eNeuro

Research Article: New Research | Sensory and Motor Systems

The Role of Visual and Semantic Properties in the Emergence of Category-Specific Patterns of Neural Response in the Human Brain

Role of Visual and Semantic Properties in Neural Response

David D. Coggan, Daniel H. Baker and Timothy J. Andrews

Department of Psychology and York Neuroimaging Centre, University of York, York, YO10 5DD, United Kingdom

DOI: 10.1523/ENEURO.0158-16.2016

Received: 13 June 2016

Accepted: 27 June 2016

Published: 21 July 2016

Funding: Wellcome Trust
100004440
105624

Conflict of Interest: Authors report no conflict of interest.

D.C., D.H.B., and T.A. designed research; D.C., D.H.B., and T.A. performed research; D.C., D.H.B., and T.A. analyzed data; D.C., D.H.B., and T.A. wrote the paper.

Wellcome Trust, 100004440 105624

Correspondence should be addressed to either Timothy J. Andrews, E-mail: timothy.andrews@york.ac.uk; or Daniel H. Baker, E-mail: daniel.baker@york.ac.uk

Cite as: eNeuro 2016; 10.1523/ENEURO.0158-16.2016

THE ROLE OF VISUAL AND SEMANTIC PROPERTIES IN THE EMERGENCE OF CATEGORY-SPECIFIC PATTERNS OF NEURAL RESPONSE IN THE HUMAN BRAIN

Short Title: ROLE OF VISUAL AND SEMANTIC PROPERTIES IN NEURAL RESPONSE

David D. Coggan, Daniel H. Baker* & Timothy J. Andrews*

Department of Psychology and York Neuroimaging Centre,

University of York, York, YO10 5DD, United Kingdom

*Corresponding authors:

timothy.andrews@york.ac.uk

daniel.baker@york.ac.uk

Figures: 6

Abstract: 235

Introduction: 647

Discussion: 828

Conflict of interest: The authors declare no competing financial interests.

Acknowledgements: Supported in part by the Wellcome Trust (ref: 105624) through the Centre for Chronic Diseases and Disorders (C2D2) at the University of York.

ABSTRACT

Brain imaging studies have found distinct spatial and temporal patterns of response to different object categories across the brain. However, the extent to which these categorical patterns of response reflect higher-level semantic or lower-level visual properties of the stimulus remains unclear. To address this question, we measured patterns of EEG response to intact and scrambled images in the human brain. Our rationale for using scrambled images is that they have many of the visual properties found in intact images, but do not convey any semantic information. Images from different object categories (bottle, face, house) were briefly presented (400 msec) in an event-related design. A multivariate pattern analysis (MVPA) revealed categorical patterns of response to intact images emerged ~80-100 msec after stimulus onset and were still evident when the stimulus was no longer present (~800 msec). Next, we measured patterns of response to scrambled images. Categorical patterns of response to scrambled images also emerged ~80-100 msec after stimulus onset. However, in contrast to the intact images, distinct patterns of response to scrambled images were mostly evident while the stimulus was present (~400 msec). Moreover, scrambled images were only able to account for all the variance in the intact images at early stages of processing. This direct manipulation of visual and semantic content provides new insights into the temporal dynamics of object perception and the extent to which different stages of processing are dependent on lower-level or higher-level properties of the image.

Significance Statement

Previous studies have shown distinct spatial and temporal patterns of response to different object categories. However, the extent to which these patterns are based on lower-level visual properties compared to high-level semantic information remains unclear. To address this question, we used scrambled objects that preserve visual properties, but do not convey any semantic information. We found distinct patterns of response to intact images from different object categories. Patterns of response to scrambled images from different categories emerge in a similar way to intact images but do not persist for the same duration. These findings demonstrate the relative importance of both lower-level visual and higher-level semantic properties in the neural response to objects.

INTRODUCTION

A variety of evidence has shown that spatially distinct regions of the ventral occipito-temporal cortex are selective for different categories of objects (Kanwisher, 2010). Lesions to this region often result in difficulties in recognizing and naming specific object categories (Farah, 1990; McNeil and Warrington, 1993; Moscovitch et al., 1997). The notion that discrete areas of the temporal lobe are specialized for different categories of objects receives support from functional imaging studies that show that some regions in the temporal lobe are more responsive to faces than to other categories (Allison et al., 1994; Kanwisher et al., 1997). Other imaging studies have found similar category-specific responses for inanimate objects (Malach et al., 1995), scenes (Epstein and Kanwisher, 1998) and human body parts (Downing et al., 2001). Although specialized regions have only been reported for a limited number of objects, the spatial pattern of response across the entire ventral stream can distinguish a wider range of categories (Haxby et al., 2001; Kriegeskorte et al., 2008; Connolly et al., 2012; Clarke and Tyler, 2014; Rice et al., 2014).

A full understanding of object perception requires the ability to discriminate object-specific brain states with both spatial *and* temporal resolution. Recently, reliable patterns of neural response to images from different object categories have been shown with MEG and EEG (Carlson et al., 2011, 2013; Cichy et al., 2014; Cauchoix et al., 2014; Clarke et al., 2015). These techniques complement previous MRI studies by providing temporal information about when these categorical patterns of response emerge and how long they are sustained. Temporal properties are important, as they place constraints on models of object recognition (Mur and Kriegeskorte, 2014). Such models suggest a dynamic process in which there is a transformation from a visual representation (based on the statistics of the image) to a semantic representation (reflecting the meaning of the object; Clarke and Tyler,

2015). It is thought that the initial component of the response reflects fast feed-forward processing that is related to visual properties, whereas later patterns reflect recurrent processing that might be related to semantic properties of the stimulus (Lamme and Roelfsema, 2000; Hochstein and Ahissar, 2002; Bar et al., 2006; DiCarlo and Cox, 2007).

The aim of this study was to investigate the relative importance of visual and semantic properties of objects in the emergence of categorical patterns of neural response. However, a fundamental problem in this endeavour is that the visual and semantic properties of objects often covary, making it difficult to resolve the relative contribution of these sources of information to patterns of neural response. So, it is not clear from many previous studies whether the distinct patterns of response to different object categories reflect visual or semantic properties (Carlson et al., 2011, 2013; Cichy et al., 2014; Cauchoix et al., 2014). In a recent MEG study, Clarke and colleagues (2015) addressed this issue by showing that the categorization of objects based on the neural response could be predicted by the visual properties of the image. However, they also found that accuracy could enhanced by including semantic properties, particularly at later stages of processing. Although this suggests that visual and semantic properties are both important for the neural representation of objects, this approach does not able to show a causal link.

To address this issue, we measured patterns of EEG response to intact images from different object categories, as well as versions of these images that had been phase-scrambled on a global or local basis. Our rationale for using scrambled images is that they have many of the visual properties found in intact images, but they do not convey any semantic information (Coggan et al., 2016). This allows us to determine the extent to which the preserved visual properties contribute to the neural representation of objects in the absence of any semantic content. The comparison between the locally-scrambled and

106 globally-scrambled images also allows us to explore the importance of spatial image
107 properties, which are preserved in the locally-scrambled condition. In a recent fMRI study,
108 we found similar spatial patterns of response to intact and scrambled images across the
109 ventral visual pathway (Coggan et al., 2016). This study demonstrated the importance of
110 low-level visual properties in the patterns of response in the ventral visual pathway. By
111 comparing the similarity of the responses to intact and scrambled images using EEG, we aim
112 to determine the relative contribution of visual properties to categorical patterns of
113 response at different time-points.
114

MATERIALS AND METHODS

Stimuli

105 images of three object categories (face, bottle, house) were taken from an object image stimulus set (Rice et al., 2014). All images were gray-scale, superimposed on a mid-gray background, and had a resolution of 400x400 pixels (Figure 1). For each of these original images, two different phase-scrambled versions were generated. A global-scrambling method involved a typical Fourier-scramble, i.e. keeping the global power of each two-dimensional frequency component constant while randomizing the phase of the components. A local-scrambling method involved windowing the original image into an 8x8 grid and applying the phase-scramble to each 50x50 pixel window independently. In a previous study, we showed that these scrambling effectively removes any semantic or categorical content in the images (Coggan et al., 2016). Stimuli were presented using a gamma corrected VIEWPixx display (VPixx Technologies Inc., Quebec, Canada) with a resolution of 1920x1200 pixels and a refresh rate of 120Hz. Images were viewed at a distance of approximately 57cm and subtended a retinal angle of 8°.

Participants

Twenty participants (3 males, mean age = 20.6, SD = 2.6 years) with normal or corrected-to-normal vision took part in the experiment. Participants gave written, informed consent. The study was approved by the University of York Department of Psychology Ethics Committee. The data for one participant (female) was removed from the analysis due to partial data loss.

Design and Procedure

The experiment involved 3 runs: The first run contained globally scrambled images, the second run contained locally scrambled images and the third run contained intact images. Therefore, participants were unaware of the object categories in our stimulus set prior to viewing the scrambled images. Each run contained 35 blocks. There were 10 trials in each block. In each trial, an image from one of the three object categories was presented for 400ms. There was a jittered inter-trial interval that had a mean duration of 1 second and a standard deviation of 200ms. The duration of the inter-block interval was 3 seconds. Participants fixated a cross in the center of the screen between trials. To maintain attention, participants were instructed to click a mouse whenever a red dot appeared on an image. One image in each block contained a red dot. Self-timed rests were taken between runs.

EEG Recording

EEG waveforms were recorded from 64 scalp locations laid out according to the 10/20 system in a WaveGuard cap (ANT Neuro, Netherlands). Data from each electrode was referenced against a whole-head average. We also monitored blinks through bipolar electrooculogram electrodes placed above and below the left eye. Signals were amplified and digitised at 1000Hz and recorded using the ANT Neuroscan software (ANT Neuro, Netherlands). Stimulus-contingent triggers were sent from the VIEWPixx device to the EEG amplifier using a 25-pin parallel port with microsecond-accurate synchronisation to the display refresh sequence. The PsychToolbox routines (Brainard, 1997; Pelli, 1997) running in Matlab were used to control the display hardware and send triggers.

EEG Pre-processing

The EEG traces from each run were concatenated and band-pass filtered between 0.01-30Hz prior to epoching. Blink artefacts were corrected using independent components analysis (ICA). This involved running ICA across data from all electrodes including the vEOG, and manually selecting the component(s) that captured blink artefacts. These components were then subtracted from the EEG trace at each electrode site according to their weighting. This approach meant that no trials were rejected. The EEG trace was then divided into epochs ranging from 200ms before stimulus onset to 800ms after stimulus onset. All trials containing a red dot were removed prior to further analysis.

EEG MVPA Analysis

All data processing was performed in Matlab using custom scripts. To measure the spatial patterns of EEG response for each participant, trials were collapsed into mean ERPs for odd and even trials for each condition and at each electrode site. These condition-averaged ERPs were then baselined by subtracting the mean amplitude during the 200ms prior to stimulus onset (across both odd and even trials) from the response at each time-point. From these ERPs, a 64-value vector representing the spatial pattern of response across all electrodes was extracted for odd and even trials for each object category at each time-point.

Pattern vectors were normalized within each participant using the following method. First, vectors were selected from one time-point and one image type. This gave a total of 6 patterns (odd/even x face/bottle/house). For each electrode site, the mean amplitude across all 6 patterns was subtracted from its amplitude in each pattern. This process was repeated for each image type at each time-point.

To see whether different object categories evoke distinct patterns of EEG response, we ran a correlation-based MVPA separately for each image type and time-point (Figure 2A).

This involved measuring the correlation between pattern vectors within and between the three object categories. For within-category correlations (e.g. face vs face), we measured the correlation between odd and even trials. For between-category correlations (e.g. bottle vs house), we used the mean correlation between odd trials of the first category and even trials of the second, and between even trials of the first category and odd trials of the second. The distinctiveness of the patterns of EEG response was then measured by subtracting between-category correlations from within-category correlations. 95% confidence intervals for this difference were then obtained by bootstrapping across participants. Points at which different object categories evoked significantly distinct patterns of EEG response were defined by the lower confidence interval being above zero.

To measure the similarity between responses to intact and scrambled images from the same object category, we first collapsed patterns across odd and even trials to create one pattern per condition per time-point. We then correlated the patterns of response at each time-point separately for the intact-locally scrambled and intact-globally scrambled contrasts for each category. A group mean was calculated across categories and 95% confidence intervals were obtained by bootstrapping across participants.

To determine whether the response to intact images could be explained by the response to scrambled images, we calculated a noise ceiling. This estimates that maximum correlation that could be expected. The noise ceiling was calculated by measuring the correlation between the responses at odd and even trials within each category in the intact condition. At the individual level, we take a mean of the within-category correlations (face-face, bottle-bottle, house-house) for each timepoint. We then average across subjects to obtain one noise ceiling estimate at each timepoint. Timepoints at which this value fell

within the 95% CI for the correlation between intact and scrambled images demonstrate when all the variance in the intact images was explained by the scrambled images.

The correlation-based method was complemented with a classification-based approach involving a support vector machine, producing similar results. To see whether different object categories evoked distinct patterns of response, classification was performed separately for each participant, image type and time-point (Figure 6). First, patterns of EEG response were extracted for each trial of each category. Two 'training' patterns and one 'testing' pattern for each category were generated by randomly dividing the 105 trials into three equal sets and taking an average. A support vector machine was then trained on the six training patterns, and tested on the three testing patterns. This procedure was repeated 100 times, with different subsets of trials used for training and test in each iteration. To see whether similar patterns of response were evoked by intact and scrambled images from the same category, the classifier was altered so that test patterns were substituted with those from another image type. This was performed for each pairwise contrast between image types, and accuracy was averaged across both directions (e.g. train on intact, test on locally scrambled; and train on locally scrambled, test on intact).

Finally, to examine transient and persistent neural activity in response to each condition, we conducted a temporal cross-correlation. This involved measuring the correlation between response patterns for odd and even trials for the same condition, iterating over each possible pair of time-points. Correlations were represented in a 1000 x 1000 similarity matrix and data were averaged across the positive diagonal. Matrices were then collapsed across categories to give one matrix per image type.

RESULTS

First, we asked whether different intact object categories produced distinct spatial patterns of EEG response (Figure 2). To address this question, we compared the similarity of patterns of response to images from the same category (e.g. face vs face) with the similarity of patterns to images of different categories (e.g. face vs house). Categorical patterns of response were demonstrated when the within-category correlations were significantly greater than the between-category correlations. Categorical patterns of response to intact images emerged 80 msec after stimulus onset. The patterns were maximally distinct at about 150 msec and persisted until at least 800 msec (Figure 2B). A classification-based approach was then used to complement the correlation-based method. In this analysis, a classifier was trained on a subset of the data and tested on the remaining data. This showed a similar pattern to the correlation-based analysis. Above chance accuracy emerged 80 msec after stimulus onset, peaked at about 150 msec and persisted until 800 msec (Figure 3A).

To measure the extent to which these category-specific patterns of response were based on lower-level visual properties, we first asked whether locally scrambled and globally scrambled images also produced distinct category-specific patterns of EEG response using both the correlation-based (Figure 2C-D) and classification (Figure 3B-C) analyses. Distinct category-specific patterns of response for locally scrambled images emerged after about 80 msec after stimulus onset. They were maximally distinct at about 110 msec and persisted until about 400-500 msec. Distinct category-specific patterns of response for globally scrambled images, emerged at about 100 msec after stimulus onset. They were maximally distinct at about 190 msec and persisted until about 300 msec.

Although distinct patterns of response were evident to scrambled images from different categories (i.e. within-category > between-category correlations), it is not clear

whether the patterns were similar to those elicited from the intact images. To address this question, we correlated patterns of response to the same object category across different levels of scrambling at different time points. Figure 4A (blue horizontal bar) shows that the correlation between intact and locally-scrambled images became significant at about 80 msec after stimulus onset, peaked at about 110 msec and 190 msec. The percentage duration that the locally-scrambled patterns were correlated with the intact patterns was greater during the stimulus period (0-400 msec: 27%) compared to the post-stimulus period (400-800 msec: 10%). A similar pattern of results was evident when we trained a classifier on intact or locally-scrambled images and then tested on locally-scrambled or intact images, respectively (Fig. 5A). The duration of above chance accuracy with the locally-scrambled and intact conditions was similar during the stimulus period (0-400 msec: 40%) and the post-stimulus period (400-800 msec: 49%).

Next, we explored the similarity between the intact and globally-scrambled images (Fig. 4B, 5B). The correlation between responses to intact and globally scrambled images became significant (blue horizontal bar) about 90 msec after stimulus onset, peaked at about 110 msec and persisted until around 120 msec. The percentage duration that the locally-scrambled patterns were correlated with the intact patterns was greater during the stimulus period (0-400 msec: 4%) compared to the post-stimulus period (400-800 msec: 0%). A similar pattern of results was evident when we trained a classifier on intact or locally-scrambled images and then tested on locally-scrambled or intact images, respectively (Fig. 5A). The duration of above chance accuracy with the locally-scrambled and intact conditions was greater during the stimulus period (0-400 msec: 4%) compared to the post-stimulus period (400-800 msec: 0%).

To directly compare similarity between intact images and either locally-scrambled or globally-scrambled images, the average correlation (Fig. 4) or accuracy (Fig. 5) was compared across individuals. The average correlation between intact and locally scrambled images was significantly higher than the correlation between intact and globally scrambled images ($t(18) = 3.29$, $p < .005$). Similarly, the average accuracy (see Fig. 5) with intact and locally-scrambled images was significantly higher than with intact and globally-scrambled images ($t(18) = 5.34$, $p < .0001$).

We then asked whether the explainable variance in intact responses was fully accounted for by the responses to scrambled images, given the level of noise in the data. This was achieved by calculating a noise ceiling (Nili et al., 2014). This involved measuring the correlation to intact images from the same category across odd and even trials of the same category. The noise ceiling was not fixed, but varied across time. We then determined whether the correlation between intact and scrambled images was not significantly different to the noise ceiling for each time-point. For locally-scrambled images, the 95% confidence intervals of the correlations overlapped until approximately 120 msec after stimulus onset (Fig. 4A). The percentage duration that the locally-scrambled patterns were not significantly different from the noise ceiling was similar during the stimulus (0-400 msec: 9%) and post-stimulus period (400-800 msec: 9%). For globally scrambled images the confidence intervals overlapped until about 100 msec after stimulus onset (Fig 4B). The percentage duration that the globally-scrambled patterns were not significantly different from the noise ceiling was greater during the stimulus period (0-400 msec: 1%) compared to the post-stimulus period (400-800 msec: 0%).

Finally, we investigated the stability of the category-specific patterns of response for each image manipulation (Cichy et al., 2014). This involved measuring the correlation

between patterns of EEG response within each condition across different time-points. The results were then averaged across categories for each image type and represented in time-time similarity matrices (Figure 6). Here, the diagonal for intact images corresponds to the noise-ceiling estimate used in Figure 4. For intact images, the pattern of response from 100-150 msec was positively correlated with patterns found from ~250-600 msec. The continuation of this neural activity far beyond stimulus offset suggests that this does not reflect prolonged visual input during image presentation. The locally scrambled matrix shows no evidence of persistent neural activity as seen in the intact matrix, but does exhibit transient neural activity between ~100-250ms after stimulus onset. Interestingly, time-point combinations of ~150ms and ~200ms show negative correlations, suggesting a polarity reversal in the potentials between these latencies. The globally scrambled matrix shows weak correlations across all combinations of time-points.

DISCUSSION

The aim of this study was to determine the contribution of lower-level visual and higher-level semantic properties to the emergence of categorical patterns of neural response. To address this question, we compared patterns of EEG response to intact and scrambled images from different object categories. Scrambled images were used, because they contain similar visual properties to intact images but do not convey any semantic information (Coggan et al., 2016). Our results show similar category-specific patterns of response at early stages of processing. However, these patterns were sustained for a longer time with intact images compared to scrambled images. These results show the importance of visual properties in the emergence of categorical patterns of response, but also show the importance of semantic properties in the recurrent processing that sustains these patterns.

The emergence of category-specific patterns of EEG response to intact images is comparable to previous studies using MEG that found categorical distinctions can be decoded prior to 100ms after stimulus onset and become maximally distinct at about 140 msec (Carlson et al., 2013; Cichy et al., 2014; Cauchoix et al., 2014). However, most previous studies have not directly determined whether these patterns of response reflect lower-level visual properties or higher-level semantic properties of the image. Recently, Clarke and colleagues (2015) addressed this issue with MEG showing that visual properties can explain patterns of response to different categories of objects. However, they also showed that the semantic properties of objects were able to explain additional variance in the pattern of response particularly at later stages of the response. In our study, we were also able to show that the patterns of response to images from different object categories are driven predominantly by the lower-level visual properties at early stages of visual

processing (up to 150 msec). Visual properties were also able to partially account for the variance in the response to intact images at later stages of processing.

Patterns of response to intact images were correlated more strongly and for longer with responses to locally scrambled images than with globally scrambled images. One key difference between these two conditions is that the spatial properties, such as the shape (or spatial envelope) of the image, are somewhat preserved in the locally scrambled images, but not in the globally scrambled images. In a recent fMRI study, we showed that the spatial pattern of response in the ventral stream to different categories of intact objects was more similar to the pattern elicited by locally scrambled objects compared to globally scrambled objects. The greater similarity between responses to intact locally scrambled images is consistent with previous studies that have shown a modulatory effect of spatial properties on patterns of response in the ventral visual pathway (Levy et al., 2001; Golomb and Kanwisher, 2012; Bracci and Op de Beeck, 2015; Silson et al., 2015; Watson et al., 2016).

Although lower-level image properties account for the majority of the variance in responses to intact images at early stages, there remains a significant amount of variance to be explained at later stages of processing. For example, although category-specific patterns of response to intact images persisted well beyond the duration of the stimulus, patterns of response to scrambled images were only evident when the stimulus was present. The persistence of these neural responses to intact images suggests an important role for recurrent processing of the image, which is likely to be driven by top-down semantic representations (Lamme and Roelfsema, 2000; DiCarlo and Cox, 2007; Kriegeskorte et al., 2008; Naselaris et al., 2009; Connolly et al., 2012; Mur and Kriegeskorte, 2014). Indeed, Clarke and colleagues (2015) showed that accuracy in categorization using MEG data was enhanced by combining visual and semantic models.

366 It is also possible that differences in the patterns of response between intact and
367 scrambled images reflect sensitivity to image properties that are disrupted by either
368 scrambling process. An important property of natural images is that they contain strong
369 statistical dependencies, such as location-specific combinations of orientation and spatial
370 frequency corresponding to image features such as edges (Marr and Hildreth, 1980). Indeed,
371 the character and extent of these statistical dependencies is likely to be diagnostic for
372 different classes of images (O'Toole et al., 2005; Rice et al., 2014). The scrambling
373 procedure disrupts many of the statistical relationships between the elements. So, it is
374 possible that image manipulations that can preserve these higher-level visual properties
375 (Freeman and Simoncelli, 2011) might generate responses that are more similar to the
376 intact images. Indeed, it is possible that neural representations underlying higher-level
377 visual properties and the corresponding semantic properties that they convey may be the
378 same.

379 In conclusion, we have found that distinct category-specific patterns of neural
380 response emerge at about 80 msec after stimulus onset and can persist for at least 800
381 msec. Using scrambled images, we show that early stages of these category-specific
382 patterns can be explained by lower-level image properties. However, the differences in the
383 neural responses to intact and scrambled images at later stages of processing also reveal
384 the importance of higher-level semantic properties.

REFERENCES

Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD (1994) Face recognition in human extrastriate cortex. J Neurophysiol 71:821–825.

Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E (2006) Top-down facilitation of visual recognition. Proc Natl Acad Sci U S A 103:449–454.

Bracci S, Op de Beeck H (2015) Dissociations and associations between shape and category representations in the two visual pathways. J Neurosci 36:432–444.

Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10:433–436.

Carlson TA, Tovar D, Alink A, Kriegeskorte N (2013) Representational dynamics of object vision: The first 1000 ms. J Vis 13:1–19.

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J (2011) High temporal resolution decoding of object position and category. J Vis 11:1–17.

Cauchoix M, Barragan-Jason G, Serre T, Barbeau EJ (2014) The neural dynamics of face detection in the wild revealed by MVPA. J Neurosci 34:846–854.

Cichy RM, Pantazis D, Oliva A (2014) Resolving human object recognition in space and time. Nat Neurosci 17:455–462.

Clarke A, Devereux BJ, Randall B, Tyler LK (2015) Predicting the time course of individual objects with MEG. Cereb Cortex 25:3602–3612.

Clarke A, Tyler L (2014) Object-Specific Semantic Coding in Human Perirhinal Cortex. J Neurosci 34:4766–4775.

Coggan DD, Liu W, Baker DH, Andrews TJ (2016) Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. Neuroimage, 135: 107-114.

409 Connolly AC, Guntupalli JS, Gors J, Hanke M, Halchenko YO, Wu Y-C, Abdi H, Haxby JV.
410 (2012) The Representation of Biological Classes in the Human Brain. J Neurosci
411 32:2608–2618.
412 DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cogn Sci 11:333–
413 341.
414 Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual
415 processing of the human body. Science 293:2470–2473.
416 Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment.
417 Nature 392:598–601.
418 Farah MJ (1990) Visual Agnosia. Cambridge, MA: MIT Press.
419 Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. Nat Neurosci 14:1195–
420 1201.
421 Golomb JD, Kanwisher N (2012) Higher level visual cortex represents retinotopic, not
422 spatiotopic, object location. Cereb Cortex 22:2794–2810.
423 Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and
424 overlapping representations of faces and objects in ventral temporal cortex. Science
425 293:2425–2430.
426 Hochstein S, Ahissar M (2002) View from the top: Hierarchies and reverse hierarchies in the
427 visual system. Neuron 36:791–804.
428 Kanwisher N (2010) Functional specificity in the human brain: a window into the functional
429 architecture of the mind. Proc Natl Acad Sci U S A 107:11163–11170.
430 Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human
431 extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.
432 Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA (2008)

433 Matching Categorical Object Representations in Inferior Temporal Cortex of Man and
434 Monkey. Neuron 60:1126–1141.
435 Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and
436 recurrent processing. Trends Neurosci 23:571–579.
437 Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center – periphery organization of
438 human object areas. Nat Neurosci 4:533–539.
439 Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ,
440 Rosen BR, Tootell RBH (1995) Object-related activity revealed by functional magnetic
441 resonance imaging in human occipital cortex. Neurobiology 92:8135–8139.
442 Marr D, Hildreth E (1980) Theory of Edge Detection. Proc R Soc London Ser B, Biol Sci
443 207:187–217.
444 McNeil JE, Warrington EK (1993) Prosopagnosia: A face-specific disorder. Q J Exp Psychol
445 Sect A 46:1–10.
446 Moscovitch M, Winocur G, Behrmann M (1997) What Is Special about Face Recognition?
447 Nineteen Experiments on a Person with Visual Object Agnosia and Dyslexia but Normal
448 Face Recognition. J Cogn Neurosci 9:555–604.
449 Mur M, Kriegeskorte N (2014) What's there, distinctly, when and where? Nat Neurosci
450 17:332–333.
451 Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian Reconstruction of
452 Natural Images from Human Brain Activity. Neuron 63:902–915.
453 Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A Toolbox for
454 Representational Similarity Analysis. PLoS Comput Biol 10:e10003553.
455 O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects
456 and faces in ventral temporal cortex. J Cogn Neurosci 17:580–590.

457 Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers
458 into movies. Spat Vis 10:437–442.
459 Rice GE, Watson DM, Hartley T, Andrews TJ (2014) Low-Level Image Properties of Visual
460 Objects Predict Patterns of Neural Response across Category-Selective Regions of the
461 Ventral Visual Pathway. J Neurosci 34:8837–8844.
462 Silson EH, Chan AW-Y, Reynolds RC, Kravitz DJ, Baker CI (2015) A Retinotopic Basis for the
463 Division of High-Level Scene Processing between Lateral and Ventral Human
464 Occipitotemporal Cortex. J Neurosci 35:11921–11935.
465 Watson DM, Young AW, Andrews TJ (2016) Spatial properties of objects predict patterns of
466 neural response in the ventral visual pathway. Neuroimage 126:173–183.
467
468

Figure Legends:

Figure 1 Exemplars of intact, locally scrambled and globally scrambled images from the different object categories.

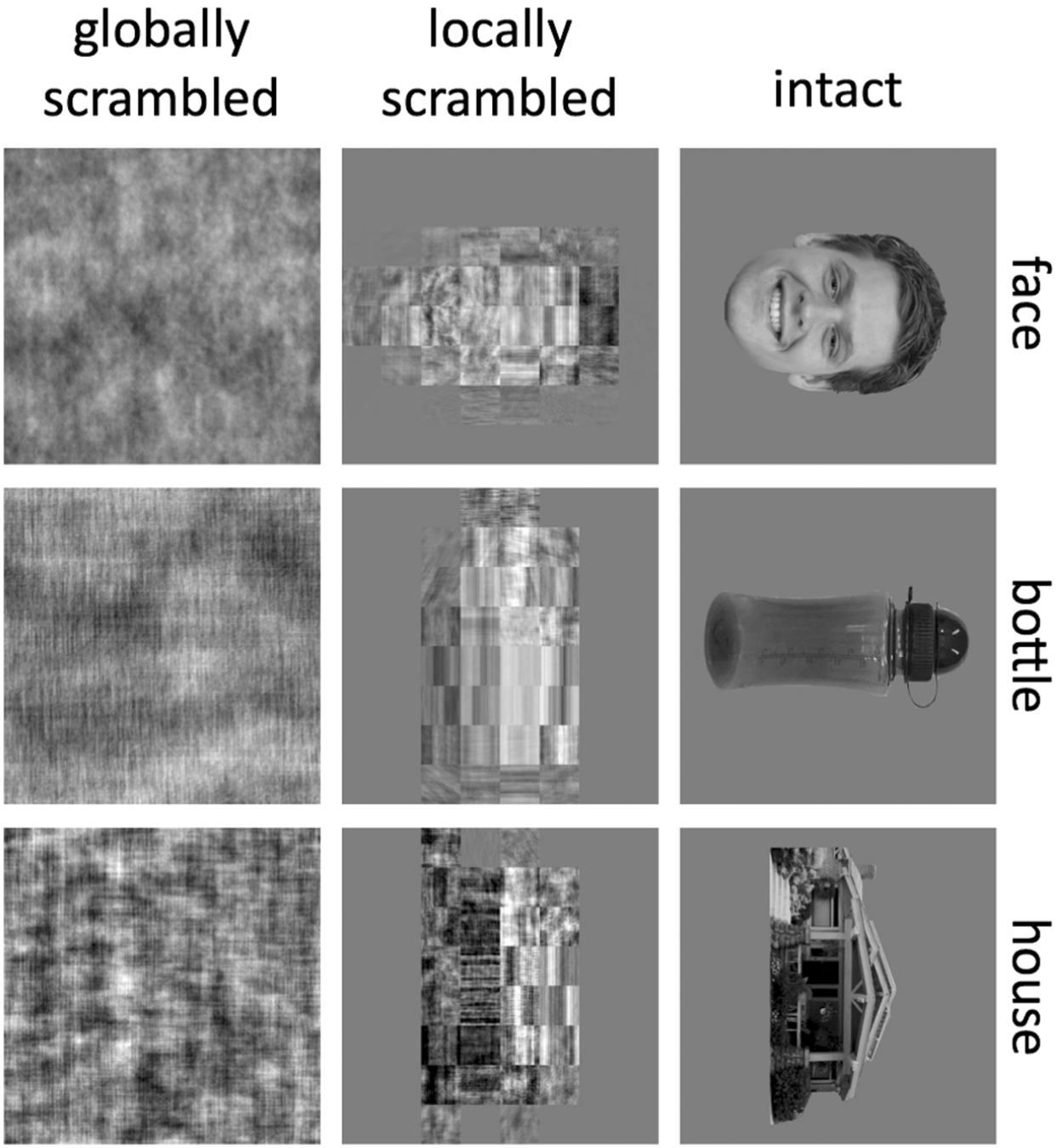
Figure 2 Category-specific patterns of EEG response to intact and scrambled images. (A) For each time-point, normalized patterns of response to odd and even trials of each category were compared across 64 electrodes. The correlation coefficients were then represented in a similarity matrix for that time-point. Distinct category-specific patterns of response were defined by higher within-category (e.g. face-face) compared to between-category (e.g. face-bottle) correlations. Correlation time-courses are shown for the (B) intact, (C) locally-scrambled and (D) globally-scrambled image types. Shaded region represents 95% confidence intervals obtained by bootstrapping across participants. Group mean correlation matrices at 100ms intervals are shown above the plot. Grey box at the base of the plot represents the time-points at which the stimulus was present. Blue bar at the base of the plot represents time-points at which the lower bound of the confidence interval is above zero, indicating significantly higher within- than between-category correlations.

Figure 3 Classifier accuracy for between-category discrimination (blue line) with (A) intact, (B) locally-scrambled and (C) globally-scrambled images (chance = 33%, grey line). Blue shaded regions represent 95% confidence intervals obtained through bootstrapping across participants. The blue bar at the top of the plot represents time-points at which the lower bound of the confidence interval is above chance. Grey box on the axes of the plot represents the stimulus duration.

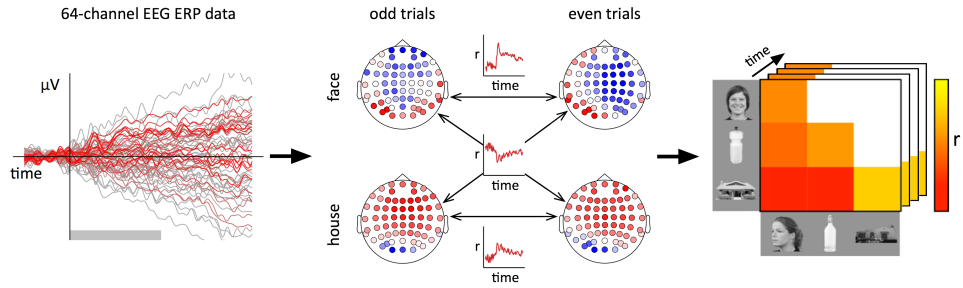
Figure 4 Similarity between patterns of EEG response to intact images and locally-scrambled (A) or globally-scrambled (B) images from the same object category. Blue shaded regions represent 95% confidence intervals across participants. Blue bar at the top of the plot indicates time-points at which the correlation is significantly above zero. Orange bar indicates time-points at which the correlation is not significantly different from the noise ceiling. Grey box represents the stimulus duration.

Figure 5 Classifier performance across different image types. (A) Accuracy in classifying responses to either intact or locally-scrambled images when trained on locally-scrambled or intact images, respectively. (B) Accuracy in classifying responses to either intact or globally-scrambled images when trained on globally-scrambled or intact images, respectively. Blue line indicates classifier accuracy across time, with shaded regions representing 95% confidence intervals obtained through bootstrapping across participants. Blue bar at the top of the plot represents time-points at which the lower bound of the confidence interval is above chance. Grey box shows stimulus duration.

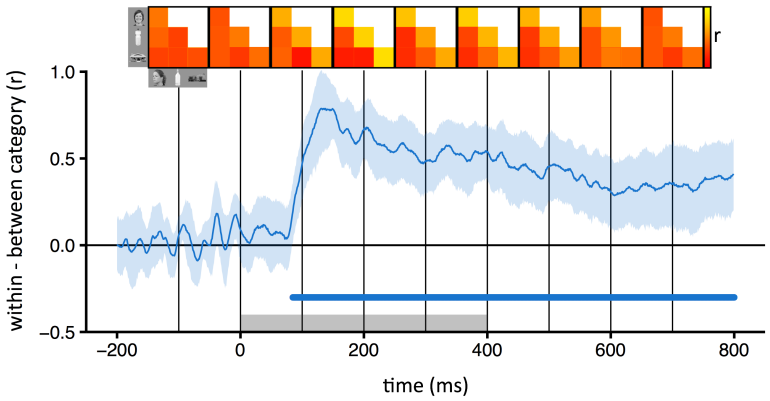
Figure 6 Temporal cross-correlation matrices for each image type. Responses to trials of the same condition were correlated over each combination of time-points. Correlations were collapsed across categories to give one matrix per image type (A - intact, B - locally scrambled, C - globally scrambled). Colourbar represents Pearson's correlation coefficient. Matrices were thresholded by obtaining 95% confidence intervals at each coordinate by bootstrapping across participants. Coordinates at which these intervals overlapped with zero are shown in white. Grey box represents the stimulus duration.



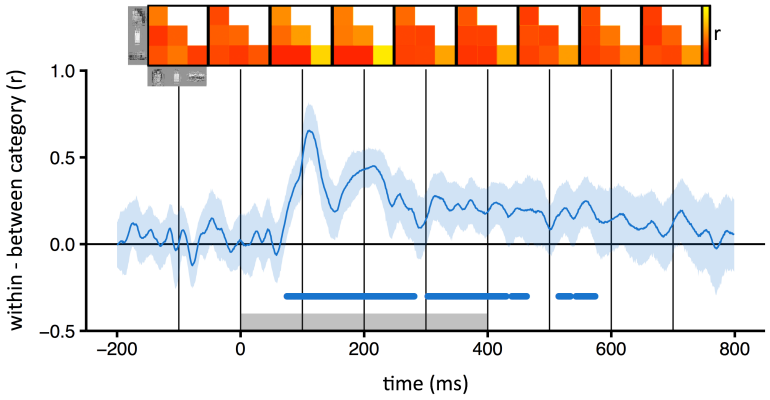
A



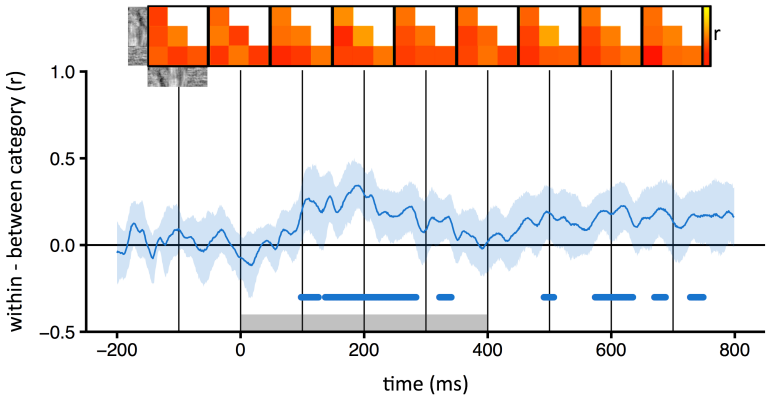
B

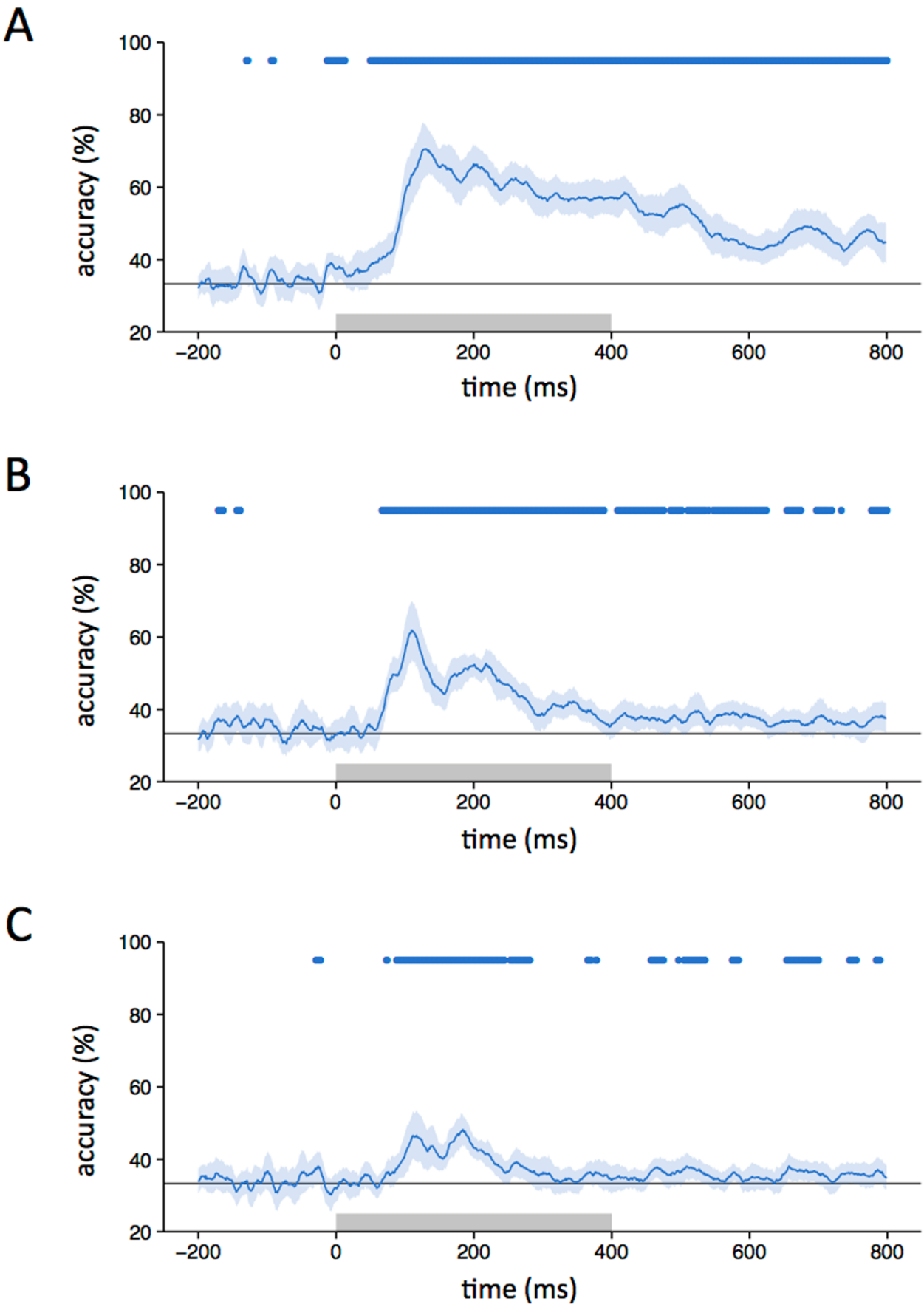


C

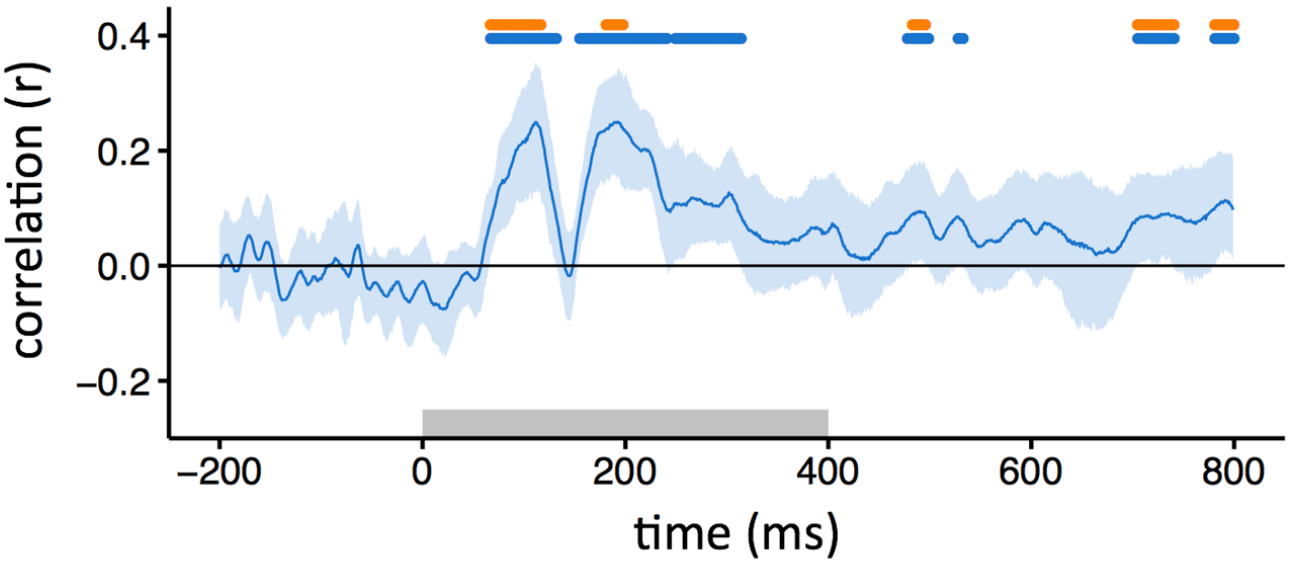


D





A



B