. In psycholinguistics, shorter latencies for the semantic processing of words are well established: in a visual-world task Allopenna, Magnuson, and Tanenhaus (1998) show that eye-movements that depend on interpretation can begin as early as 200 ms after the auditory onset of a word, while the effect of semantics during lexical processing becomes apparent in EEG between 100 and 200 ms (Pulvermüller, 2002, p. 62).

The similarities to the time intervals (Rossion et al., 2003, 2004; Rousselet et al., 2004) and spectral frequencies (Rousselet et al., 2007) seen in the face-processing literature are particularly striking. [12] As discussed in the background section, it is unlikely that these effects are exclusively due to perceptual processing (they are seen when faces are compared to non-face stimuli that have been matched for low-level visual properties; and are also found for non-human animate stimuli). So while we cannot rule out that we are also detecting high-level visual categorisation processes, these *follow* initial categorisation, and so are at a level of abstraction (e.g. having handles vs having legs; being animate/inanimate) where visual and linguistic descriptions are closely aligned, if at all separable. This conclusion stands, irrespective of ones assumptions as to whether the conceptual system has separate visual and linguistic modules or not. In this respect it is important that the concepts we have used as stimuli are highly imageable, as the linguistic or psychological features that are most salient (e.g. according to feature norms such as McRae et al., 2005, or as might be yielded by corpus-based models of semantics like Baroni et al., 2010; Poesio & Almuhareb, 2008), are predominantly ones that are readily visualisable (colour, shape, size of ears, number of legs, etc).

The scalp locations at which these differences are maximal point to brain regions in which the processing load is modulated by semantic category. It is difficult to draw precise anatomical conclusions on the basis of this data as there is no principled basis on which to decide on the number of dipoles to expect in a source localisation analysis, nor whether the point-sources such a model assumes are appropriate when interpreting distributed activations. But it is clear that semantic processing and the resulting representations are widely spread across the brain, in a fashion that is somewhat shared between the participants that took part in the study. In particular, the extensive dorsal activations seen can be interpreted as an effect of motor planning induced by the visual processing of manipulable objects (Goodale & Milner, 1992; Mishkin & Ungerleider, 1982), or rather a reflection of conceptual

[12] We would like to thank the second reviewer for bringing these studies to our attention.

representations proper, as have been suggested more recently both in the literature on visual processing (Almeida, Mahon, Nakayama, & Caramazza, 2008) and lexical comprehension (Pulvermüller, 2005). In the case of the second interpretation, this constitutes further support for theories of neuronal networks of word meanings that arise from our experience of the world and language (Barsalou, 2003; Spitzer, 1999; Tyler & Moss, 2001).

Acknowledgments

We are very grateful to Lisandro Kaunitz, Francesco Vespignani, Laura Tonelli and Stefano Bertamini for assistance in data collection and analysis. We would also like to thank the three reviewers who gave very helpful observations and suggestions on the earlier draft of this paper. The work described here was funded by CIMeC, the Autonomous Province of Trento, the Fondazione Cassa Risparmio Trento e Rovereto, and a University of Essex Research Incentive Fund grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bandl.2010.09.013.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419–439.

Almeida, J., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2008). Unconscious processing dissociates along categorical lines. *Proceedings of the National Academy of Sciences of the United States of America, 105*(39), 15214–15218.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation, 43*(3), 209–226.

Baroni, M., Murphy, B., Barbu, E., & Poesio, M. (2010). Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science, 34*(2), 222–254.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*, 513–562.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by probability distributions. *Bulletin of the Calcutta Mathematical Society, 35*, 99–109.

Bloom, P. (2002). *How children learn the meaning of words*. Cambridge: MIT Press.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In D. Haussler (Ed.), *5th Annual ACM workshop on COLT* (pp. 144–152). Pittsburgh: ACM Press.

Bruzzone, L., Roli, F., & Serpico, S. B. (1995). An extension of the Jeffreys–Matusita distance to multiclass cases for feature selection. *IEEE Transactions on Geoscience and Remote Sensing, 33*(6), 1318–1321.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences, 7*, 354–361.

Caramazza, A., & Shelton, J. (1998). Domain-specific knowledge systems in the brain: The animate–inanimate distinction. *Journal of Cognitive Neuroscience, 10*, 1–34.

Chang, C.-C., & Lin, C.-J. (2001). *LIBSVM: A library for support vector machines*.

Chao, L., Haxby, J., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*, 913–919.

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category related cortical activity. *Cerebral Cortex, 12*, 545–551.

Collins, A. M., & Quillian, M. R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference. In A. F. Sanders (Ed.), *Attention and performance III. Acta psychologica* (Vol. 33, pp. 304–314). Amsterdam: North-Holland.

Connolly, A., & Haxby, J. (2010). Similarity-based multi-voxel pattern analysis reveals an emergent taxonomy of animal species along the object vision pathway. *Journal of Vision, 10*(7), 964.

Curran, E. A., & Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain and Cognition, 51*(3), 326–336.

Dalponte, M., Bovolo, F., & Bruzzone, L. (2007). Automatic selection of frequency and time intervals for classification of EEG signals. *Electronics Letters, 43*, 1406–1408.

DeDeyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., et al. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods, 40*(4), 1030–1048.

Delorme, A., & Makeig, S. (2003). EEGLAB: An open source toolbox for analysis of single-trial dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.

Fillmore, C.J. (1982). Frame semantics. In L.S. Korea (Ed.), *Linguistics in the morning calm*. Hanshin, Seoul (pp. 111–138).

Forsythe, A., Mulhern, G., & Sawey, M. (2008). Confounds in pictorial sets: The role of complexity and familiarity in basic-level picture processing. *Behaviour Research Methods, 40*(1), 116–129.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge: MIT Press.

Gilbert, J., Shapiro, L., & Barnes, G. (2009). Processing of living and nonliving objects diverges in the visual processing system: Evidence from MEG. In *Proceedings of the cognitive neuroscience society annual meeting*.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15*, 20–25.

Haxby, J., Gobbini, M., Furey, M., Ishai, A., Schouten, J., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425–2430.

Itier, R., & Taylor, M. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14*(2), 132.

Jensen, O., Kaiser, J., & Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences, 30*(7), 317–324.

Kahana, M. J. (2006). The cognitive correlates of human brain oscillations. *The Journal of Neuroscience, 26*, 1669–1672.

Kamp, H., & Partee, B. (1995). Prototype theory and compositionality. *Cognition, 57*(2), 129–191.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects in object categorization: Event-related potentials during picture and word categorization. *Memory and Cognition, 29*(1), 100–116.

Kirchner, H., Barbeau, E., Thorpe, S., Regis, J., & Liegeois-Chauvel, C. (2009). Ultra-rapid sensory responses in the human frontal eye field region. *Journal of Neuroscience, 29*(23), 7599–7606.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46*(11), 1762–1776.

Koles, Z. J., Lazar, M. S., & Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain Topography, 2*(4), 275–284.

Kriegeskorte, N., Simmons, W., Bellgowan, P., & Baker, C. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience, 12*(5), 535–540.

Landauer, T., & Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of COLING-ACL* (pp. 768–774).

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron, 62*(2), 281–290.

Lund, K., Burgess, C., & Atchley, R. (1995). Semantic and associative priming in high dimensional semantic space. In *Proceedings of the 17th cognitive science society meeting* (pp. 660–665).

Maguire, M., Brier, M., & Ferree, T. (2010). EEG theta and alpha responses reveal qualitative differences in processing taxonomic versus thematic semantic relationships. *Brain and Language, 114*, 16–25.

Mahon, B., Anzellotti, S., Schwarzbach, J., Zampini, M., & Caramazza, A. (2009). Category-specific organization in the human brain does not depend on visual experience. *Neuron, 63*(3), 397–405.

Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems* (Vol. 8, pp. 145–151). MIT Press.

Mandler, J. M. (1992). How to build a baby I: Conceptual primitives. *Psychological Review, 99*, 587–604.

Martin, A., & Chao, L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinions in Neurobiology, 11*, 194–201.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379*, 649–652.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*, 310–322.

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, Instruments, & Computers, 37*(4), 547–559.

Meeren, H., Hadjikhani, N., Ahlfors, S., Hämäläinen, M., & De Gelder, B. (2008). Early category-specific cortical activation revealed by visual stimulus inversion. *PLoS One, 3*(10).

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioral Brain Research, 6*(1), 57–77.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., et al. (2008). Predicting human brain activity associated with the meanings of nouns. *Science, 320*, 1191–1195.

Model, D., & Zibulevsky, M. (2006). Learning subject-specific spatial and temporal filters for single-trial EEG classification. *NeuroImage, 32*(4), 1631–1641.

Morrow, L. I., & Duffy, M. F. (2005). The representation of ontological category concepts as affected by ageing: Normative data and theoretical implications. *Behavior Research Methods, 37*(4), 608–625.

Murphy, B., Baroni, M., & Poesio, M. (2009). EEG responds to conceptual stimuli and corpus semantics. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 619–627). The Association for Computational Linguistics.

Murphy, B., & Poesio, M. (2010). Detecting semantic category in simultaneous EEG/MEG recordings. In *First workshop on computational neurolinguistics, NAACL HLT 2010* (pp. 36–44). Los Angeles: Association for Computational Linguistics.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Mahwah: Erlbaum.

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology, 24*, 1565–1567.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424–430.

Olivetti, E., Mognon, A., Greiner, S., & Avesani, P. (2010). Brain decoding: Biases in error estimation. In *Proceedings of the 1st ICPR workshop on brain decoding. 20th International conference on pattern recognition*. Istanbul: IEEE Computer Science Society.

Padó, S., & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics, 33*(2), 161–199.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage, 28*, 326–341.

Patwardhan, S., Banerjee, S., & Pedersen, T. (2003). Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the fourth international conference on intelligent text processing and computational linguistics* (pp. 241–257). Mexico City.

Paz-Caballero, D., Cuetos, F., & Dobarro, A. (2006). Electrophysiological evidence for a natural/artifactual dissociation. *Brain Research, 1067*(1), 189–200.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S. F., & Fazio, F. (1999). Word and picture matching: A PET study of semantic category effects. *Neuropsychologia, 37*(3), 293–306.

Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage, 45*(1), S199–S209.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology, 110*, 1842–1857.

Philiastides, M., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience, 26*(35), 8965.

Plaut, D. (1995). Semantic and associative priming in a distributed attractor network. In *Proceedings of the seventeenth annual conference of the cognitive science society* (pp. 37–42).

Poesio, M., & Almuhareb, A. (2005). Identifying concept attributes using a classifier. In *Proceedings of the ACL workshop on deep lexical semantics* (pp. 18–27).

Poesio, M., & Almuhareb, A. (2008). Extracting concept descriptions from the Web: The importance of attributes and values. In P. Buitelaar & P. Cimiano (Eds.), *Bridging the gap between text and knowledge* (pp. 29–44). Amsterdam: IOS.

Pulvermüller, F. (2002). *The neuroscience of language: On brain circuits of words and serial order*. Cambridge: Cambridge University Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*, 576–582.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge: MIT Press.

Quillian, M. R. (1967). Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science, 12*(5), 410–430.

Ramoser, H., Gerking, J. M., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering, 8*(4), 441–446.

Rapp, R. (2004). A freely available automatically generated thesaurus of related words. In *Proceedings of LREC 2004* (pp. 395–398).

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 7*, 328–350.

Rossion, B., Joyce, C., Cottrell, G., & Tarr, M. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage, 20*(3), 1609–1624.

Rousselet, G., Husk, J., Bennett, P., & Sekuler, A. (2007). Single-trial EEG dynamics of object and face visual processing. *Neuroimage, 36*(3), 843–862.

Rousselet, G., Mace, M., & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision, 4*(1), 13–21.

Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., et al. (2008). Automatic processing of semantic relations in fMRI: Neural activation during semantic priming of taxonomic and thematic categories. *Brain Research, 1218*, 194–205.

Shinkareva, S., Mason, R., Malave, V., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PloS One, 3*(1).

Simons, J., Koutstaal, W., Prince, S., Wagner, A., & Schacter, D. (2003). Neural mechanisms of visual object priming: Evidence for perceptual and semantic distinctions in fusiform cortex. *Neuroimage, 19*(3), 613–626.

Siri, S., Tettamanti, M., Cappa, S., Rosa, P., Saccuman, C., Scifo, P., et al. (2007). The neural substrate of naming events: Effects of processing demands but not of grammatical class. *Cerebral Cortex, 18*(1), 171–177.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214–241.

Spitzer, M. (1999). *The mind within the net: Models of learning, thinking and acting*. Cambridge: MIT Press.

Székely, A., & Bates, E. (2000). Objective visual complexity as a variable in studies of picture naming. Tech. Rep. 12-2. Center for Research in Language Newsletter, University of California, San Diego.

Tallon-Baudry, C., & Bertand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Science, 3*(4), 151–162.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567.

Thorpe, S. J. (2009). The speed of categorization in the human visual system. *Neuron, 62*(2), 168–170.

Tyler, L. K., & Moss, H. E. (2001). Towards a distributed account of conceptual knowledge. *Trends in Cognitive Sciences, 5*(6), 244–252.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

Vinson, D., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*(1), 183–190.

Vuilleumier, P., Henson, R., Driver, J., & Dolan, R. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neuroscience, 5*(5), 491–499.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*(3), 829–853.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers, 20*(1), 6–11.