Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data

Tijl Grootswagers[1,2,3], Susan G. Wardle[1,2], and Thomas A. Carlson[2,3]

Abstract

■ Multivariate pattern analysis (MVPA) or brain decoding methods have become standard practice in analyzing fMRI data. Although decoding methods have been extensively applied in brain–computer interfaces, these methods have only recently been applied to time series neuroimaging data such as MEG and EEG to address experimental questions in cognitive neuroscience. In a tutorial style review, we describe a broad set of options to inform future time series decoding studies from a cognitive neuroscience perspective. Using example MEG data, we illustrate the effects that different options in the decoding analysis pipeline can have on experimental results where the aim is to "decode" different perceptual stimuli or cognitive states over time from dynamic brain activation patterns. We show that decisions made at both preprocessing (e.g., dimensionality reduction, subsampling, trial averaging) and decoding (e.g., classifier selection, cross-validation design) stages of the analysis can significantly affect the results. In addition to standard decoding, we describe extensions to MVPA for time-varying neuroimaging data including representational similarity analysis, temporal generalization, and the interpretation of classifier weight maps. Finally, we outline important caveats in the design and interpretation of time series decoding experiments. ■

INTRODUCTION

The application of "brain decoding" methods to the analysis of fMRI data has been highly influential over the past 15 years in the field of cognitive neuroscience (Kamitani & Tong, 2005; Carlson, Schrater, & He, 2003; Cox & Savoy, 2003; Haxby et al., 2001; Edelman, Grill-Spector, Kushnir, & Malach, 1998). In addition to their increased sensitivity, the introduction of fMRI decoding methods offered the possibility to address questions about information processing in the human brain, which have complemented traditional univariate analysis techniques. Although decoding methods for time series neuroimaging data such as MEG/EEG have been extensively applied in brain–computer interfaces (BCI; Müller et al., 2008; Curran & Stokes, 2003; Wolpaw, Birbaumer, McFarland, Pfurtscheller, & Vaughan, 2002; Kübler, Kotchoubey, Kaiser, Wolpaw, & Birbaumer, 2001; Farwell & Donchin, 1988; Vidal, 1973), they have only recently been applied in cognitive neuroscience (Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011; Duncan et al., 2010; Schaefer, Farquhar, Blokland, Sadakata, & Desain, 2010).

The goal of this article is to provide a tutorial style guide to the analysis of time series neuroimaging data for cognitive neuroscience experiments. Although introductions to BCI exist (Blankertz, Lemm, Treder, Haufe, & Müller, 2011; Lemm, Blankertz, Dickhaus, & Müller, 2011), the aims of time series decoding for cognitive neuroscience are distinct from those that drive the application of these methods in BCI, thus requiring a targeted introduction. Although there are many reviews and tutorials for fMRI decoding (Haynes, 2015; Schwarzkopf & Rees, 2011; Mur, Bandettini, & Kriegeskorte, 2009; Pereira, Mitchell, & Botvinick, 2009; Formisano, De Martino, & Valente, 2008; Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006; Cox & Savoy, 2003), there are no existing tutorial introductions to decoding time-varying brain activity. Although the approaches are conceptually similar, there are important distinctions that stem from fundamental differences in the nature of the neuroimaging data between fMRI and MEG/EEG. In this article, we provide a tutorial introduction using an example MEG data set. Although there are many possible analyses targeting time series data (e.g., oscillatory [Jafarpour, Horner, Fuentemilla, Penny, & Duzel, 2013] or induced responses), we restrict the scope of this article to decoding information from evoked responses, with statistical inference at the group level on single time points or small time windows. As with most neuroimaging analysis techniques, the number of possible permutations for a given set of analysis decisions is very large, and the particular choice of analysis pipeline is guided by the experimental question at hand. Here we aim to provide a broad demonstration of how the analysis may be approached, rather than prescribing a particular analysis pipeline.

[1] Macquarie University, Sydney, Australia, [2] ARC Centre of Excellence in Cognition and its Disorders, [3] University of Sydney

Journal of Cognitive Neuroscience 29:4, pp. 677–697
doi:10.1162/jocn_a_01068

Early studies using time-resolved decoding methods have revealed significant potential for experimental investigation using this approach with MEG/EEG (see MVPA for MEG/EEG section). However, compared with the popularity of decoding methods in fMRI, to date only a small number of studies have applied multivariate pattern analysis (MVPA) techniques to EEG or MEG. Accordingly, the aims of this article are to (a) introduce the critical differences between decoding time series (e.g., MEG/EEG) versus spatial (e.g., fMRI) neuroimaging data, (b) illustrate the time series decoding approach using a practical tutorial with example MEG data, (c) demonstrate the effect that selecting different analysis parameters has on the results, and (d) outline important caveats in the interpretation of time series decoding studies. In summary, this article will provide a broad overview of available methods to inform future time-resolved decoding studies. This tutorial is presented in the context of MEG; however, the methods and analysis principles generalize to other time-varying brain recording techniques (e.g., ECoG, EEG, electrophysiological recordings.). As this review is targeted at providing a broad overview to a general audience, we avoid formal mathematical definitions and implementation details of the methods and instead focus on the rationale behind the decoding approach as applied to time series data.

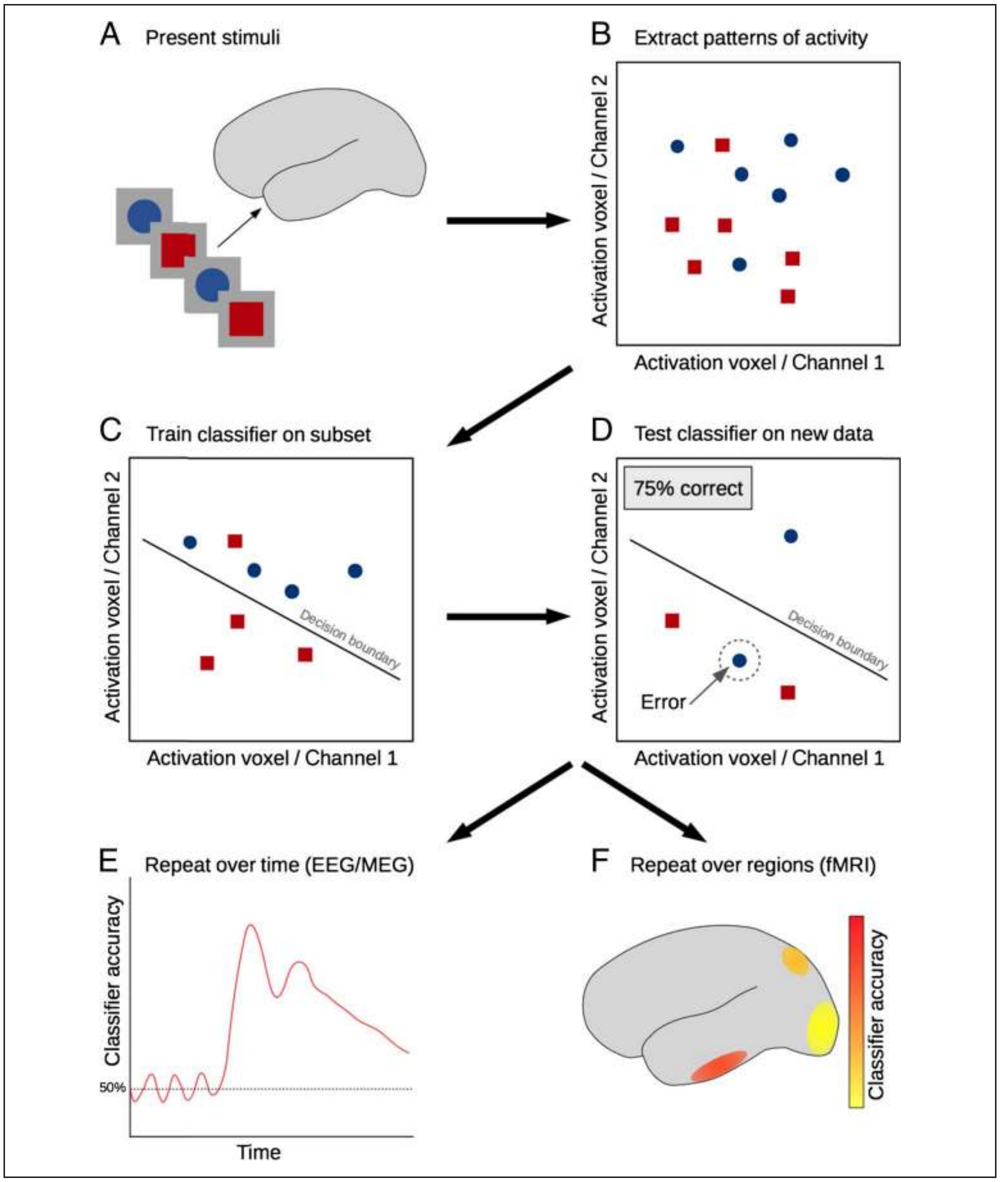
MVPA for MEG/EEG

The term "multivariate pattern analysis" (or MVPA) encompasses a diverse set of methods for analyzing neuroimaging data. The common element that unites these approaches is that they take into account the relationships between multiple variables (e.g., voxels in fMRI or channels in MEG/EEG), instead of treating them as independent and measuring relative activation strengths. The term "decoding" refers to the prediction of a model from the data ("encoding" approaches do the reverse, predicting the data from the model, reviewed in Naselaris, Kay, Nishimoto, & Gallant, 2011; see also, e.g., Ding & Simon, 2012, for an example of encoding models for MEG). The most common application of decoding in cognitive neuroscience is the use of machine learning classifiers (e.g., correlation classifiers (Haxby et al., 2001) or discriminant classifiers (Carlson et al., 2003; Cox & Savoy, 2003) to identify patterns in neuroimaging data, which correspond to the experimental task or stimulus. The most popular applications of MVPA are decoding (for recent reviews on fMRI decoding, see e.g., Haynes, 2015; Pereira et al., 2009) and, more recently, representational similarity analysis (RSA: Kriegeskorte & Kievit, 2013). Within the broad category of MVPA analyses, the central focus of this article is on decoding methods applied to evoked responses and the increasingly popular RSA framework (see Representational Similarity Analysis section).

The decoding approach is illustrated in Figure 1 for a simple experimental design in which the participant viewed pictures of blue circles or red squares while their brain activity was recorded. The goal of the decoding analysis is to test whether we can predict if the participant was viewing a blue circle or a red square based on their patterns of brain activation. If the experimental stimuli can be successfully "decoded" from the participant's patterns of brain activation, we can conclude that some information relevant to the experimental manipulation exists in the neuroimaging data. First, brain activation patterns in response to the different stimuli (or experimental conditions) are recorded using standard neuroimaging (MEG, fMRI, etc.) techniques (Figure 1A). The activation levels of the variables (e.g., voxels in fMRI, channels in MEG/EEG) in different experimental conditions are represented as complex patterns in high-dimensional space (each voxel, channel, or principal component is one dimension). For simplicity, in Figure 1B, these patterns are shown in two-dimensional space. Each point in the plot represents an experimental observation corresponding to the simultaneous activation level in two example voxels/channels in response to one of the experimental conditions (blue circles or red squares).

The first step in a decoding analysis involves training a classifier to associate brain activation patterns with the experimental conditions using a subset of the data (Figure 1C). In effect, during training the classifier finds the decision boundary in higher-dimensional space that best separates the patterns of brain activation corresponding to the two experimental categories into two distinct groups. As neuroimaging data are inherently noisy, this separation is not necessarily perfect (note the red square on the wrong side of the decision boundary in Figure 1C). Next, the trained classifier is used to predict the condition labels for new data that were not used for training the classifier (Figure 1D). The classifier predicts whether the new (unlabeled) data are more similar to the pattern of activation evoked by viewing a blue circle or a red square. If the classifier performs higher than that expected by chance (in this case 50% is the guessing rate as there are two stimuli), it provides evidence that the classifier can successfully generalize the learned associations to labeling new brain response patterns. Consequently, it is assumed that the patterns of brain activation contain information that distinguishes between the experimental conditions (i.e., the conditions blue circle/red square can be "decoded" from the neuroimaging data). Decoding accuracy can then be compared across brain regions (in fMRI) or time points (in MEG/EEG) to probe the location or time course of information processing in the brain. This is achieved by repeating the classification multiple times for different data, that is, different time points in MEG/EEG (Figure 1E) for examining the time course, or for different brain regions in fMRI (Figure 1F) for examining the spatial distribution of information in the brain. Thus, the main practical differences between decoding from MEG/EEG versus fMRI data lie in the methods used to obtain the patterns of information (Figure 1A, B) and the nature of the conclusions drawn from successful decoding performance (Figure 1E, F).

Figure 1. The general decoding approach. (A) Brain responses to stimuli (e.g., blue circles and red squares) are recorded with standard neuroimaging techniques. (B) Patterns of activation evoked by the two stimulus conditions (red square and blue circle) are represented in multiple dimensions (channels in EEG/MEG or voxels in fMRI); here only two dimensions are illustrated for simplicity. (C) A classifier is trained on a subset of the neuroimaging data, with the aim of distinguishing a reliable difference in the complex brain activation patterns associated with each stimulus class. (D) The performance of the classifier in distinguishing between the stimulus classes is evaluated by testing its predictions on independent neuroimaging data (not used in training) to obtain a measure of decoding accuracy. (E, F) Steps B–D may then be repeated for different time points (when using EEG/MEG) to study the temporal evolution of the decodable signal or repeated for different brain areas (in fMRI) to examine the spatial location of the decodable information.


A Present stimuli
B Extract patterns of activity
Activation voxel / Channel 2
Activation voxel / Channel 1
C Train classifier on subset
Activation voxel / Channel 2
Decision boundary
Activation voxel / Channel 1
D Test classifier on new data
75% correct
Activation voxel / Channel 2
Decision boundary
Error
Activation voxel / Channel 1
E Repeat over time (EEG/MEG)
Classifier accuracy
50%
Time
F Repeat over regions (fMRI)
Classifier accuracy

Decoding time series neuroimaging data is becoming increasingly popular. To date, most studies have applied the methods to understanding the temporal dynamics of the processing of visual stimuli and object categories. For example, time-resolved decoding has been used to study the emergence of object representations at the category and exemplar level using MEG (Carlson, Tovar, Alink, & Kriegeskorte, 2013), EEG (Cauchoix, Barragan-Jason, Serre, & Barbeau, 2014), and neuronal recordings (Zhang et al., 2011; Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Hung, Kreiman, Poggio, & DiCarlo, 2005); how invariant object representations emerge over time (Kaiser, Azzalini, & Peelen, 2016; Isik, Meyers, Leibo, & Poggio, 2014; Carlson et al., 2011); and how objects are represented in other (e.g., written or auditory) modalities (Simanova, van Gerven, Oostenveld, & Hagoort, 2010, 2015; Murphy et al., 2011; Chan, Halgren, Marinkovic, & Cash, 2010). Other studies have also used this approach to decode the orientation and spatial frequency of gratings from MEG (Wardle, Kriegeskorte, Grootswagers, Khaligh-Razavi, & Carlson, 2016; Cichy, Ramirez, & Pantazis, 2015; Ramkumar, Jas, Pannasch, Hari, & Parkkonen, 2013) and to study decision-making (Stokes et al., 2013; Bode et al., 2012), illusions (Hogendoorn, Verstraten, & Cavanagh, 2015), or working memory (Wolff, Ding, Myers, & Stokes, 2015; van Gerven et al., 2013). Notably, classifiers have been extensively applied to EEG (Guimaraes, Wong, Uy, Grosenick, & Suppes, 2007) for a different goal, as the low cost and portability of EEG is ideal for the development of BCI. These applications use classifiers to predict brain states to operate computers or robots (Müller et al., 2008; Allison, Wolpaw, & Wolpaw, 2007; Hill et al., 2006; Müller, Anderson, & Birch, 2003; Vidal, 1973, p. 2008). However, the goal of BCI is to achieve the maximum possible usability, that is, optimal prediction accuracy, robust real-time classification, and generalizability. The performance measures of BCI systems are therefore often compared across studies (and in competitions; see, e.g., Tangermann et al., 2012). This contrasts with decoding in neuroscience, where the goal is to understand brain processing by statistical inference on the availability of information (Hebart, Görgen, & Haynes, 2015), and accuracy differences between studies are generally not taken as meaningful.

Although the field is relatively new, there have already been several methodological extensions to standard decoding analysis applied to time series neuroimaging data (see Additional Analyses section). Following its application in fMRI, RSA (Kriegeskorte & Kievit, 2013) has been used with MEG data to correlate the temporal structure of brain representations with behavior (Wardle et al., 2016; Redcay & Carlson, 2015). RSA has also been used to link neuroimaging data from different modalities. For example, for object representations, the representational structure that appears early in the MEG data corresponds to representations in primary visual cortex measured with fMRI, whereas later stages instead reflect the representation in inferior temporal cortex (Cichy, Pantazis, & Oliva, 2014, 2016). A strength of time series decoding is that the dynamic evolution of brain representations can be examined. One example of this is the temporal generalization approach (see The Temporal Generalization Method section), which has been used in MEG to reveal that local and global responses to auditory novelty exhibit markedly different patterns of temporal generalization (King, Gramfort, Schurger, Naccache, & Dehaene, 2014). Furthermore, insights into the spatiotemporal dynamics can also be gained by combining source reconstruction methods with the decoding approach (van de Nieuwenhuijzen et al., 2013; Sudre et al., 2012) or by comparing the interaction between subsets of sensors (e.g., Goddard, Carlson, Dermody, & Woolgar, 2016). Thus, although relatively few time series neuroimaging studies to date have applied decoding methods, these have already provided valuable insights, illustrating the rich potential for future applications.

Recently, several toolboxes have been developed that implement the methods described in the rest of this paper; the PyMVPA toolbox (Hanke, Halchenko, Sederberg, Hanson, et al., 2009; www.pymvpa.org) handles both fMRI and M/EEG data using the open-source Python language (Hanke, Halchenko, Sederberg, Olivetti, et al., 2009), MNE (Gramfort et al., 2013, 2014; martinos.org/mne) is a Python toolbox (and can be accessed in MATLAB [The MathWorks, Natick, MA]) designed for M/EEG analyses, the Neural Decoding Toolbox (Meyers, 2013; www.readout.info) is a MATLAB toolbox created specifically for time-varying input, and the MATLAB toolbox CoSMoMVPA (Oosterhof, Connolly, & Haxby, 2016; www.cosmomvpa.org) handles both fMRI and M/EEG and was inspired by (and interfaces with) pyMVPA.

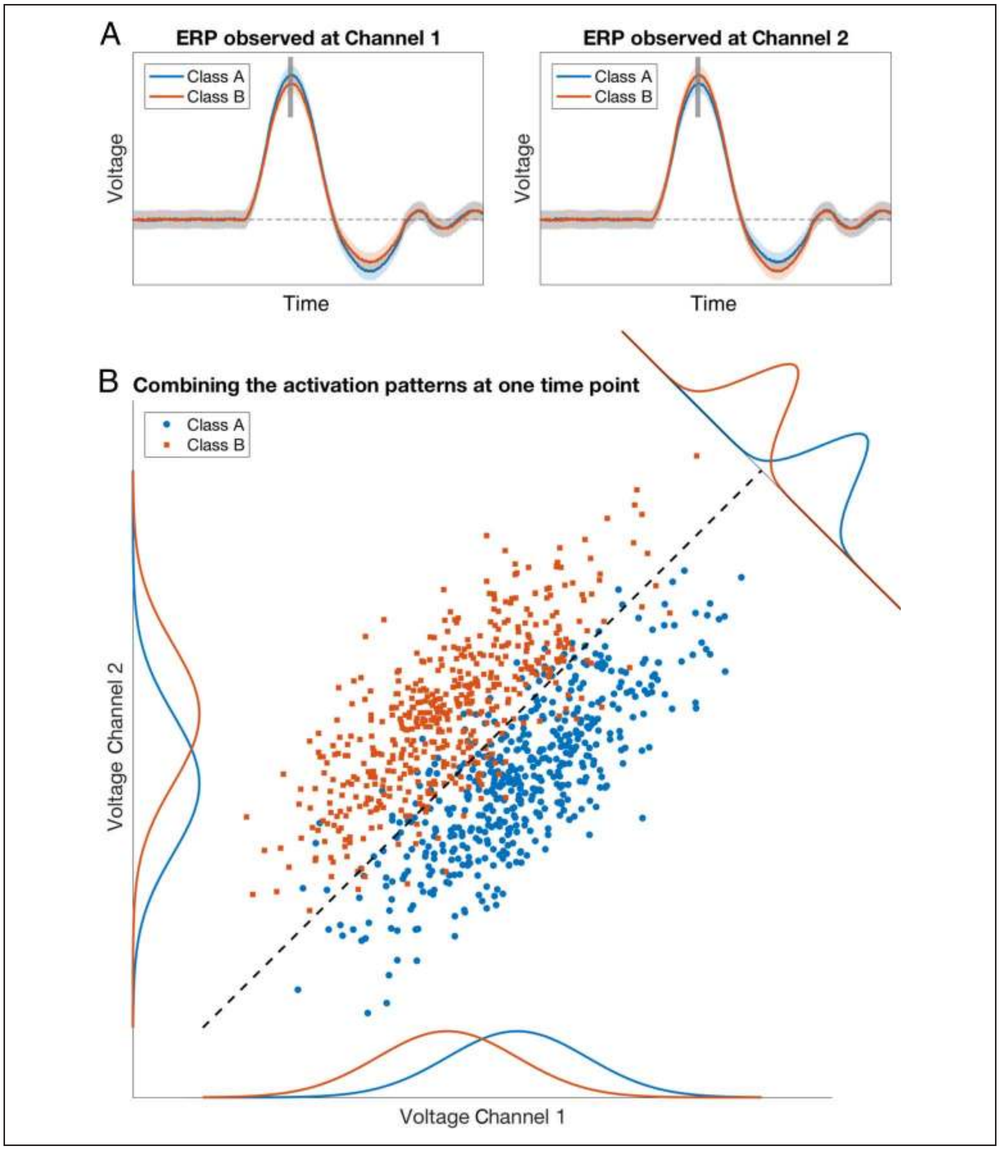
Decoding and other variants of MVPA are an alternative and complementary approach to univariate MEG/EEG analysis. This article will not cover univariate methods for MEG and EEG (which are well established; see, e.g., Cohen, 2014; Luck, 2005), and as always, the choice of analysis method must be guided by the experimental question. One of the central differences between univariate and multivariate methods is that the classifiers used in decoding approaches can use information that would not be detected when comparing the averaged signals in a univariate analysis (see Figure 2 for an illustration). This can lead to increased sensitivity for detecting differences between conditions (and on a single-trial basis). For example, decoding analysis can result in earlier detection of differences in the signals (Cauchoix et al., 2014; Cauchoix, Arslan, Fize, & Serre, 2012), and the differences found by classifiers can differ from those found in components (Ritchie, Tovar, & Carlson, 2015). Beyond sensitivity, the central distinction between univariate and MVPA analyses are the conceptual differences (activation-based vs. information-based) in the experimental questions each approach is suited to addressing. We anticipate that time series decoding approaches will continue to evolve alongside univariate methods, as has occurred with the adoption of decoding in fMRI, where both methods are used fruitfully.

The main aim of this article is to describe a typical analysis pipeline for decoding time series data in a tutorial format. The article is organized as follows. We begin by describing the experiment and the data recording procedures used to obtain the example MEG data (see Description of experiment section). Next, we illustrate how the recordings are preprocessed using a combination of principal component analysis (PCA), subsampling and averaging (see Preprocessing section). This is followed by the decoding analysis (see Decoding section). For all analysis stages, we provide comparisons of how different choices made at each stage may affect the results. Following the decoding tutorial, in the Additional Analyses section we describe three extensions to the method: (1) temporal generalization (King & Dehaene, 2014), (2) RSA (Kriegeskorte, Mur, & Bandettini, 2008), and (3) classifier weights projection (Haufe et al., 2014). Finally, we outline important caveats and limitations of the decoding approach in the General Discussion section. See Figure 3 for an overview of the analysis pipeline and the structure of the article, including the relevant section heading titles.

DESCRIPTION OF EXPERIMENT

In this tutorial, we use MEG data to illustrate the effect that different choices made at several analysis stages have on the decoding results. Object animacy has been shown to be a reliably decoded categorical distinction in studies using both fMRI (Proklova, Kaiser, & Peelen, 2016; Sha et al., 2015; Kriegeskorte, Mur, Ruff, et al., 2008; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006) and MEG data (e.g., Cichy et al., 2014; Carlson, Tovar, et al., 2013). Here we use this robust paradigm as a basis for comparing the

Figure 2. An illustration of how multivariate analysis can result in increased sensitivity compared with univariate analysis. (A) Example average ERPs in response to two stimuli (Class A and Class B) are shown in two channels (left and right). The responses to the two classes in the individual channels overlap substantially and potentially nonsignificant in a univariate analysis. (B) The same responses represented as points in two-dimensional space, showing the activation in the two channels at one time point (i.e., location of the vertical gray bar in the ERP plots). When combining the information from both channels as in a decoding analysis, it is possible to define a boundary (dashed line) separating the two classes (distributions plotted orthogonal to the dashed line).



consequences of different analysis decisions in a decoding pipeline.

Twenty healthy volunteers (4 men) with a mean age of 29.3 years (ranging between 24 and 35 years) participated in the study. Informed consent in writing was obtained from each participant before the experiment, and the study was conducted with the approval of the Macquarie University human research ethics committee. The stimuli were images of 48 visual object exemplars (24 animate and 24 inanimate), segmented and displayed on a phase-scrambled background (see Figure 4).[1] Stimulus presentation was controlled by custom-written MATLAB scripts using functions from Psychtoolbox (Kleiner et al., 2007; Brainard, 1997; Pelli, 1997). The images were shown briefly for 66 msec (at 9° visual angle) followed by a fixation cross with a random ISI between 1000 and 1200 msec. Participants were instructed to categorize the stimulus as "animate" or "inanimate" as fast and accurately as possible, using a button press. The response button mapping alternated between 7-min blocks to avoid confounding the response with stimulus category (see Common Pitfalls section). This resulted in 32 trials per exemplar, 768 trials per category (animate/inanimate), and 1536 trials total per participant. All trials were included in the analysis, regardless of response, eye blinks, or other movement artifacts.

Data Collection

The MEG signal was continuously sampled at 1000 Hz from 160 axial gradiometers[2] using a whole-head MEG system (Model PQ1160R-N2, KIT, Kanazawa, Japan) inside a magnetically shielded room (Fujihara Co. Ltd., Tokyo, Japan) while participants lay in a supine position. Recordings were filtered online with a high-pass filter of 0.03 Hz and a low-pass filter of 200 Hz. The recordings were imported into MATLAB using the Yokogawa MEG Reader Toolbox for MATLAB (v1.04.01; Yokogawa Electric Corporation). The first step in the pipeline was to slice the data into epochs (i.e., trials), time-locked to a specific event. We extracted from −100 to 600 msec of MEG data relative to the stimulus onset. The first 100 msec of signal taken before trial onset serves as a sanity check for decoding accuracy (see Cross-validation section).

Analysis Summary

The effect of different choices on the decoding results will be described by systematically varying one parameter relative to a set of fixed parameters. Three caveats of this approach are that (1) as these parameters are not independent, interactions between analysis decisions are likely; (2), the effects of these analysis decisions will vary between data sets; and (3) drawing conclusions on differences in decoding performances is only valid when the noise level is the same in all cases. Consequently, the following results should be interpreted as illustrative rather than provide prescriptive analysis guidelines. All analysis code for the examples was written in MATLAB, using only standard functions unless otherwise specified. To illustrate the effects of different parameters on the results, they are consistently shown at the final stage plotted as a function of classifier accuracy over time. The default methods and fixed parameters are listed here for reference, and unless otherwise specified, the results in Figures 6–10 are obtained using this default pipeline:

- Preprocessing (see Preprocessing section)
    - Subsampling 200 Hz
    - Averaging four trials
    - PCA retaining 99% of the variance
- Decoding (see Decoding section)
    - Naive Bayes classifier
    - Leave-one-exemplar-out cross-validation

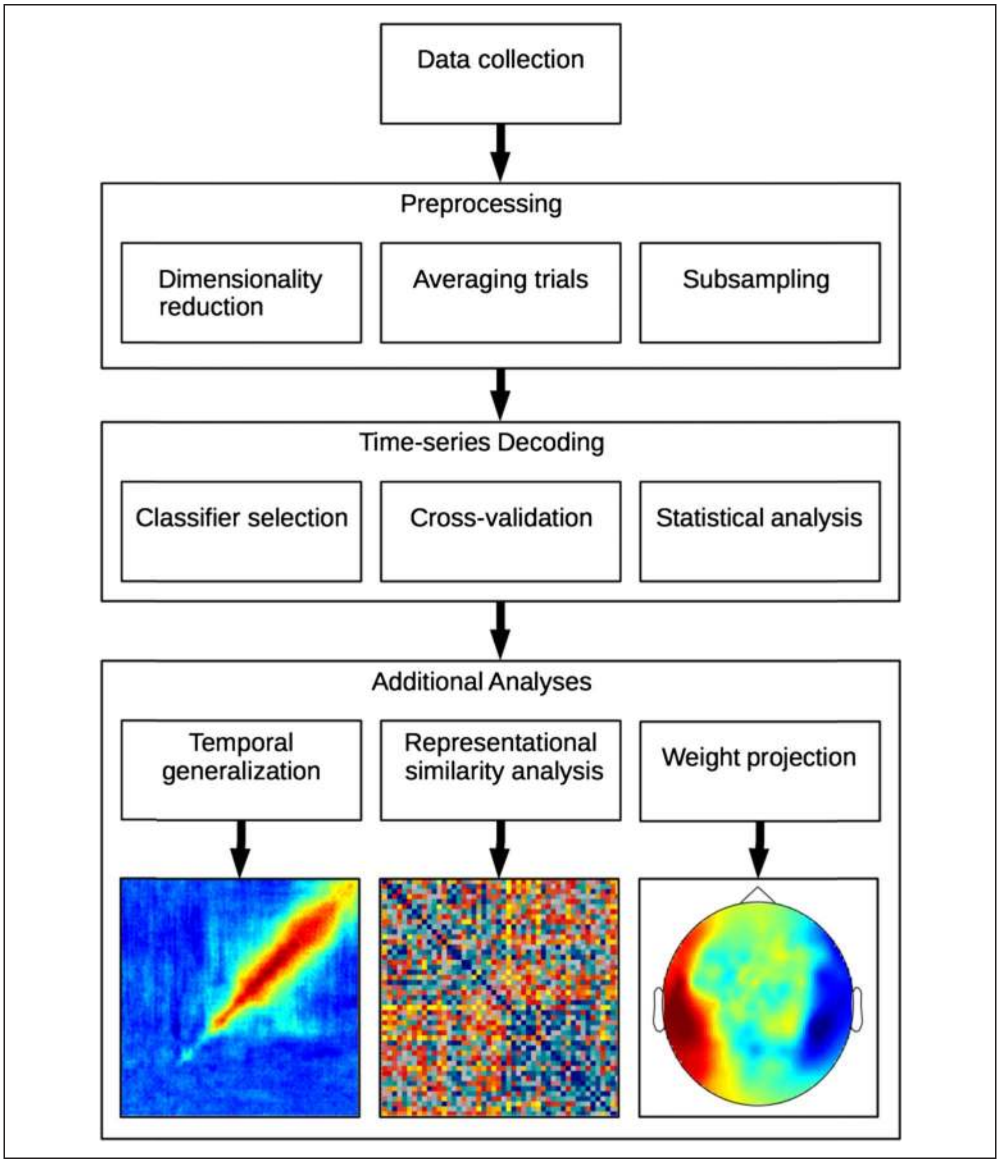


Figure 3. A schematic overview of a typical analysis pipeline. Refer to the relevant sections in the article for further details. This overview illustrates a general pipeline for decoding studies. The practical differences between decoding with MEG/EEG data versus fMRI data arise in both the preprocessing and analysis stages.

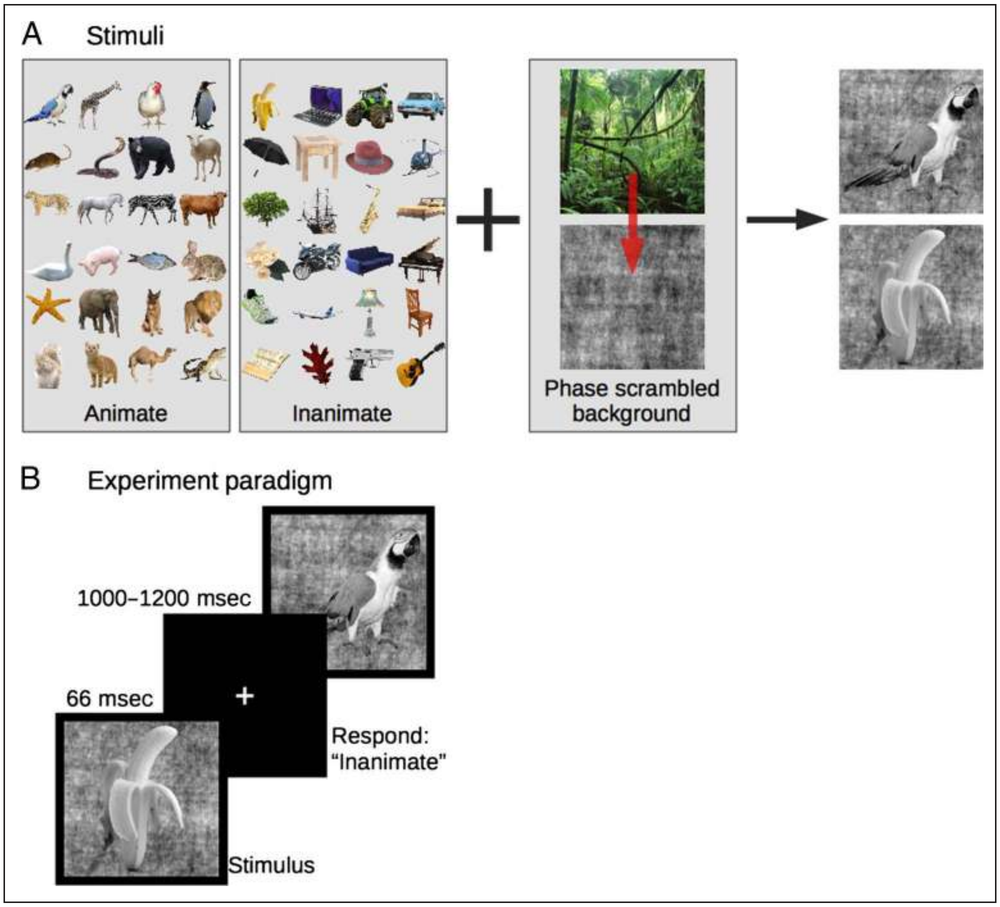


Figure 4. Illustration of the experimental design. (A) The stimuli consisted of 24 animate and 24 inanimate visual objects, converted to gray-scale and overlayed on a phase-scrambled natural image background. (B) Stimuli were presented in random order for 66 msec followed by a random ISI between 1000 and 1200 msec. Participants categorized the animacy of the stimulus during the ISI with a button press.

The results are reported as time-varying decoding accuracy, that is, higher accuracies reflect better decoding (prediction) of stimulus animacy from the MEG data. To assess whether accuracy was higher than chance, a Wilcoxon signed-rank test on the grand mean of decoding performance ($n = 20$) was performed at each time point. The resulting p values were corrected for multiple comparisons by controlling the false discovery rate (FDR; Benjamini & Hochberg, 1995). Note that these statistics were chosen for their simplicity and ease of use; we discuss commonly used options for assessing classifier performance and statistics in the Evaluation Of Classifier Performance and Group Level Statistical Testing section.

Figure 5 shows the result of this default pipeline. As expected, before stimulus onset (−100 to 0 msec), decoding performance is at chance (50%), confirming that there is no animacy information present in the signal. Then, approximately 80 msec after stimulus presentation, the classifier's performance rises significantly above chance for almost the entire time window (to 600 msec). Thus, at these time points, we are able to successfully decode from the MEG activation patterns whether the presented stimulus in a given trial was animate (parrot, dog, horse, etc.) or inanimate (banana, chair, tree, etc.). This indicates that the MEG signal contains information related to the animacy of the stimulus. The next sections will describe this pipeline in detail while comparing the effect of different analysis decisions.

PREPROCESSING

Neuroimaging data are often noisy. The signals in imaging data are weak compared with, for example, environmental noise, baseline activity levels, or fluctuations caused by eye blinks or other movements. Therefore, a set of standard procedures is used to increase the signal-to-noise ratio. Furthermore, neuroimaging data are high dimensional, and it is common practice to restrict the analysis to fewer dimensions. In MEG decoding, the dimensions of the data are generally reduced in the number of features (i.e., channels) that are input to the classifier. In addition, temporal smoothing is commonly applied. There are multiple ways to achieve these preprocessing steps, the most common are described in this section.

Data Transformation and Dimensionality Reduction

A standard step in preprocessing is to reduce the dimensionality of the data. Some classifiers require more training samples than features, and others might overfit to noise in the data if provided with too many features

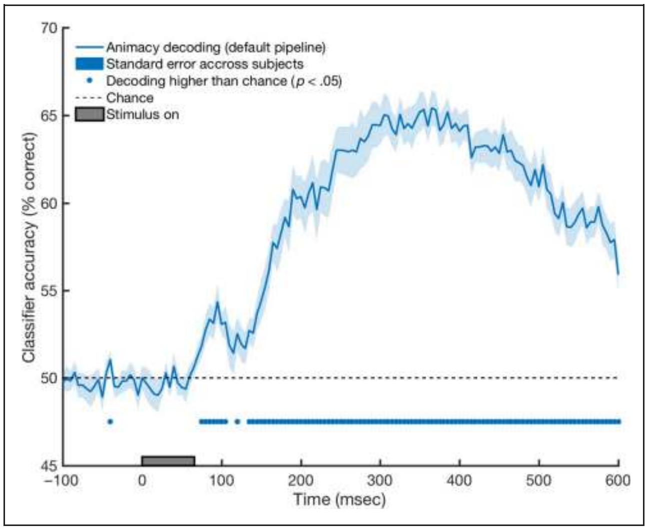


Figure 5. Decoding animacy from MEG data using the default analysis pipeline. Classifier accuracy (percent correct averaged across participants) is shown as a function of time relative to stimulus onset at 0 msec. The dashed line marks chance classification accuracy at 50%. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

(Misaki, Kim, Bandettini, & Kriegeskorte, 2010; De Martino et al., 2008; Bishop, 2006) or require longer computation time. Raw MEG recordings consist of many channels, typically 160 or more, and there is considerable redundant information, for example, in adjacent channels. It is therefore common practice to reduce the dimensionality of the data by feature selection before decoding, which can be accomplished in multiple ways. One approach is to select the channels that are most informative (Hanke, Halchenko, Sederberg, Hanson, et al., 2009; De Martino et al., 2008). For example, Isik et al. (2014) use an ANOVA significance test to select the MEG channels that contain significant stimulus-specific information.

Alternatively, one can use unsupervised, data-driven approaches such as PCA, which transforms the data into linearly uncorrelated components with the same number of feature dimensions, ordered by the amount of variance explained by each component (for a detailed introduction to PCA, see Jackson, 1991). The use of PCA for MEG has a number of advantages: First, retaining only the components that account for most of the variance substantially reduces the dimensionality of the data. In the example data (160 channels), on average 48.16 (*SD* = 7.05, range = 26–79) components accounted for 99% of the variance in the data. Second, PCA can separate out noise and artifacts such as eye blinks (see Improving Signal to Noise section) into their own components. These components can then be suppressed by the classifier because they do not contain class-specific information. Third, as the resulting PCA components are uncorrelated, it allows for using simpler (i.e., faster) classifiers that assume no feature covariance (e.g., naive Bayes; see Classifiers section).

Figure 6 illustrates the effect of the described dimensionality reduction methods on decoding performance for the example data. For this data set and classifier, PCA yields much better performance compared with using the raw channels (cf. Isik et al., 2014). Note that these differences are classifier dependent (as shown in the Classifiers section). Here, the PCA transformation was computed on the training data and applied on the test data, separately for each time point and separately for each training fold. Alternatively, one could compute one transformation for the whole time series and/or do this on all data before the cross-validation process. However, this is only viable if the goal of the analysis is statistical inference (Hebart et al., 2015), as it could result in more optimistic decoding accuracies that would not generalize to new data.[3]

An alternative method is to transform the sensor level data into activations in virtual source space. Instead of decoding channel level activations, source reconstruction (e.g., beamformer [Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997] or minimum norm estimate [Hämäläinen & Ilmoniemi, 1994]) can be applied during preprocessing. Classification is then performed in source space rather than channel space (Sandberg et al., 2013; van de Nieuwenhuijzen et al., 2013; Sudre et al., 2012). Using source space for decoding has the potential to improve classification accuracies (Sandberg et al., 2013; van de Nieuwenhuijzen et al., 2013), as source reconstruction algorithms can ignore channel level noise. Inferences about the spatial origin of the decoded discrimination can be made by restricting the classifier to considering signals from predetermined ROIs (e.g., Sudre et al., 2012)

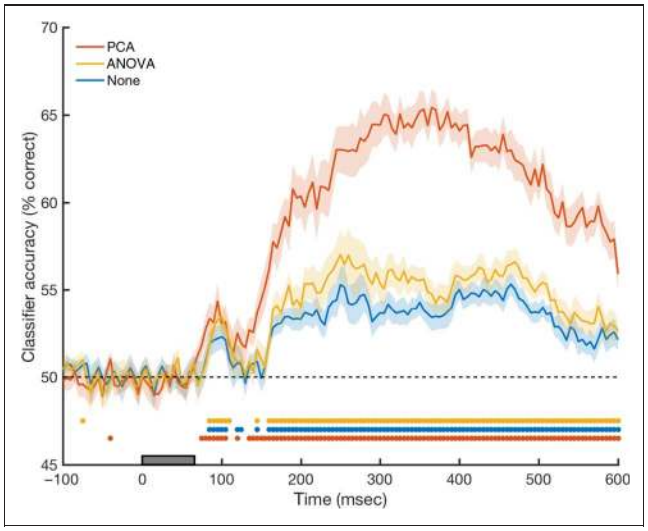


Figure 6. The effect of dimensionality reduction methods on decoding performance. The effect of channel selection using ANOVA (yellow line) is marginally better than using the raw data (blue line). Using PCA (red line) yields the largest gain in performance. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

or by using the complete source space reconstruction and projecting the classifier weights (see Weight Projection section) into source space (e.g., van de Nieuwenhuijzen et al., 2013). The second approach relies on interpreting classifier weights, and therefore, the reliability of the sources depends not only on the reconstruction quality but also on decoding performance (see Weight Projection section). Source reconstruction methods are still developing, and reconstruction accuracies are likely to improve in the future, making source space decoding an attractive option. However, as source space decoding has not been widely used to date, we will not cover it in the rest of this tutorial.

Improving Signal to Noise

MEG data are generally sampled at high frequencies (e.g., 1000 Hz), and a common strategy to improve signal-to-noise ratio (the strength of the signal compared with the strength of the background noise) is by collapsing data over time. The two main approaches are to classify on more than one time point using a sliding window (e.g., Ramkumar et al., 2013) or down-sample the data to lower frequencies (see Figure 7). The difference between the methods is that, when using a sliding window, the classifier has access to all time points in the window (the number of features is increased), whereas in subsampling, it receives the average (the number of features at each time point stays the same). For the example data, subsampling has a small effect on decoding performance but also benefits the analysis by reducing the computation time for the decoding analysis as there are fewer time points to classify. The sliding window approach also improves performance, but the benefit is marginal especially considering that the computation time increases significantly with larger sliding windows, as the classifier is still trained and tested at each time point. The optimal parameters will depend on the particular data set and desired temporal resolution. An important caveat for both approaches is that estimates of both decoding onset and the time of peak decoding are affected by the choice of subsampling or sliding window. When using a sliding time window, the last time bin in the window should be used for determining the onset (as in Figure 7) to avoid shifting the onset forward in time. It is recommended to apply a low-pass filter before resampling (e.g., subsampling using the decimate function in MATLAB) as subsampling can cause aliasing. Low-pass filtering, however, can cause an artifact whereby significant decoding emerges even when no signal exists in the original data (Vanrullen, 2011). For the example data, we subsampled by a factor of 5 to obtain a sampling rate of 200 Hz.

Another source of noise originates from artifacts. Eye blinks, eye movements, heartbeats, and muscle movement can cause significant artifacts. Typically, in classical M/EEG analyses trials containing such artifacts are manually inspected and excluded from the analysis, or independent component analysis is used to separate out these artifacts into their own components, which are then removed manually or automatically (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). Experiments can also be designed in a way to reduce the number of artifacts, for example, by instructing participants to blink in response to a particular stimulus that is not part of the analysis (Cichy et al., 2014). We did not perform any artifact rejection on our data and found classification performance to be well above chance, but this can vary across data sets. As classifiers have the capacity to learn to ignore bad channels or supress noise during training, artifact correction is likely less critical in decoding analyses. However, note that if artifacts are confounded with a condition (e.g., if more eye movements occurred in one condition than the other due to some property of

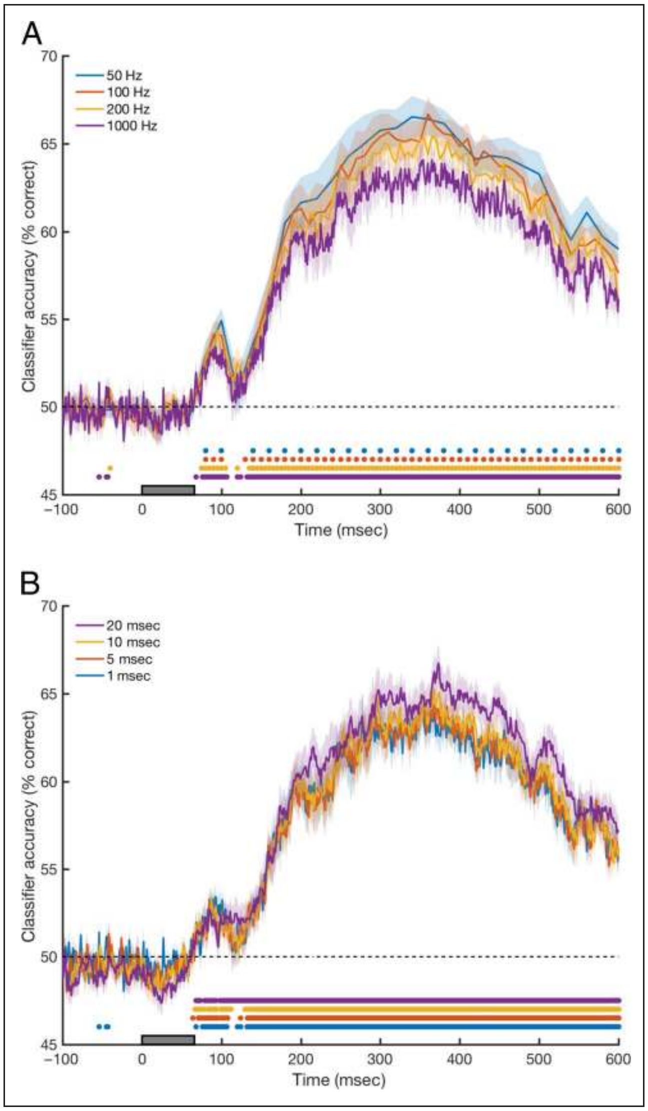


Figure 7. The effect of (A) subsampling and (B) sliding window approaches to improving signal-to-noise ratio on classifier accuracy. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

the stimulus), this would make the artifacts a potential source of discrimination information for the classifier. If this is the case, it would not be possible to determine whether the classifier was decoding the experimental condition, or the correlated difference in artifacts (see also Common Pitfalls section).

Increased signal-to-noise ratio can also be achieved by averaging trials belonging to the same exemplar before decoding (Isik et al., 2014). Averaging increases general decoding performance and makes signatures (e.g., onsets, maxima or minima) more pronounced. This effect is shown in Figure 8, where different numbers of trials (belonging to the same exemplar) are averaged. Interestingly, the first onset of decoding is similar regardless of the number of trials that are averaged. The greatest increase in performance (in our example data) is observed when averaging four trials. Averaging more trials does not increase decoding performance by the same factor, suggesting that here four trials is a good trade-off between signal-to-noise ratio and trials per exemplar. The trade-off to consider when selecting the number of trials to average is that reducing the trials per exemplar (e.g., averaging 32 trials here produces only one trial per exemplar) typically increases the variance in (within-subject) classifier performance. Alternatively, when not enough trials are available, the trials used for training the classifier could be sampled with replacement (bootstrapped). The optimal number of trials to average will differ for different data (e.g., in Isik et al., 2014, averaging 10 trials was used). Note that trial averaging does not affect model testing (e.g., RSA, see the Representational Similarity Analysis section), as relative decoding performance is scaled similarly between exemplars or time points.

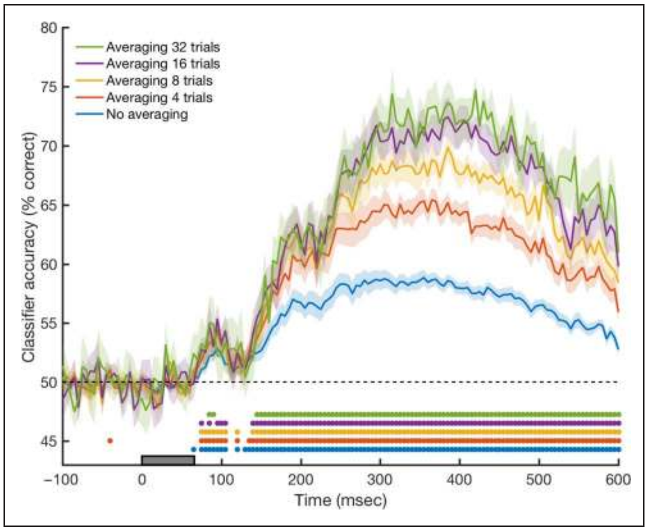


Figure 8. The effect of averaging trials on decoding performance. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

DECODING

Decoding analysis is performed on the preprocessed data. To summarize, in preprocessing the raw MEG signal is sliced into epochs from −100 to 600 msec relative to stimulus onset, then down-sampled to 200 Hz. Groups of four single trials are averaged to boost signal-to-noise ratio, resulting in eight pseudotrials for each object exemplar. These preprocessed pseudotrials are the input to the classifier in the decoding analysis.

To decode the class information (animacy) from the MEG data, a pattern classifier (see Classifiers section) is trained to distinguish between two classes of stimuli (animate and inanimate objects). The classifier's ability to generalize this distinction to new data is assessed using cross-validation (see Cross-validation section). If the classifier's performance after cross-validation is significantly above chance, this indicates that the MEG patterns contain class-specific information, and we conclude that the class can be decoded from the MEG data. In time-resolved MEG decoding studies, this process is repeated on all time points in the data. Then, for example, one can examine when the peak in decoding performance occurs, that is, at what time point the information in the signal allows for the best class distinction. Another feature often used is the onset of significant decoding performance to determine the earliest time that class-specific information becomes available. These signatures can then be compared across experimental conditions.

Classifiers

There are numerous types of classifiers that originate from the machine learning literature. Classifier choice has the potential to influence experimental results, as different classifiers make different assumptions about the data. In addition, the goal of classification in machine learning is high predication accuracy, which drives the development of increasingly sophisticated classifier algorithms. In contrast, prediction is not the main goal of decoding in neuroscience, and classifier choice instead favors simplicity and ease of interpretation over optimizing prediction accuracies. Therefore, for brain decoding studies, linear classifiers are generally preferred, as they are simpler in nature, making interpretation less complex (Schwarzkopf & Rees, 2011; Misaki et al., 2010; Müller et al., 2003). The default classifiers used in fMRI decoding are typically linear support vector machines (SVM) or, to a lesser extent, correlation classifiers. However, fMRI data typically have many features/dimensions. SVM is generally better than other classifiers when dealing with many features and is therefore a popular choice. In comparison with fMRI data, time series data often has fewer features (e.g., our example MEG data set uses only ~50 components following PCA). Consequently, it is possible that there are differences in the suitability of different classifiers for fMRI versus time series decoding analysis.

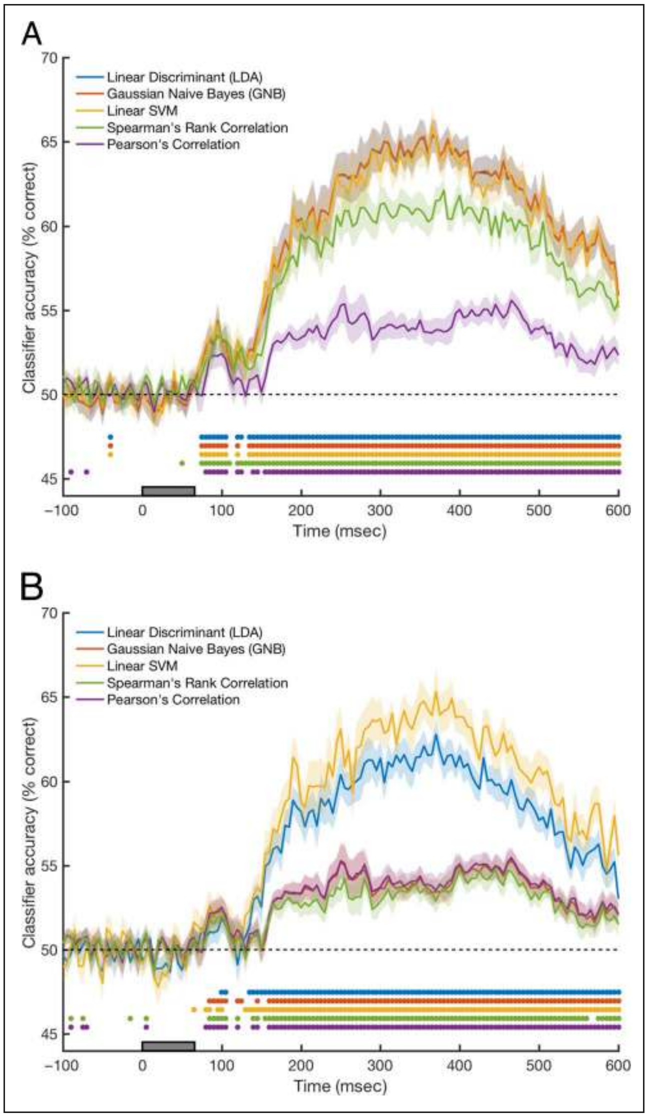


Figure 9. Comparison of classification accuracy as a function of classifier type. (A) Using the standard decoding pipeline. (B) Using the standard pipeline without performing PCA. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

Here we compare the performance of SVM, correlation classifiers, and two common alternatives (linear discriminant analysis [LDA] and Gaussian naive Bayes [GNB]) on the example MEG data (Figure 9), using their built-in MATLAB implementations (and default parameters). Notably, LDA, GNB, and SVM have the best overall performance. Taking the complexity of the classifier into account, which affects the computational requirements and given that classification is generally repeated many times (e.g., on multiple time points), this argues in favor of the discriminant classifiers (GNB and LDA), which are faster to train than SVM. Interestingly, despite their relative popularity in fMRI, the correlation classifiers did not perform as well on our data. However, Isik et al. (2014) reported correlation classifier performance for their MEG data on par with other classifiers. This difference could be due to many factors, for example, different choices in the preprocessing pipeline or experimental design. To illustrate that classifier performance depends on preprocessing, we tested the same classifiers using different preprocessing decisions. For example, Figure 9B shows that not performing PCA has a large effect on GNB performance, but a smaller effect on the performance of LDA and SVM. These dependencies highlight the difficulty in attempting to make universal recommendations for decoding analyses. Furthermore, each classifier has a number of parameters that may be optimized; however, most neuroscience studies use standard classifier implementations.

Cross-validation

An essential step in decoding analysis is cross-validation: This provides an evaluation of classifier generalization performance. In standard *k*-fold cross-validation, the data are divided into *k* subsets (i.e., folds), where each subset contains a balanced amount of trials from each class (e.g., animate and inanimate exemplars in our example experiment). The classifier is trained using all-but-one subsets (the training set). Next, the trained classifier is used to predict the class of the trials from the remaining subset (the test set). This process is repeated for all subsets, and the average classifier performance across all folds is reported. This method makes maximal use of the available data, as all trials are used for testing the classifier. Note that in fMRI decoding the sets are often based on experimental runs (leave-one-run-out cross-validation), as the trials within each run are not independent (e.g., due to the slow hemodynamic response). In MEG decoding, individual trials are generally assumed to be independent (Oosterhof et al., 2016), and trials are randomly assigned to train and test sets. The theoretical optimal performance is obtained by leave-one-trial-out cross-validation, where the classifier is trained on all-but-one trial. It is however computationally more intensive, especially with many trials (which is typically the case in MEG).

As with other analysis decisions, the most appropriate implementation of cross-validation is guided by the experimental design. Standard *k*-fold cross-validation assigns individual trials to training and testing sets. Depending on the research question, this may produce a confound in the class distinction that the classifier learns from the training data. For example, for decoding animacy, standard cross-validation would entail that trials belonging to the same exemplar (e.g., "car") are assigned to both training and test sets. Consequently, it may be possible for the classifier to learn to distinguish the classes based on the activation patterns evoked by visual properties of specific exemplars. This makes it unclear whether the classification boundary is based on animacy or visual features. To avoid this, when decoding categories composed of many exemplars, we recommend leave-one-exemplar-out cross-validation (see Carlson, Tovar, et al., 2013), where all trials belonging to one exemplar (e.g., car) are assigned to the test set and the classifier is trained on the data from the other exemplars (e.g., "dog" and "chair"). This is repeated for all exemplars (i.e., every exemplar is assigned to the test set once).

Figure 10 shows decoding accuracy for different forms of cross-validation, including an invalid analysis without cross-validation. Note that without cross-validation, classifier performance is above chance before stimulus onset. This nonsensical result arises from the test data being used to train the classifier, violating the constraint of independence. Time-resolved decoding methods have a convenient built-in check for this: Above-chance decoding performance before stimulus onset suggests that an error exists in either the preprocessing or cross-validation stages. In our data, 10-fold and leave-one-trial-out cross-validation yielded very similar results, suggesting that the optimal split is data specific. Furthermore, by comparing performance between traditional cross-validation (e.g., *k*-fold) and leave-one-exemplar-out, it is possible to estimate to what degree classifier performance is driven by individual stimulus properties (e.g., low-level visual properties of the exemplar images). The difference between *k*-fold and leave-one-exemplar-out cross-validation is observed early in the time series (consistent with the timing of early visual feature processing) and is reduced later in the time course (Figure 10). Taken together, a valid form of cross-validation with independent training and test data is essential. Although there are several ways of splitting up the data into training and test sets, the particular version of cross-validation implemented must be compatible with the research question.

Evaluation of Classifier Performance and Group Level Statistical Testing

Statistical evaluation of decoding analyses is a complex issue, and there is no consensus yet on the optimal approach (Allefeld, Görgen, & Haynes, 2016; Noirhomme et al., 2014; Schreiber & Krekelberg, 2013; Stelzer, Chen, & Turner, 2013; Nichols & Holmes, 2002). The statistical approach used in our example analysis is common in the literature (e.g., Ritchie et al., 2015; Carlson, Tovar, et al., 2013) and was chosen for its simplicity; however, there are several alternative methods that are also valid. For example, we report classifier performance as accuracy (percent correct). Accuracy is a less appropriate measure when dealing with unbalanced data (more trials exist for one class than for the other), as a trained classifier could exploit the uneven distribution and achieve high accuracy simply by predicting the more frequent class. For unbalanced data, a measure of performance that is unaffected by class bias such as d' is more appropriate. Alternatively, "balanced accuracy" includes the mean of the accuracies for each class and thus is also unaffected by any class imbalance in the data.

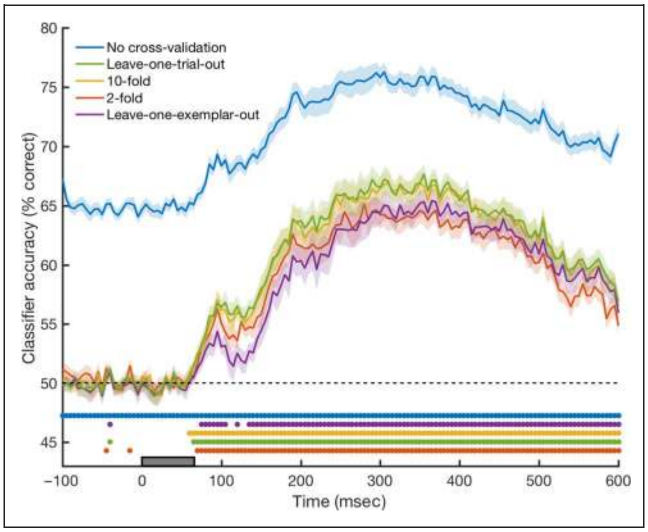


Figure 10. Classification accuracy as a function of cross-validation method. The shaded area is the standard error across participants. Discs above the *x* axis indicate the time points where decoding performance is significantly higher than chance.

Several options exist for assessing whether classifier performance is significantly above chance. The nonparametric Wilcoxon signed-rank test (Wilcoxon, 1945) was used in our example (see also Ritchie et al., 2015; Carlson, Tovar, et al., 2013), as it makes minimal assumptions about the distribution of the data. Alternatively, the Student's *t* test is also commonly used (but see Allefeld, Görgen, & Haynes, 2016). Another popular alternative is the permutation test, which entails repeatedly shuffling the data and recomputing classifier performance on the shuffled data to obtain a null distribution, which is then compared against observed classifier performance on the original set to assess statistical significance (see, e.g., Kaiser et al., 2016; Cichy et al., 2014; Isik et al., 2014). Permutation tests are especially useful when no assumptions about the null distribution can be made (e.g., in the case of biased classifiers or unbalanced data), but they take much longer to run (e.g., repeating the analysis ~10,000 times).

Importantly, as is the case in fMRI analyses, time series neuroimaging analyses also require addressing the problem of multiple comparisons (Nichols, 2012; Bennett, Baird, Miller, & Wolford, 2011; Bennett, Wolford, & Miller, 2009; Pantazis, Nichols, Baillet, & Leahy, 2005) as typically multiple tests are conducted across different time points. The FDR adjustment used in our example analysis is straightforward, but a limitation is that it does not incorporate the relation between time points (Chumbley & Friston, 2009). Alternatively, cluster-based multiple-comparison correction involves testing whether clusters of time points show above-chance decoding and therefore can result in increased sensitivity to smaller, but more sustained effects (Oosterhof et al., 2016; Mensen & Khatami, 2013; Nichols, 2012; Smith & Nichols, 2009).

ADDITIONAL ANALYSES

In the sections above, we illustrated the standard approach to decoding time series neuroimaging data. Here we outline three extensions for decoding analysis. The first is temporal cross-decoding (see The Temporal Generalization Method section), which tests the degree to which activation patterns in response to the experimental conditions are sustained or evolve over time. The second is the RSA framework (see Representational Similarity Analysis section), which facilitates the testing of models of the structure of decodable information over time. Finally, we outline a method that involves projection of the classifier weights to determine the spatial source of the signal driving the classifier in sensor space (see Weight Projection section).

The Temporal Generalization Method

An advantage of time series decoding is that it has the potential to reveal the temporal evolution of brain activation patterns, rather than providing a single, static estimate of decodability for a stimulus or task. One method is to train a classifier on a particular time point and then test its decoding performance on different time points. This form of cross-decoding reveals to what degree the activation patterns for a particular stimulus or task evolve. Classifiers effectively carve up multidimensional space to distinguish between the experimental conditions; thus, when a classifier that is trained on one time point can successfully predict class labels for data at other time points, it suggests that the structure of the multidimensional space is similar across time. Conversely, if cross-decoding is unsuccessful across two time points, it suggests that the multidimensional space has changed sufficiently for the boundary between classes determined at one time point to be no longer meaningful by the second time point. Beyond temporal characterization of the decoding results, this method has the potential utility to test cognitive models, which make theoretical predictions about the generalizability of representations (see also Figure 4 in King & Dehaene, 2014). For example, the temporal generalization of classifiers can be tested between two completely separate data sets. Isik et al. (2014) tested the temporal generalization performance of a classifier that was trained on stimuli that were presented foveally and then tested on peripherally presented stimuli. Similarly, Kaiser et al. (2016) used this method to distinguish category-specific responses from shape-specific responses.

Figure 11A shows cross-validated temporal cross-decoding performed on the example MEG data. The diagonal in this figure is analogous to the standard one-dimensional time series decoding plot (e.g., Figures 5–10). Significant points (shown in Figure 11B) off the diagonal indicate that the classifier, when trained on data from time point A, can generalize to data from time point B. The generalization accuracy normally drops off systematically away from the diagonal. In this case, classifier performance generalizes well for neighboring time points (red region on the diagonal) as expected and, additionally, to some extent between 150–200 and 300–500 msec, indicating that the MEG activation patterns are similar in these windows.

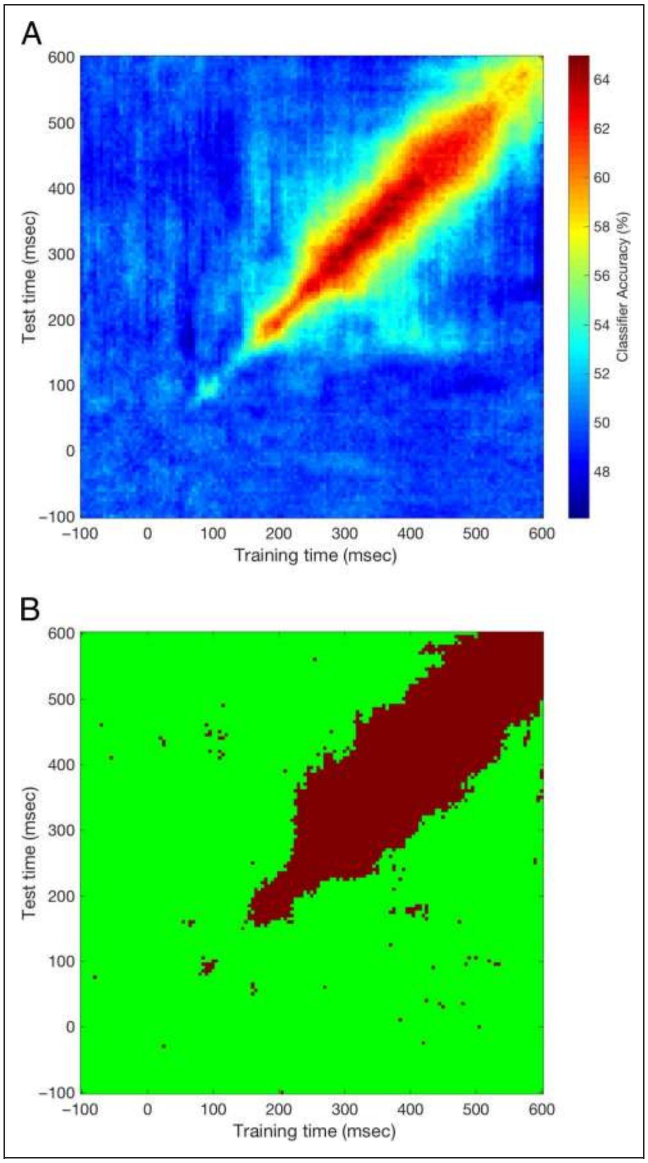


Figure 11. (A) Temporal generalization of decoding performance. A classifier is trained at one time point and tested at a different time point. This is repeated for all pairs of time points. The figure shows the generalization accuracy averaged over participants. (B) Map of time point pairs where the generalization was significantly different (red area) from chance (Wilcoxon signed rank test, controlled for multiple comparisons using FDR).

Representational Similarity Analysis

Standard decoding analysis reveals whether class-specific information is present in the neuroimaging signal. Approaches such as cross-decoding (e.g., temporal generalization) can begin to probe the underlying representational structure of the information in the brain activation patterns used by the classifier. RSA takes this concept further and provides a framework for testing hypotheses about the structure of this information (Kriegeskorte, Mur, & Bandettini, 2008). RSA is based on the assumption that stimuli with more similar neural representations are more difficult to decode. Conversely, stimuli with more distinct representations are expected to be easier to decode. Thus, the central idea is that representational similarity can be indexed by the degree of decodability. By comparing the decodability of all possible pairwise combinations of stimuli, a representational dissimilarity matrix (RDM) is calculated. That is, for each pair of stimuli, the distance between their activation patterns is computed using one of several distance metrics (e.g., correlation between the activation patterns or difference in classifier performance (Walther et al., 2016).

An example RDM is shown in Figure 12A, in which each cell in the matrix corresponds to the dissimilarity of two of the object stimuli in the MEG animacy experiment. For data with high temporal resolution such as MEG, a series of RDMs can be created for each time point

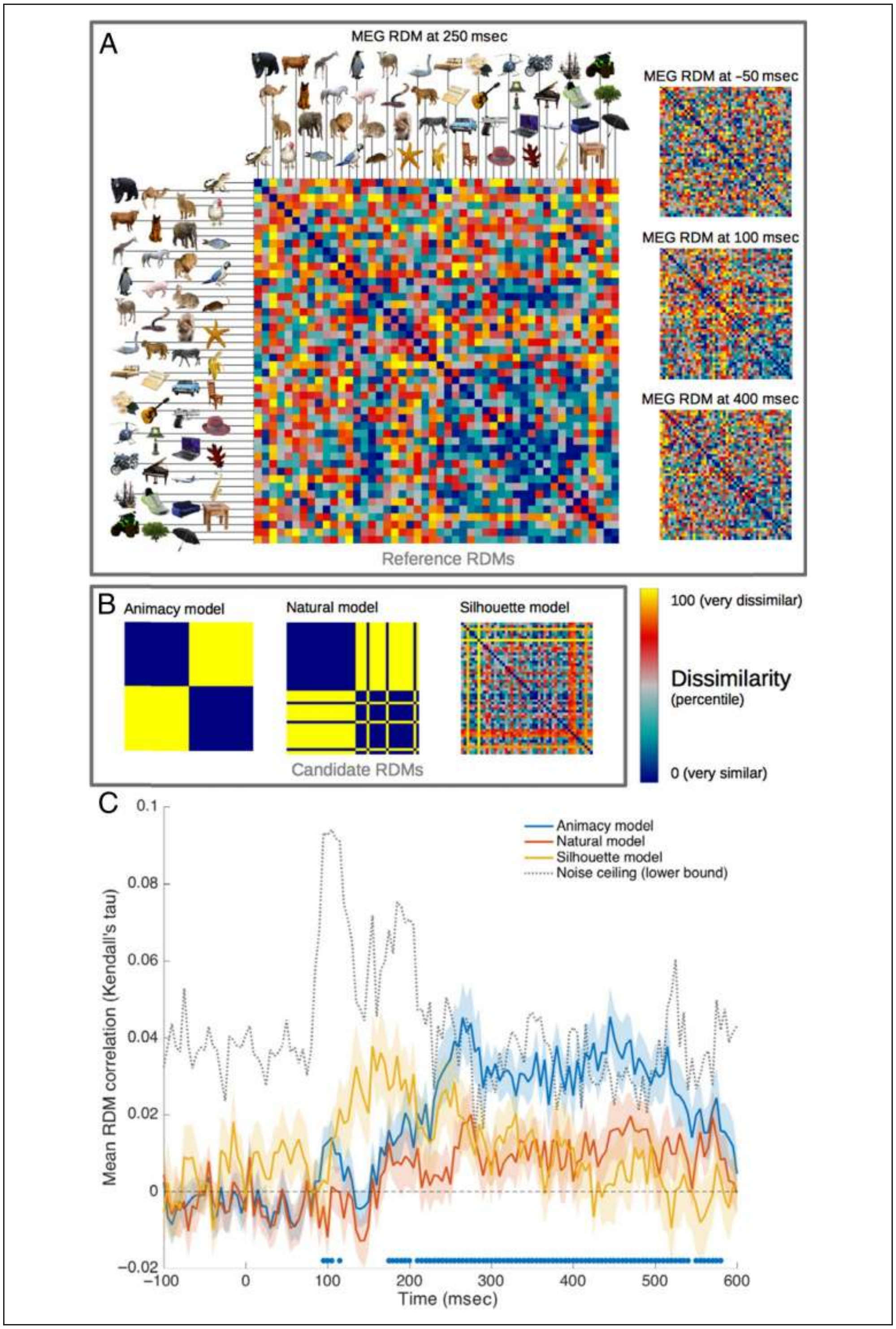


Figure 12. Model evaluation within the RSA framework. (A) The empirical MEG RDMs averaged across participants. One cell in the matrix represents the dissimilarity between the MEG activation patterns for one pair of object exemplars. RDMs are shown for four time points: −50 msec, 100 msec, 250 msec, and 400 msec. (B) Three model RDMs, which predict the representational similarity of the brain activation patterns for all object pairs based on different stimulus properties: an Animacy model (Animate vs. Inanimate objects), a Natural model (Natural vs. Artificial objects), and a Silhouette model (based on the visual similarity of the objects' silhouettes). (C) RSA model evaluation. At each time point, the empirical RDMs for each participant are correlated with the three candidate model RDMs in B. The strength of the average correlations shows how well the candidate models fit the data. Shaded areas represent the standard error over participants, and the marks above the *x* axis indicate time points where the mean correlation was significantly higher than zero (Wilcoxon signed-rank test, controlled for multiple comparisons using FDR). The gray dotted line represents the lower bound of the "noise ceiling" at each time point, which is the theoretical lower bound of the maximum correlation of any model with the reference RDMs at each time point, given the noise in the data (Nili et al., 2014).

and used to investigate the temporal dynamics of representations over time. The time-varying RDMs in Figure 12A are constructed by decoding all pairwise stimuli using the same pipeline (using twofold cross-validation, as leave-one-exemplar-out is not possible when decoding between two exemplars); thus, one square in the RDM represents the decoding accuracy for classifying between one pair. Following calculation of the RDM (either time-varying or static) from the empirical data, the empirical RDM can be compared with model RDMs that make specific predictions about the relative decodability of the stimulus pairs. In RSA studies to date, model RDMs have been constructed from predictions based on a wide range of sources, including behavioral results, computational models, stimulus properties, or neuroimaging data from a complementary imaging method such as fMRI (e.g., Cichy et al., 2014, 2016; Wardle et al., 2016; Redcay & Carlson, 2015; Carlson, Simmons, Kriegeskorte, & Slevc, 2013; Kriegeskorte, Mur, Ruff, et al., 2008).

Figure 12 shows the results of RSA model evaluation for the example MEG data. For each time point, the empirical RDMs (Figure 12A) are correlated with three theoretical models (Figure 12B); a model of stimulus animacy, a model that distinguishes artificial versus natural stimuli, and a control model based on the visual similarity of the exemplar's silhouettes (which correlates well with early stimulus discriminability; see, e.g., Redcay & Carlson, 2015; Carlson et al., 2011). Each of these models predicts the relative (dis)similarity of the MEG activation patterns for each exemplar pair based on their specific stimulus features. The extent of the correlation between the model and empirical MEG RDMs is interpreted as reflecting the degree to which the "representational structure" characterized by each model exists in the brain activation patterns. The results in Figure 12C are plotted as the correlation between the three model RDMs with the MEG RDM over time. The Animacy model (blue line) has a better fit to the MEG data than the Natural model (orange line), and both models have a better fit than the Silhouette model (yellow line) later in the time series. The Silhouette model has the best fit early in the time series, which is expected as it represents early visual features. This suggests that animacy is a relatively good predictor of the similarity of the MEG activation patterns for the exemplar pairs: object pairs from the same category (e.g., both animate) are more difficult to decode than object pairs from different categories (e.g., one animate and one inanimate). Within the RSA framework, this is interpreted as evidence that animacy is a key organizing principle in the representational structure of the object exemplars.

Despite its strengths, a current limitation of the RSA approach is that valid statistical comparison of different candidate models is difficult (Thirion, Pedregosa, Eickenberg, & Varoquaux, 2015; Kriegeskorte & Kievit, 2013). A recent development proposes evaluating model performance by comparing it to the highest possible performance given the noise in the data, called the "noise ceiling" (Nili et al., 2014). When applied to MEG data, the performance of various models relative to the noise ceiling (computed from the empirical data as described in Nili et al., 2014) can be evaluated over time, as shown in Figure 12C. Despite the present limitations in directly comparing different models, RSA is a useful tool for investigating the structure of the decodable signal in neuroimaging data, which will undoubtedly continue to evolve in its sophistication and utility. For a more detailed introduction, see Nili et al. (2014), Kriegeskorte and Kievit (2013), and Kriegeskorte, Mur, and Bandettini (2008).

Weight Projection

Following successful classification of experimental conditions, it is sometimes of interest to examine the extent to which different voxels (fMRI) or sensors (MEG/EEG) drive classifier performance. During standard classification analysis, each feature (e.g., MEG sensors) is assigned a weight corresponding to the degree to which its output is used by the classifier to maximize class separation. Therefore, it is tempting to use the raw weight as an index of the degree to which sensors contained class-specific information. However, this is not straightforward, as higher raw weights do not directly imply more class-specific information than lower weights. Similarly, a nonzero weight does not imply that there is class-specific information in a sensor (for a full explanation, proof, and example scenarios, see Haufe et al., 2014). This is because sensors may be assigned a nonzero weight not only because they contain class-specific information but also when their output is useful to the classifier in suppressing noise or distractor signals (e.g., eyeblinks or heartbeats). An elegant solution to this issue was recently introduced by Haufe et al. (2014) and has been applied to MEG decoding (Wardle et al., 2016). This consists of transforming the classifier weights back into activation patterns. Following this transformation, the reconstructed patterns are interpretable (i.e., nonzero values imply class-specific information) and can be projected onto the sensors. It is important to note, however, that the reliability of the patterns depends on the quality of the weights. That is, if decoding performance is low, weights are likely suboptimal, and reconstructed activation patterns have to be interpreted with caution (Haufe et al., 2014).

Here we summarize this transformation for MEG data and plot the results in Figure 13. First, the classifier weights (we used LDA instead of GNB in this example, as this method only applies to classifiers that consider the feature covariance) are transformed into activation patterns by multiplying them with the covariance in the data: $A = \mathrm{cov}(X) \times w$; where X is the $N \times M$ matrix of MEG data with N trials and M features (channels) and w is a classifier weight vector of length M. A is the resulting vector of length M containing the reconstructed activation patterns (i.e., the transformed classifier weights).

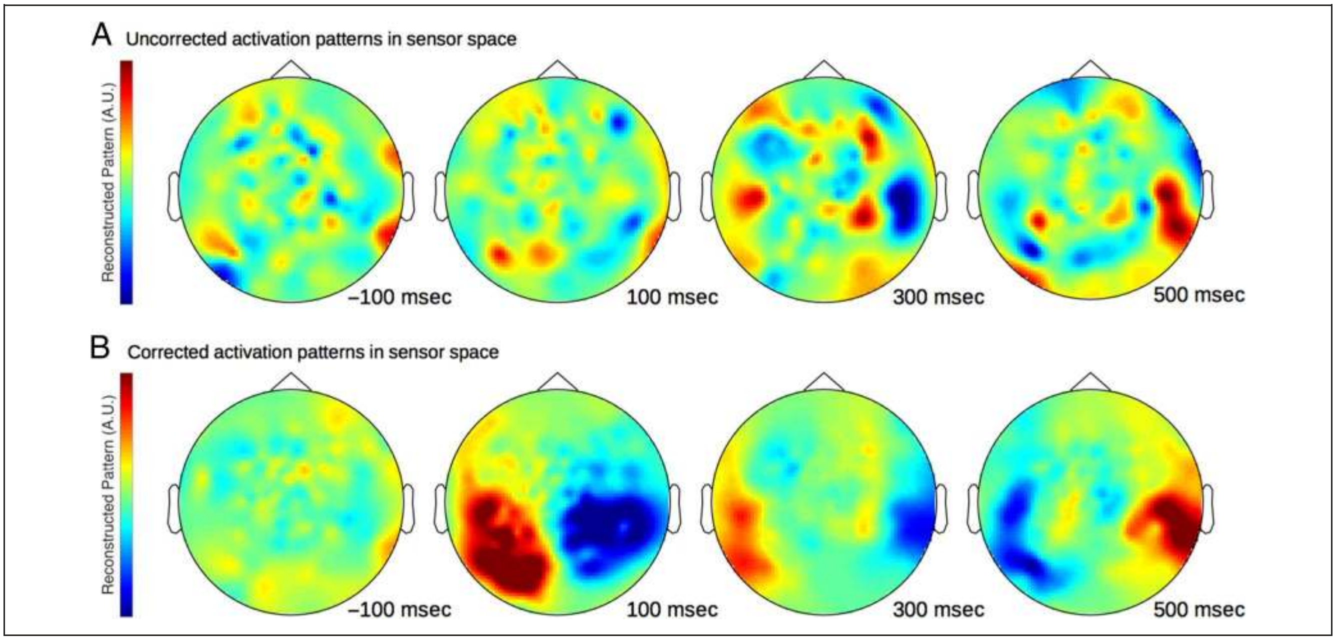

A Uncorrected activation patterns in sensor space
Reconstructed Pattern (A.U.)
−100 msec
100 msec
300 msec
500 msec
B Corrected activation patterns in sensor space
Reconstructed Pattern (A.U.)
−100 msec
100 msec
300 msec
500 msec

Figure 13. Classifier weights projected onto MEG sensor space. The corresponding time points are shown beneath the scalp topographies. Darker colors indicate channels that contribute to animacy decoding. (A) Uncorrected (raw) weights projections cannot be interpreted directly, as classifiers can assign nonzero weights to channels that contain no class-specific information. (B) The activation patterns computed from transformed weights (following the method of Haufe et al., 2014) can be interpreted.

For display purposes, the reconstructed activation patterns can be projected onto the scalp location of the channels. Figure 13B shows the result for the example MEG data at four time points (using the FieldTrip toolbox for MATLAB: Oostenveld, Fries, Maris, & Schoffelen, 2010); here the results are scaled by the inverse of the source covariance ($A \times \mathrm{cov}(X \times \mathrm{w})^{-1}$) to allow for comparison across time points. Note that this method cannot be directly used if multiple time points are used for classification (e.g., the sliding window approach described in the Improving Signal to Noise section). The uncorrected (raw) weight projections are shown for comparison in Figure 13A. We can now observe that, for the activation patterns in Figure 13B, the information source is located approximately around the occipital lobes (back sensors) at 100 msec and later around the temporal lobes (side sensors) at 300 msec, as expected from the visual processing hierarchy. Notably, this pattern is not as easily identifiable in the raw weight topographies shown in Figure 13A. For an in-depth explanation (with examples) of the weights interpretation problem and its solution, see Haufe et al. (2014).

GENERAL DISCUSSION

Time series decoding methods provide a valuable tool for investigating the temporal dynamics and organization of information processing in the human brain. In the previous sections, we outlined an example decoding analysis pipeline for time series neuroimaging data, illustrated effects of different methods and parameters (and their interactions), and introduced extensions of the method such as temporal generalization (see The Temporal Generalization Method section), RSA (see Representational Similarity Analysis section), and weights projection (see Weight Projection section). In the final section, we discuss some important aspects to consider when performing these analyses and interpreting the results. One of the central issues concerns the interpretation of classifier accuracy. Classifiers are extremely sensitive and will exploit all possible information in the data. This means that careful experimental design and interpretation of the results is required to draw meaningful conclusions from decoding studies (see, e.g., de-Wit, Alexander, Ekroll, & Wagemans, 2016; Carlson & Wardle, 2015; Naselaris & Kay, 2015). The next section outlines a number of such pitfalls to avoid in the implementation of time series decoding methods.

Common Pitfalls

The first caveat applies to all studies using classifiers and is well described in the literature (Kriegeskorte, Lindquist, Nichols, Poldrack, & Vul, 2010; Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Pereira et al., 2009). It is important that the classifier has no access to class-specific information about the data contained in the test set, as this will artificially inflate classifier performance. This analysis confound is referred to as “double dipping” and was demonstrated in the analysis without cross-validation in Figure 10 (see Cross-validation section). One advantage of time series decoding is that, in most cases, data obtained before stimulus onset serve as a first check. If classifier

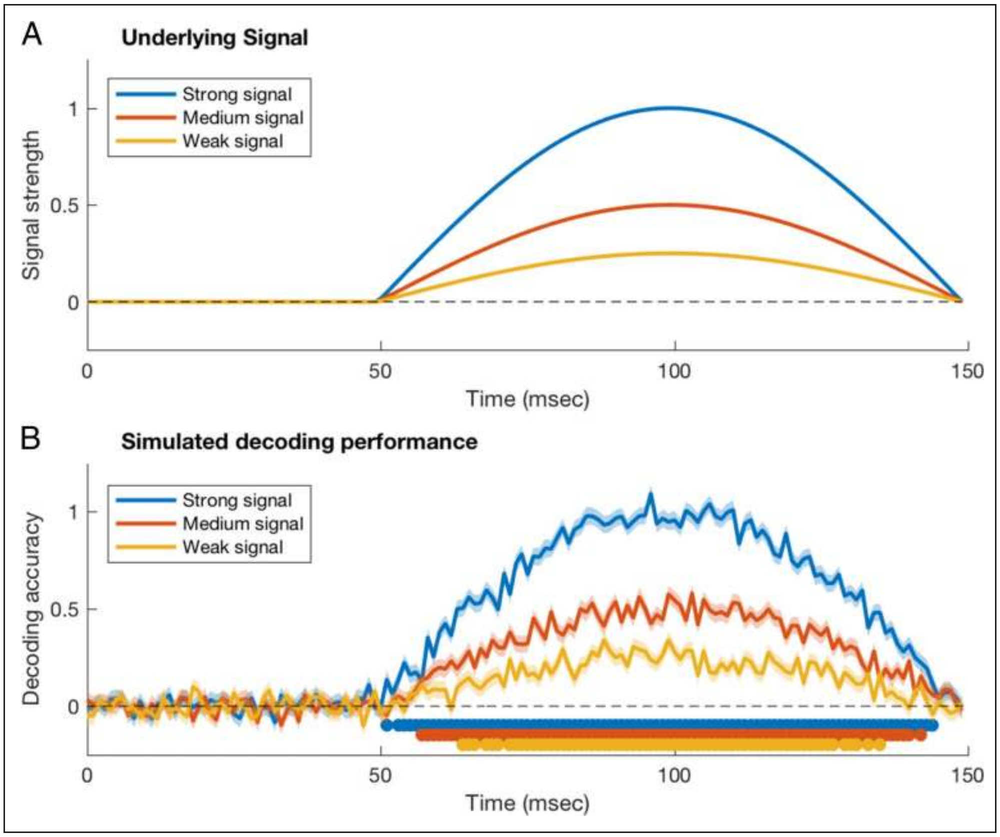


Figure 14. Demonstration of how the strength of peak decoding affects decoding onsets using stimulated data. (A) Three data sets were simulated to have the same onset and peak decoding latencies, but different peak strengths. (B) Gaussian noise was added to the underlying signals in each set (500 trials per set, σ = 1), and significant decoding (above zero) was assessed across the time course (signed-rank test, FDR-corrected). Colored discs above the *x* axis indicate time points with significant decoding.

accuracy is above chance before stimulus onset, it indicates possible contamination from double dipping.

A second caveat specific to time series decoding is that caution is required when interpreting (differences in) onsets of significant decoding. The time at which decoding is first significant for an experimental condition is determined by the underlying strength of the signal. For example, when the strength of peak decoding differs between two conditions (e.g., one is much easier to decode than the other), this will also affect the relative onset of decoding. This is illustrated in Figure 14. Three simulated data sets were constructed to have the same decoding onset (50 msec) and peak latency of decoding (100 msec), but different signal strengths (see Figure 14A). To evaluate how signal strength influences decoding onset, Gaussian noise was added to each data set and significance testing was conducted to find the onset of decoding (signed-rank test across time points, FDR-corrected). The outcome of the simulation is plotted in Figure 14B. Note that, although these simulated data sets were constructed to have an identical "true" onset of decoding, the onset of significant decoding is earlier for the set with a strong signal and much later for the set with the weak signal. This underscores the ambiguity in interpreting onset differences: It cannot be assumed that an earlier decoding onset reflects a true onset difference in the availability of decodable information between conditions. Isik et al. (2014) addresses this issue by using less data for the condition that had higher peak decoding and by equalizing the peaks across conditions before determining decoding onset.

Third, as noted earlier, filtering the signal can smear out information over time. An extreme example (using a step function) is illustrated in Figure 15, using simulated

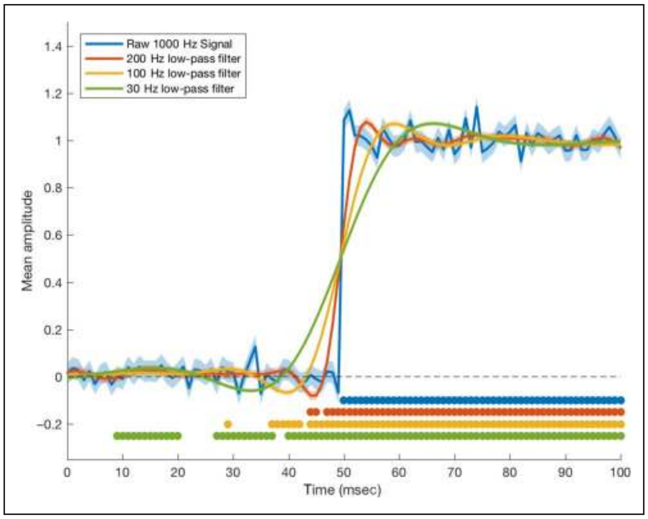


Figure 15. The effect of low-pass filtering on decoding onset. In this example, a signal with onset at 50 msec was simulated with added Gaussian noise (500 trials, σ = 1). The signal was then low-pass filtered using different cutoff frequencies. Time points where the trial average differed significantly from zero (signed-rank test, FDR-corrected) are indicated by the colored discs above the *x* axis.

data with a signal occurring at 50 msec. To demonstrate the effect of filtering, Gaussian noise was added to the signal, and low-pass filters were applied with different cutoff frequencies using the *ft_preproc_lowpassfilter* function (using the default Butterworth fourth-order two-pass IIR filter) from the FieldTrip toolbox (Oostenveld et al., 2010). The result of lowering the cutoff frequency is increased signal distortion. Applying a 30-Hz low-pass filter resulted in a signal that was significantly different from zero 40 msec earlier in the time series, compared with the simulated "true" onset at 50 msec. However, the effect is substantially reduced by applying much higher filter cutoffs, for example, 200 Hz. Therefore, interpretations based on the timing of decoding signatures relative to the stimulus should be avoided when using filters with a low cutoff frequency (Vanrullen, 2011).

Finally, decoding studies require careful experimental design to avoid confounds in the classifier analysis. The considerations vital to designing decoding studies are not necessarily the same as that for univariate analysis. Accordingly, care must be taken when reanalyzing data not originally intended for a decoding analysis. The high sensitivity of classifiers means that, if there are any differences between classes other than the intended manipulations, it is likely that the classifier will exploit this information, making it easy to introduce experimental confounds. An example is the effect of the participant's behavioral responses. In our example MEG experiment, the response buttons (to respond "animate" and "inanimate") were switched every block. If response mapping were uniform across blocks, response would be confounded with stimulus category, as a left button response would always correspond to "animate" and right for "inanimate." The physical pressing of the button would generate corresponding brain signals, for example, in motor areas, and this would provide a signal in the whole-brain MEG data that would correlate perfectly with the class conditions. In this case, it would be unclear whether the classifier decoded the intended experimental manipulation of "animacy" or simply the participant's motor responses. Alternatively, a classifier may distinguish between two conditions or categories of stimuli based on a confounding factor that covaries with class membership (e.g., differential attention to two conditions, leading to greater overall signal for one class) rather than the manipulation (e.g., difference in visual features or task difficulty) intended by the experimental design.

Furthermore, even with carefully controlled designs, the interpretation of decoding studies must be executed with caution. Decoding studies may conclude that Condition A is decodable from Condition B; however, the source of decodable information usually remains elusive (Carlson & Wardle, 2015; Naselaris & Kay, 2015). One notable example of this is the current debate surrounding the source of orientation decoding in fMRI (e.g., Pratte, Sy, Swisher, & Tong, 2016; Carlson & Wardle, 2015; Clifford & Mannion, 2015; Carlson, 2014; Alink, Krugliak, Walther, & Kriegeskorte, 2013; Freeman, Ziemba, Heeger, Simoncelli, & Movshon, 2013; Freeman, Brouwer, Heeger, & Merriam, 2011; Mannion, McDonald, & Clifford, 2009; Kamitani & Tong, 2005). Despite a decade of orientation decoding in early visual cortex with fMRI, it is still debated whether any information at the subvoxel level (e.g., within-voxel biases in orientation-specific columnar responses) contributes to the decodable signal (Op de Beeck, 2010). The interpretation of the source of decodable signals in neuroimaging remains one of the central challenges facing the application of MVPA techniques to advancing our understanding of information processing in the human brain (de-Wit et al., 2016).

Acknowledgments

This research was supported by an Australian Research Council Future Fellowship (FT120100816) and an Australian Research Council Discovery project (DP160101300) awarded to T. A. C. S. G. W. was supported by an Australian NHMRC Early Career Fellowship (APP1072245).

Reprint requests should be sent to Tijl Grootswagers, Department of Cognitive Science, Australian Hearing Hub, 16 University Avenue, Macquarie University, NSW 2109 Australia, or via e-mail: t.grootswagers@gmail.com.

Notes

1. The main study consisted of two conditions, stimuli in a clear or degraded state; however, for the purpose of this article, we only use the data for stimuli in the clear state (normal photographs of objects).
2. Other MEG systems also include magnetometers, and there are possible differences in decodability from gradiometers and magnetometers (Kaiser et al., 2016).
3. Note that when comparing PCA performed inside the cross-validation loop on separate time points with PCA performed before the cross-validation on all time points, we did not find any difference in classifier accuracy (data not shown), but this may not hold for different data sets.

REFERENCES

Alink, A., Krugliak, A., Walther, A., & Kriegeskorte, N. (2013). fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli. *Frontiers in Psychology, 4,* 493.

Allefeld, C., Görgen, K., & Haynes, J.-D. (2016). Valid population inference for information-based imaging: From the second-level *t* test to prevalence inference. *Neuroimage, 141,* 378–392.

Allison, B. Z., Wolpaw, E. W., & Wolpaw, J. R. (2007). Brain–computer interface systems: Progress and prospects. *Expert Review of Medical Devices, 4,* 463–474.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, Methodological, 57,* 289–300.

Bennett, C. M., Baird, A., Miller, M. B., & Wolford, G. L. (2011). Neural correlates of interspecies perspective taking in the post-mortem Atlantic salmon: An argument for proper multiple comparisons correction. *Journal of Serendipitous and Unexpected Results, 1,* 1–5.

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience, 4,* 417–422.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4). New York: Springer.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K.-R. (2011). Single-trial analysis and classification of ERP components—A tutorial. *Neuroimage, 56,* 814–825.

Bode, S., Sewell, D. K., Lilburn, S., Forte, J. D., Smith, P. L., & Stahl, J. (2012). Predicting perceptual decision biases from early brain activity. *Journal of Neuroscience, 32,* 12488–12498.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Carlson, T. A. (2014). Orientation decoding in human visual cortex: New insights from an unbiased perspective. *Journal of Neuroscience, 34,* 8373–8383.

Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision, 11,* 9.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience, 15,* 704–717.

Carlson, T. A., Simmons, R. A., Kriegeskorte, N., & Slevc, L. R. (2013). The emergence of semantic meaning in the ventral temporal pathway. *Journal of Cognitive Neuroscience, 26,* 120–131.

Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision, 13,* 1.

Carlson, T. A., & Wardle, S. G. (2015). Sensible decoding. *Neuroimage, 110,* 217–218.

Cauchoix, M., Arslan, A. B., Fize, D., & Serre, T. (2012). The neural dynamics of visual processing in monkey extrastriate cortex: A comparison between univariate and multivariate techniques. In G. Langs, I. Rish, M. Grosse-Wentrup, & B. Murphy (Eds.), *Machine learning and interpretation in neuroimaging* (pp. 164–171). Berlin/Heidelberg: Springer.

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The neural dynamics of face detection in the wild revealed by MVPA. *Journal of Neuroscience, 34,* 846–854.

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2010). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *Neuroimage, 54,* 3028–3039.

Chumbley, J. R., & Friston, K. J. (2009). False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage, 44,* 62–70.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17,* 455–462.

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex, 26,* 3563–3579.

Cichy, R. M., Ramirez, F. M., & Pantazis, D. (2015). Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *Neuroimage, 121,* 193–204.

Clifford, C. W. G., & Mannion, D. J. (2015). Orientation decoding: Sense in spirals? *Neuroimage, 110,* 219–222.

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. Cambridge, MA: MIT Press.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19,* 261–270.

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,* 326–336.

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage, 43,* 44–58.

de-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic Bulletin & Review, 23,* 1415–1428.

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology, 107,* 78–89.

Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.

Duncan, K. K., Hadjipapas, A., Li, S., Kourtzi, Z., Bagshaw, A., & Barnes, G. (2010). Identifying spatially overlapping local cortical networks with MEG. *Human Brain Mapping, 31,* 1003–1016.

Edelman, S., Grill-Spector, K., Kushnir, T., & Malach, R. (1998). Toward direct visualization of the internal shape representation space by fMRI. *Psychobiology, 26,* 309–321.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology, 70,* 510–523.

Formisano, E., De Martino, F., & Valente, G. (2008). Multivariate analysis of fMRI time series: Classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging, 26,* 921–934.

Freeman, J., Brouwer, G. J., Heeger, D. J., & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *Journal of Neuroscience, 31,* 4792–4804.

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience, 16,* 974–981.

Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows. *Neuroimage, 128,* 385–397.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Brain Imaging Methods, 7,* 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage, 86,* 446–460.

Guimaraes, M. P., Wong, D. K., Uy, E. T., Grosenick, L., & Suppes, P. (2007). Single-trial classification of MEG recordings. *IEEE Transactions on Biomedical Engineering, 54,* 436–443.

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing, 32,* 35–42.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics, 7,* 37–53.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Fründ, I., Rieger, J. W., et al. (2009). PyMVPA: A unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics, 3,* 3.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage, 87,* 96–110.

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

Haynes, J.-D. (2015). A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives. *Neuron, 87,* 257–270.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7,* 523–534.

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics, 8,* 88.

Hill, N. J., Lal, T. N., Schröder, M., Hinterberger, T., Widman, G., Elger, C. E., et al. (2006). Classifying event-related desynchronization in EEG, ECoG and MEG signals. In K. Franke, K.-R. Müller, B. Nickolay, & R. Schäfer (Eds.), *Pattern recognition* (pp. 404–413). Berlin/Heidelberg: Springer.

Hogendoorn, H., Verstraten, F. A. J., & Cavanagh, P. (2015). Strikingly rapid neural basis of motion-induced position shifts revealed by high temporal-resolution EEG pattern classification. *Vision Research, 113,* 1–10.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology, 111,* 91–102.

Jackson, J. E. (1991). *A user's guide to principal components* (Vol. 587). New York: John Wiley & Sons.

Jafarpour, A., Horner, A. J., Fuentemilla, L., Penny, W. D., & Duzel, E. (2013). Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia, 51,* 772–780.

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology, 115,* 2246–2250.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8,* 679–685.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences, 18,* 203–210.

King, J.-R., Gramfort, A., Schurger, A., Naccache, L., & Dehaene, S. (2014). Two distinct dynamic modes subtend the detection of unexpected sounds. *PLoS ONE, 9,* e85791.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in Psychtoolbox-3. *Perception, 36,* 1.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences, 17,* 401–412.

Kriegeskorte, N., Lindquist, M. A., Nichols, T. E., Poldrack, R. A., & Vul, E. (2010). Everything you never wanted to know about circular analysis, but were afraid to ask. *Journal of Cerebral Blood Flow & Metabolism, 30,* 1551–1557.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis—Connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., et al. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60,* 1126–1141.

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

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychological Bulletin, 127,* 358–375.

Lemm, S., Blankertz, B., Dickhaus, T., & Müller, K.-R. (2011). Introduction to machine learning for brain imaging. *Neuroimage, 56,* 387–399.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.

Mannion, D. J., McDonald, J. S., & Clifford, C. W. G. (2009). Discrimination of the local orientation structure of spiral Glass patterns early in human visual cortex. *Neuroimage, 46,* 511–515.

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *Neuroimage, 67,* 111–118.

Meyers, E. M. (2013). The neural decoding toolbox. *Frontiers in Neuroinformatics, 7,* 8.

Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology, 100,* 1407–1419.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage, 53,* 103–118.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology, 48,* 229–240.

Müller, K., Anderson, C. W., & Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11,* 165–169.

Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *Journal of Neuroscience Methods, 167,* 82–90.

Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—An introductory guide. *Social Cognitive and Affective Neuroscience, 41,* 101–109.

Murphy, B., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language, 117,* 12–22.

Naselaris, T., & Kay, K. N. (2015). Resolving ambiguities of MVPA using explicit models of representation. *Trends in Cognitive Sciences, 19,* 551–554.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage, 56,* 400–410.

Nichols, T. E. (2012). Multiple testing corrections, nonparametric methods, and random field theory. *Neuroimage, 62,* 811–815.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping, 15,* 1–25.

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLOS Computational Biology, 10,* e1003553.

Noirhomme, Q., Lesenfants, D., Gomez, F., Soddu, A., Schrouff, J., Garraux, G., et al. (2014). Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions. *Neuroimage: Clinical, 4,* 687–694.

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,* 424–430.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011,* 156869.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in MATLAB/GNU octave. *Frontiers in Neuroinformatics, 10,* 27.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *Neuroimage, 49,* 1943–1948.

Pantazis, D., Nichols, T. E., Baillet, S., & Leahy, R. M. (2005). A comparison of random field theory and permutation methods for the statistical analysis of MEG data. *Neuroimage, 25,* 383–394.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

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

Pratte, M. S., Sy, J. L., Swisher, J. D., & Tong, F. (2016). Radial bias is not necessary for orientation decoding. *Neuroimage, 127,* 23–33.

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of Cognitive Neuroscience, 28,* 680–692.

Ramkumar, P., Jas, M., Pannasch, S., Hari, R., & Parkkonen, L. (2013). Feature-specific information processing precedes concerted activation in human visual cortex. *Journal of Neuroscience, 33,* 7691–7699.

Redcay, E., & Carlson, T. (2015). Rapid neural discrimination of communicative gestures. *Social Cognitive and Affective Neuroscience, 10,* 545–551.

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging object representations in the visual system predict reaction times for categorization. *PLOS Computational Biology, 11,* e1004316.

Sandberg, K., Bahrami, B., Kanai, R., Barnes, G. R., Overgaard, M., & Rees, G. (2013). Early visual responses predict conscious face perception within and between subjects during binocular rivalry. *Journal of Cognitive Neuroscience, 25,* 969–985.

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., & Desain, P. (2010). Name that tune: Decoding music from the listening brain. *Neuroimage, 56,* 843–849.

Schreiber, K., & Krekelberg, B. (2013). The statistical analysis of multi-voxel patterns in functional imaging. *PLoS ONE, 8,* e69328.

Schwarzkopf, D. S., & Rees, G. (2011). Pattern classification using functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews: Cognitive Science, 2,* 568–579.

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, J. S., Oosterhof, N. N., Halchenko, Y. O., et al. (2015). The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience, 27,* 665–678.

Simanova, I., van Gerven, M., Oostenveld, R., & Hagoort, P. (2010). Identifying object categories from event-related EEG: Toward decoding of conceptual representations. *PLoS ONE, 5,* e14465.

Simanova, I., van Gerven, M. A. J., Oostenveld, R., & Hagoort, P. (2015). Predicting the semantic category of internally generated words from neuromagnetic recordings. *Journal of Cognitive Neuroscience, 27,* 35–45.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage, 44,* 83–98.

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *Neuroimage, 65,* 69–82.

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron, 78,* 364–375.

Sudre, G., Pomerleau, D., Palatucci, M., Wehbe, L., Fyshe, A., Salmelin, R., et al. (2012). Tracking neural coding of perceptual and semantic features of concrete nouns. *Neuroimage, 62,* 451–463.

Tangermann, M., Müller, K.-R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., et al. (2012). Review of the BCI competition IV. *Neuroprosthetics, 6,* 55.

Thirion, B., Pedregosa, F., Eickenberg, M., & Varoquaux, G. (2015). Correlations of correlations are not reliable statistics: Implications for multivariate pattern analysis. In *ICML Workshop on Statistics, Machine Learning and Neuroscience (Stamlins 2015)*. Retrieved from https://hal.inria.fr/hal-01187297/.

van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van Gerven, M. A. J. (2013). MEG-based decoding of the spatiotemporal dynamics of visual category perception. *Neuroimage, 83,* 1063–1073.

van Gerven, M. A. J., Maris, E., Sperling, M., Sharan, A., Litt, B., Anderson, C., et al. (2013). Decoding the memorization of individual stimuli with direct human brain recordings. *Neuroimage, 70,* 223–232.

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering, 44,* 867–880.

Vanrullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Perception Science, 2,* 365.

Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual Review of Biophysics and Bioengineering, 2,* 157–180.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage, 137,* 188–200.

Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S.-M., & Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *Neuroimage, 132,* 59–70.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1,* 80–83.

Wolff, M. J., Ding, J., Myers, N. E., & Stokes, M. G. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience, 9,* 123.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology, 113,* 767–791.

Zhang, Y., Meyers, E. M., Bichot, N. P., Serre, T., Poggio, T. A., & Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 108,* 8850–8855.