Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions

Erika W. Contini[a,b,*], Susan G. Wardle[a,b,1], Thomas A. Carlson[a,b,c,1]

[a] Department of Cognitive Science, Macquarie University, Sydney, Australia
[b] ARC Centre of Excellence in Cognition and its Disorders and Perception in Action Research Centre, Macquarie University, Australia
[c] School of Psychology, University of Sydney, Australia

ARTICLE INFO

ABSTRACT

Visual object recognition is a complex, dynamic process. Multivariate pattern analysis methods, such as decoding, have begun to reveal how the brain processes complex visual information. Recently, temporal decoding methods for EEG and MEG have offered the potential to evaluate the temporal dynamics of object recognition. Here we review the contribution of M/EEG time-series decoding methods to understanding visual object recognition in the human brain. Consistent with the current understanding of the visual processing hierarchy, low-level visual features dominate decodable object representations early in the time-course, with more abstract representations related to object category emerging later. A key finding is that the time-course of object processing is highly dynamic and rapidly evolving, with limited temporal generalisation of decodable information. Several studies have examined the emergence of object category structure, and we consider to what degree category decoding can be explained by sensitivity to low-level visual features. Finally, we evaluate recent work attempting to link human behaviour to the neural time-course of object processing.

1. Introduction

Visual object recognition is a complex problem. In everyday life we experience an overwhelming number of objects that the brain needs to rapidly differentiate and identify. How is it that we are able to identify a chair, for example, despite large variability in lighting, colour, design, materials and viewpoint? The human brain does a remarkable job of efficiently solving this problem, and has inspired decades of behavioural, neuroscience, and computer science research. The ventral visual processing stream, which involves a number of regions throughout the occipito-temporal cortex, is well-established as the neural pathway for object recognition (e.g., Grill-Spector et al., 2001; Haxby et al., 2001; Ishai et al., 1999). Functional neuroimaging methods have identified a number of object-selective areas in human visual cortex that are preferentially activated by specific object categories including animals (Chao et al., 1999; Martin et al., 1996), faces (Ishai et al., 1999; Kanwisher et al., 1997), bodies (Downing et al., 2001), places (Epstein and Kanwisher, 1998), and tools (Beauchamp et al., 2002; Chao et al., 1999; Martin et al., 1996). However, the number of object-selective regions identified to date accounts for only a handful of the numerous object categories that exist (Biederman, 1987). This highlights that while identification of these regions has advanced our understanding of

the neural pathways involved in higher-level visual processing, we are yet to fully understand *how* the brain solves the many challenges associated with object recognition.

Recent studies in visual object recognition have been influenced by advances in neuroimaging analysis methods, which allow for a fundamental change in the type of information that can be extracted from neuroimaging data. Early neuroimaging studies employed univariate analysis techniques that focus on identifying differences in the average activation of individual voxels or sensors for different experimental conditions. In contrast, more recent multivariate analysis techniques (e.g. multivariate pattern analysis; MVPA) analyse patterns of activation associated with experimental conditions from multiple voxels/sensors simultaneously. In the field of neuroimaging, multivariate methods have the potential to detect differences in activation which are lost when averaging data for univariate analyses, making them more sensitive (Carlson et al., 2003; Cox and Savoy, 2003; Grootswagers et al., 2016; Haxby et al., 2001; Haynes and Rees, 2006). While univariate and multivariate analyses are complementary in the information they provide, there is a fundamental difference in the types of experimental questions these methods address. Univariate magneto/electro-encephalography (M/EEG) analyses evaluate differences in *activation,* quantifying relative differences in average activity

---

* Corresponding author at: Department of Cognitive Science, Macquarie University, Sydney, Australia.
  E-mail address: erika.contini@students.mq.edu.au (E.W. Contini).
[1] shared senior authorship.

between experimental conditions, while multivariate methods have the potential to examine differences in *information,* for example by comparing differences in distributed patterns of brain activation between experimental conditions (Grootswagers et al., 2016; but see also de-Wit et al., 2016).

Decoding methods are a form of MVPA that have provided important insights into how the brain processes information (Grootswagers et al., 2016; Haxby et al., 2014; Haynes, 2015; Pereira et al., 2009). In the field of object recognition, decoding methods aim to map differences in complex neural activity patterns associated with perceiving objects. Early studies adopting these methods marked an important first step in our understanding of *how* objects are processed in the brain (Carlson et al., 2011; Carlson et al., 2003; Clarke and Tyler, 2014; Cox and Savoy, 2003; Haxby et al., 2001; Kriegeskorte et al., 2008; Liu et al., 2009; O'Toole et al., 2005). A particular strength of applying decoding methods to M/EEG data with the aim of understanding visual object recognition is that it can reveal how visual object representations change over time with high temporal resolution. The focus of this review is on what has been learned about visual object processing in the human brain with the application of these recent, powerful temporal decoding methods.

To date decoding methods have been used in conjunction with a variety of neuroimaging and neurophysiology approaches to investigate the neural mechanisms underlying object recognition. For example, brain decoding using single cell recordings has revealed category structure within monkey inferior temporal cortex (IT) (Hung et al., 2005), with differentiable neural patterns associated with animate and inanimate objects, as well as more specific animate subcategories, including human and animal faces and bodies (Kiani et al., 2007). Decoding of fMRI data has shown commonalities between object response patterns in both human and monkey brains, with similar differentiation of animate/inanimate categories and face/body subcategories identified in both species (Kriegeskorte et al., 2008). Categorical representations are not limited to localised brain regions, as activity evoked by objects and faces often overlaps between the categories and is distributed throughout ventral temporal cortex (Haxby et al., 2001). Moreover, there has been a recent move towards looking beyond representing object categories in terms of dichotomies, such as the animate/inanimate distinction. Instead, a continuum has been proposed that spans from inanimate objects to humans, with objects categorised in terms of their biological similarity to humans (Connolly et al., 2012; Sha et al., 2015). Together, these studies emphasise a role for distributed patterns of activity in the neural representation of object category, building on the original observation of overlapping activation patterns in the human ventral stream (Haxby et al., 2001).

Decoding of time-series data using M/EEG has offered the potential to examine the time-course of object representations in the human brain, revealing a dynamic evolution of object category structure over time (e.g., Barragan-Jason et al., 2015; Carlson et al., 2011; Carlson et al., 2013; Cauchoix et al., 2014; Cichy et al., 2014; Clarke et al., 2014; Goddard et al., 2016; Kaiser et al., 2016a, 2016b; Simanova et al., 2010). As visual information moves through the ventral pathway, the content of visual representations changes rapidly between brain regions. fMRI decoding studies have revealed much about the representation of objects in the human brain, however the coarse temporal resolution of fMRI limits the examination of dynamic visual processes. As a complement to the static snapshot of representational structure revealed with fMRI, neuroimaging techniques with higher temporal resolution such as M/EEG facilitate investigation of the dynamic processes of visual object recognition. The aim of this review is to discuss how time-series decoding studies have advanced our understanding of the complexities of visual object recognition by focusing on the dynamic processes involved. Time-series decoding is a relatively new approach to studying object recognition, and here we highlight the potential of this new direction to inform the field. Note that we focus on

what has been learned about object processing by applying time-series decoding methods, for a more thorough discussion of the technical details of time-series decoding analyses see Grootswagers et al. (2016).

2. Decoding the object recognition time-course

2.1. Initial insights into time-series decoding

It is well-established that the process of visual object recognition requires a number of hierarchically organised stages that progress through the occipito-temporal pathway (Grill-Spector and Malach, 2004; Malach et al., 2002). Early retinotopic visual areas are more sensitive to changes in low-level stimulus properties, while higher cortical areas within the ventral temporal lobe produce more complex responses to whole objects, and appear to account for more abstract properties such as object category (Altmann et al., 2003; Grill-Spector and Malach, 2004; Van Essen et al., 1992). Single-unit recordings in macaques have shown that these different processing stages progress successively in time as information passes through the occipito-temporal pathway (Schmolesky et al., 1998). With the development of M/EEG techniques, we are able to evaluate the temporal dynamics of visual object recognition in the human brain with millisecond resolution, allowing us to delve into more specific and fine-grained processes occurring in the various stages of visual object processing.

Differences in early versus late stages in object processing have been examined by comparing MEG and fMRI data for the same stimulus set. By linking both temporal and spatial neuroimaging data, Cichy et al. (2014) showed that activity early in the MEG time-course correlated more strongly with fMRI activity in V1, while later MEG activity was more strongly associated with activity in IT. The stimulus set were 96 colour images of animate and inanimate objects used in previous studies (Kiani et al., 2007; Kriegeskorte et al., 2008). The activity patterns associated with viewing each individual object were first compared using representational similarity analysis (RSA; Kriegeskorte, 2008), where a matrix is created based on the difference in brain activation patterns for every pairwise comparison of object images. This was done separately for the fMRI and MEG data. These 'dissimilarity matrices' provide an index of the difference in the brain response between object representations. The dissimilarity matrices were then compared across imaging modalities by examining when the relative similarity between the activation patterns for each object pair in the fMRI data most closely resembled that in the MEG data. The finding that the representational structure early in the MEG data more closely resembles V1 activity while later MEG data is closer to the structure observed in IT with fMRI is consistent with the known features of the visual processing hierarchy, and thus provides a source of validation for MEG decoding methods.

Time-series decoding methods have also expanded our understanding of the temporal intricacies associated with processing low-level stimulus properties. For example, Goddard et al. (2016) investigated temporal differences in the contribution of low and high spatial frequencies to the representation of object identity. Stimuli were greyscale images of objects that were matched in their amplitude spectrum to control for low-level visual properties. Phase randomisation was applied selectively to spatial frequency bands such that object identity information was confined to a restricted spatial frequency band in each stimulus. Decoding analysis revealed that the activation pattern of the whole-brain MEG recordings contained information related to object identity, but there were critical differences in the processing of low versus high spatial frequencies. They found that low spatial frequencies provided object identity information earlier and in more occipitally located regions than high spatial frequency information, which supported decoding of object identity later in the time-course. Further, by applying Granger causality analysis, they found evidence for both an early feedforward and later feedback flow of information related to object identity (Goddard et al., 2016).

Most of the existing temporal decoding literature on object recognition has focused on decoding object categories and investigating the representational structure of object representations. These results are discussed in detail in the following sections.

2.2. Temporal decoding of high-dimensional distributed category representations

Understanding population coding is one of the overarching aims of neuroscience (Averbeck et al., 2006). A significant and influential recent development in understanding the neural mechanisms underlying object representation has been the theoretical shift away from identifying specific localised regions associated with particular object categories towards focusing on more distributed activity patterns that extend throughout higher-level visual cortex (Haxby et al., 2001). A current focus in object recognition is on interpreting the highly multidimensional activation patterns elicited by object stimuli. Specifically, one of the main themes centres on understanding the structure of object category representations. In this section we outline what temporal decoding methods have so far revealed about the neural architecture of object categories.

A useful and intuitive way to think about the highly multidimensional activation patterns evoked by viewing objects is to conceptualise them in terms of an abstract representational space (DiCarlo and Cox, 2007; Kriegeskorte and Kievit, 2013). The complex activation pattern across M/EEG sensors elicited by viewing a particular object exemplar (e.g., shoe, horse, face) can be considered as a single point in an abstract representational space. Thus in this abstract representation, the proximity between data points indicates the degree of similarity, such that a greater distance between object exemplars represents more disparate neural activation patterns, while exemplars with closer points have more similar activation patterns. Consequently, object exemplars further away from each other in representational space are easier to "decode" from each other using machine learning classification than objects which are closer together, as they share more similar activation patterns and are less separable in higher-dimensional space.

Multidimensional scaling (MDS) is a technique that can be used as a tool for visualising the representational space of objects in the brain. An example of an MDS plot for object representations is shown in Fig. 1. Carlson et al. (2013) studied the first 1000 ms of the visual

object recognition time course using MEG decoding. MEG recordings were acquired as participants viewed a series of single object images while performing an unrelated attention task (reporting whether a letter superimposed on top of the object image was a vowel or a consonant). Using MDS to visualise the differences in object representations as a function of time, we can appreciate the emergence of category structure throughout the time course (Fig. 1). Early in the time-course (~60 ms post-stimulus onset), the representations of individual exemplars are almost wholly overlapping in the representational space, reflecting poor decodability of the individual objects based on the MEG data. This is expected given the time it takes for an image on the retina to transition to a cortical representation that is accessible to MEG (see Nowak and Bullier, 1997 for a review). By ~120 ms, differences between individual exemplars have emerged such that the exemplars are spread out in the representational space. This is the time of peak decoding for this image set; the point at which individual exemplars are most easily distinguished from one another. As we progress through the time course, category structure begins to emerge. From 120 ms, some subcategories appear to start to cluster: note the grouping of faces and animals in the centre. By 240 ms a clear categorical distinction between animate and inanimate objects is apparent (diagonal black lines in Fig. 1 indicate the category boundary). Interestingly, this animacy category distinction is maintained at 360 ms, despite much less differentiation of individual object exemplars, which are clustered tightly together in the later stages of the time-course. This is an example of the advantage of time-series decoding, as the emergence of categorical clustering over time is not captured by other neuroimaging methods with lower temporal resolution such as fMRI.

As MDS has been used in several fMRI and MEG studies with the same object stimuli (Carlson et al., 2013; Cichy et al., 2014; Kriegeskorte et al., 2008) it is possible to visualise the reliability and consistency of object representations across neuroimaging modalities by comparing across studies - a rare opportunity in neuroimaging (Fig. 2). All three studies used the object set (or subset of the set) originally used by Kiani et al. (2007) in monkey IT. Kriegeskorte et al. (2008) and Cichy et al. (2014) both used fMRI to evaluate the representational geometry of objects in human IT (Fig. 2a and b). The animate/inanimate divide is clearly visible in IT, and additional subcategory groupings such as animals and humans also cluster
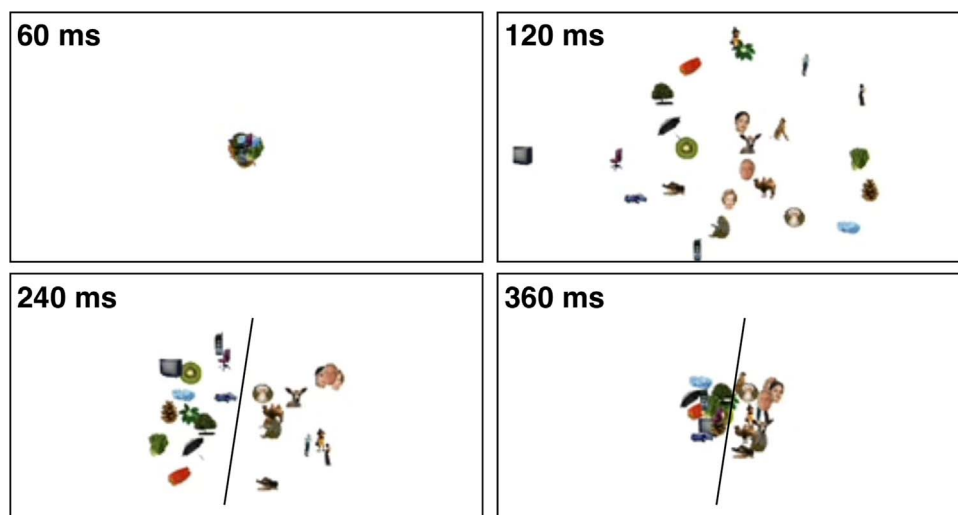

Fig. 1. MDS plots demonstrating the evolution of object representations over time with MEG. Individual MDS panels show the representational geometry of a set of objects at a particular time-point (shown as ms post-stimulus onset). Distances between objects represent the level of similarity between the neural patterns measured with MEG such that larger distances indicate greater dissimilarity (i.e., more distinct neural patterns). The representational geometry evolves from initially entirely overlapping (poorly discriminated) objects, to maximal differentiation at 120 ms (peak decoding for this sample). The animate/inanimate category distinction (black dividing line represents the boundary) peaks at 240 ms and is maintained even at 360 ms, although differences between individual object exemplars become less defined. Images adapted with permission of the Association for Research in Vision and Ophthalmology, from Carlson et al. (2013); permission conveyed through Copyright Clearance Center, Inc.
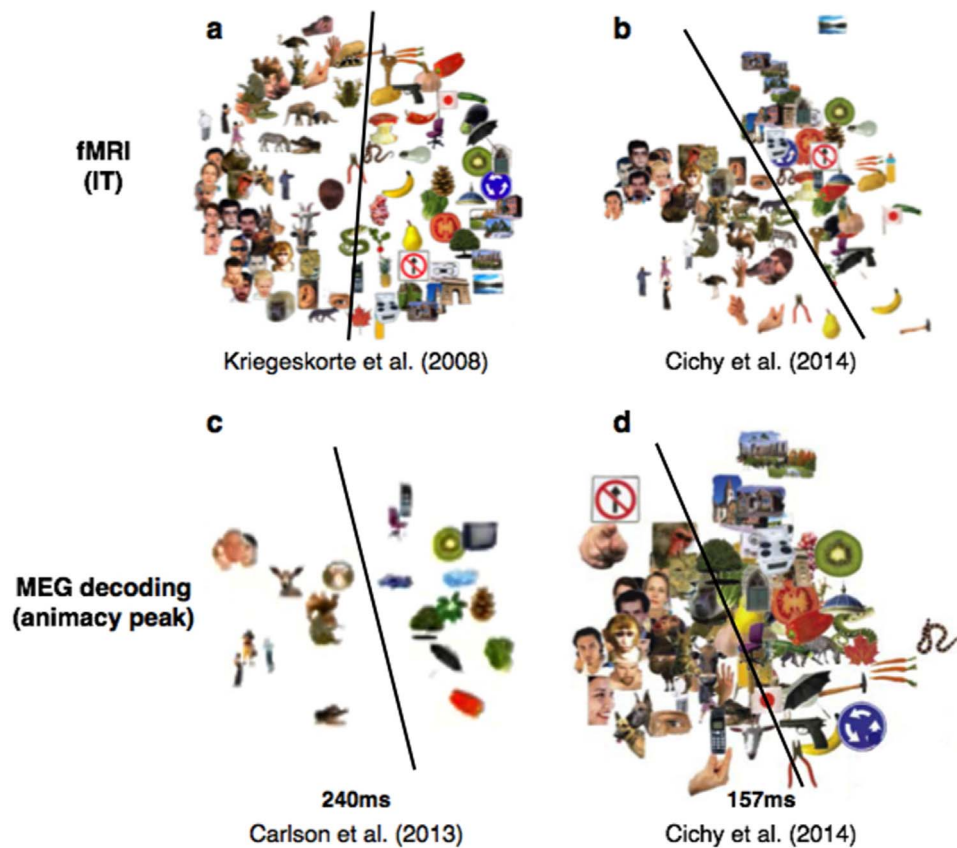
Fig. 2. Comparison of MDS plots from fMRI and MEG studies. Top MDS plots (a and b) show the representational geometry in human IT using fMRI; bottom MDS plots (c and d) show the representational geometry using MEG at the time of peak decoding for the animate/inanimate (animacy) category distinction. Animacy peak decoding times for the individual studies are listed below the MEG MDS plots. Note that (a) has been normalised and rigidly aligned (Procrustes alignment) for visualisation, producing a more regular spacing between object exemplars in the MDS plot compared to panels b-d. Black lines in each plot mark the (approximate) animacy boundary. Panel (a) image from Kriegeskorte et al. (2008), reprinted with permission from Elsevier. Panel (c) image from Carlson et al. (2013) reprinted with permission of the Association for Research in Vision and Ophthalmology; permission conveyed through Copyright Clearance Center, Inc. Panels (b) and (d) images from Cichy et al. (2014) reprinted with permission from Macmillan Publishers Ltd: Nature Neuroscience, copyright (2014).

together. These fMRI MDS plots can be compared to those from MEG time-series decoding studies (Fig. 2c and d; Carlson et al., 2013; Cichy et al., 2014). MEG data is shown at the time of peak animacy decoding in each study: 240 ms (Fig. 2c) and 157 ms (Fig. 2d). Note that the time of peak decoding is dependent on the stimulus set being evaluated as well as the noise and variability in a particular data set, complicating the comparison of specific decoding onset times across studies. The emergence of feature information or category structure is better discussed in terms of relative timing within individual studies. A critical difference between the MDS plots for fMRI versus MEG is that the MEG plots represent similarity in whole-brain activation patterns, whereas the fMRI analysis is localised to IT. Considering this substantial difference, the degree of similarity in the categorical representation is remarkable (compare top and bottom rows of Fig. 2).

Another advantage of using time-series decoding to investigate object processing is that we can examine to what degree information related to object category is maintained in the brain activation patterns over time (see King and Dehaene, 2014 for a review of temporal generalisation methods). This is achieved by performing time-point by time-point comparisons across the entire time-course, whereby a classifier is trained on data from one time-point and then tested at every other time point. If information is maintained within the signal for some period of time, then the classifier is expected to generalise and successfully classify the test data, regardless of the time difference between the training and test data sets. The results of temporal generalisation are visualised in two-dimensional heat maps indicating the magnitude of classification performance for each pair of time points (Fig. 3).

A number of MEG studies have examined the temporal generalisation of decoding to evaluate the dynamics of object representations (Fig. 3). For each plot in Fig. 3, the diagonal represents standard decoding analysis when the classifier is trained and tested on data from the same time-point, thus classification accuracy is expected to be highest along the diagonal. If decoding performance is significant off the diagonal, it is indicative of some degree of temporal generalisation of the decodable signal related to object classification. Note that the decoded category differs across studies: Isik et al. (2014) and Cichy et al. (2014) decoded object exemplars, while Carlson et al. (2013) and Grootswagers et al. (2016) decoded object animacy.

The most striking similarity between the temporal generalisation plots from these different studies is that the MEG signal underlying object decoding evolves relatively quickly, with little information generalising across time. Where there is generalisation, it tends to cluster closely around the diagonal, suggesting that the structure of object representations accessible in the whole-brain MEG signal evolves rapidly, and follows a specific neural trajectory. Another interesting feature is that the period where there is the greatest generalisation occurs relatively late after stimulus onset (see particularly Fig. 3c-d). This suggests maintenance of object representations at later stages of processing (Carlson et al., 2013). Alternatively, it may reflect the gradual accumulation of multiple related information processing stages of variable duration throughout the visual processing hierarchy, leading to a greater temporal spread of information over time. Paradoxically, there are also some periods of below chance decoding, whereby a classifier systematically categorises an object incorrectly as the opposite category (e.g. see blue regions in Fig. 3c).
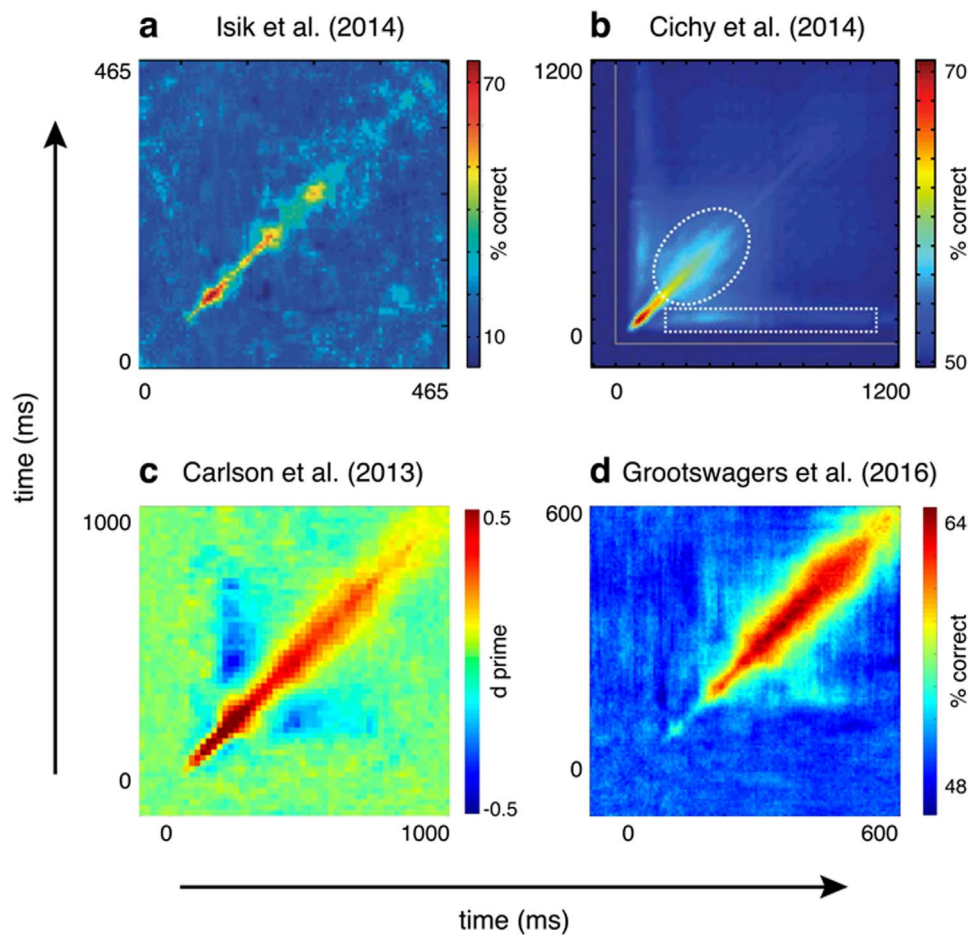
Fig. 3. Examples of temporal cross-decoding of MEG data. Temporal generalisation of MEG decoding from (a) Isik et al. (2014), (b) Cichy et al. (2014) (c) Carlson et al. (2013), and (d) Grootswagers et al. (2016). Generalisation of decoding performance across time is assessed by training and testing the classifier on MEG data from different time points. Points on the diagonal represent training and testing on data from the same time point (i.e., regular classification). Colour indicates classifier performance as a function of time for each study; hotter colours indicate higher levels of classifier performance (measured using decoding accuracy percent (a, b, d) and d-prime in (c). Panel (a) image from Isik et al. (2014) reprinted with permission from ©The American Physiological Society. Panel (b) image from Cichy et al. (2014) reprinted with permission from Macmillan Publishers Ltd: Nature Neuroscience, copyright (2014). Panel (c) image from Carlson et al. (2013) reprinted with permission of the Association for Research in Vision and Ophthalmology; permission conveyed through Copyright Clearance Center, Inc. Panel (d) image from Grootswagers et al. (2016) reprinted with permission from theMassachusetts Institute of Technology © 2016, published by the MIT Press.

Carlson et al. (2013) suggest this may represent adaptation or inhibition following a period of excitation, resulting in later inversion of the neural representation. Consistent with this explanation, antic-orrelated MEG signals have also been identified around the time of stimulus offset for both visual (Carlson et al., 2011) and auditory (Chait et al., 2007) stimuli.

An exception to the general lack of generalisation is seen in the results of Cichy et al. (2014), where there is a period of significant generalisation of early information across most of the time-course (Fig. 3b; grey dotted rectangle shows the period of significant temporal generalisation). This could be because this analysis used pair-wise object decoding as opposed to leave-one-exemplar-out category classi-fication (see Carlson et al., 2013; Grootswagers et al., 2016). With pairwise object decoding, low-level information diagnostic of object is contained in both the training and test sets for the classifier, thus maintenance of early visual information is consistent with low-level properties being useful for classification. There is a similar pattern of generalisation in Fig. 3c, and this analysis also included exemplars in both training and test sets (Carlson et al., 2013). However, in leave-one-out exemplar decoding (Fig. 3d) the test exemplar is not included in the training set so early processing reflecting sensitivity to low level image properties is less likely to be used by the classifier. Kaiser et al. (2016a) took this idea further by using temporal generalisation to directly examine decoding of low-level shape versus object category in a

matched stimulus set (see Section 2.5, Fig. 6).

The studies in this section demonstrate the utility of examining the change in representational structure of object representations over time. Studies to date have shown that the brain's response to visual objects is highly dynamic, evolving rapidly from sensitivity to low-level visual properties to more category-like representations. There is also evidence for a robust categorisation distinction between animate and inanimate objects that is sustained. In the following section we examine what has been learnt from time-series decoding about the hierarchical structure of object category representations.

2.3. Uncovering the hierarchy of object category representations

Visual object categorisation can be understood as a dynamic process of evidence accumulation over time (Mack and Palmeri, 2011; Nosofsky and Palmeri, 1997; Philiastides and Sajda, 2006). Functionally, the accumulation of evidence for object category mem-bership is likely to exploit the complex network of feedback and feedforward connections within the object-selective ventral pathway, rather than operating as a linear progression of representation from low-level features through to semantic concepts of increasing abstrac-tion (Kravitz et al., 2013). This is a recent development and extension of the earlier idea of sequential processing stages, where objects are first categorised at an intermediate/basic level (Mervis and Rosch,
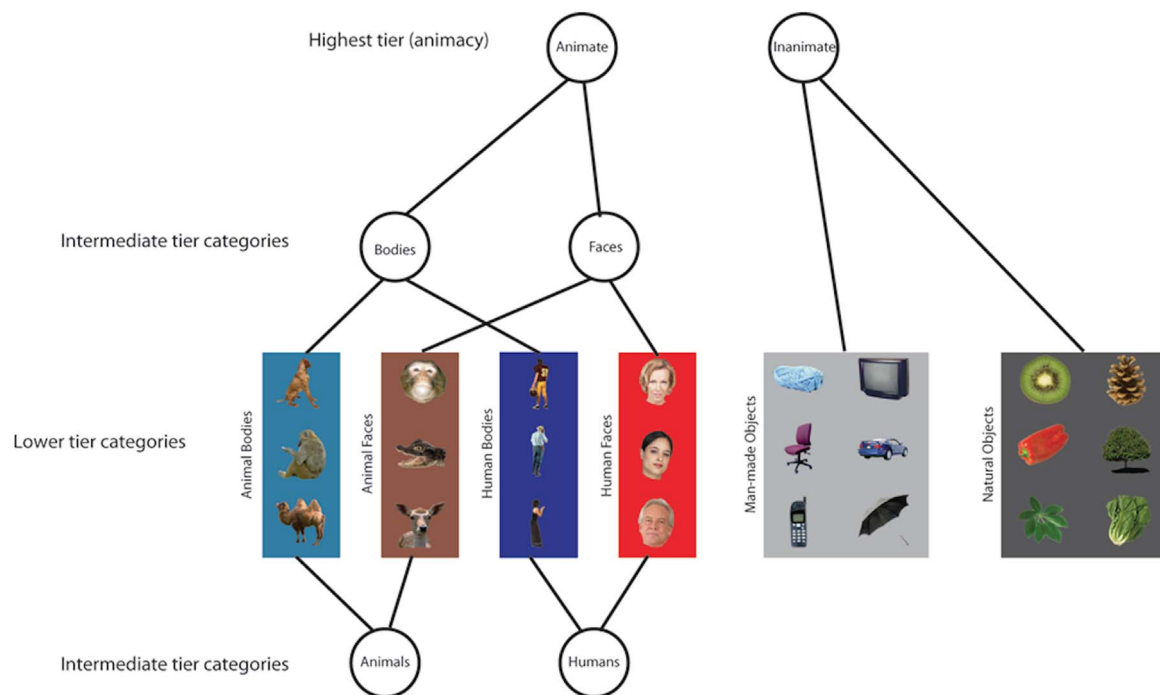
Fig. 4. Hierarchical category structure of the object stimuli used in both Carlson et al. (2013) and Cichy et al. (2014). Diagram shows the three levels of category abstraction: lower tier (animal and human faces and bodies), intermediate tier (bodies and faces), and highest tier (animacy). Note, number of stimuli differed between the two studies: Carlson et al. (2013) used 24 images (as shown), while Cichy et al. (2014) employed 92. Image reprinted with permission of the Association for Research in Vision and Ophthalmology, from Carlson et al. (2013); permission conveyed through Copyright Clearance Center, Inc.

1981) (e.g., cat), with superordinate (e.g., animal) and subordinate (e.g., Siamese cat) categorisation occurring later in the visual processing hierarchy (Mack and Palmeri, 2011).

MEG decoding methods have been applied to understanding the timing of category abstraction by selecting stimuli with a planned hierarchical category structure and assessing when these category level distinctions emerge. Results from two MEG decoding studies (Carlson et al., 2013; Cichy et al., 2014) provide insights into the category processing hierarchy by evaluating the time course of hierarchically organised object categories. The stimulus sets across both studies followed the same planned hierarchical structure (see Fig. 4): the highest (most general/abstract) category tier was the commonly evaluated animate/inanimate distinction. The animate domain was then further subdivided into intermediate categories of faces/bodies as well as humans/animals, with the lowest (most specific) category level being that of human faces/human bodies, animal faces/animal bodies. For the inanimate domain, these objects could be further subcategorised into man-made and natural objects.

To investigate the dynamics of hierarchical object category representations, these studies looked at two critical moments in the MEG object decoding time-course: the time at which decoding first reaches significance (decoding onset) and the time that the category distinction is maximally differentiated (peak decoding time). A visual comparison of the decoding time course in both Carlson et al. (2013) and Cichy et al. (2014) for object categories at different levels of the category hierarchy is shown in Fig. 5. Carlson et al. (2013) found the onset of significant decoding for individual object exemplars occurred around 80 ms after stimulus onset, which was also similar to the decoding onsets of individual categories; all falling between 80–100 ms. Cichy et al. (2014) similarly found little variability between overall exemplar decoding onset (~48 ms) and individual category onsets (~51–61 ms post-stimulus onset).

Evaluation of peak decoding times for the category hierarchy in Carlson et al. (2013) revealed a largely linear organisation of the category tiers (see Fig. 5, red bars), such that lower tier categories (e.g., human faces, animal bodies) showed peak decoding times from 120–180 ms, followed by intermediate tier categories (faces, bodies, human, animal) which ranged from 160–240 ms, with the superordinate tier animacy category (animate vs. inanimate) peaking at 240 ms. The results suggest evolving hierarchical representations of object categories that progress from specific, individual exemplar individuation through to more abstract semantic groupings. Results from Cichy et al. (2014) largely support these findings (see Fig. 5, blue bars), with peak decoding between intermediate tier categories (natural vs. artificial and faces vs. bodies, 122 and 136 ms respectively) occurring earlier than the superordinate animacy category peak (157 ms).

Although there is similar evidence in the results of both Carlson et al. (2013) and Cichy et al. (2014) for a roughly linear emergence of object categories from specific to more abstract object category representations, the data are not entirely straightforward. For example, for the subordinate tier categories in Cichy et al. (2014), this linear hierarchy does not hold, with subcategory peaks occurring later in the time-course than some of the intermediate tier category peaks: human versus animal faces peaked at 127 ms, while human versus animal bodies peaked at 170 ms. Furthermore, in Carlson et al. (2013), the intermediate tier categories of faces and bodies are best discriminated at a similar time point to the more specific lower tier categories (e.g., human faces, animal bodies), and similarly, Cichy et al. (2014) found no significant difference between the peak latency at which individual images were discriminated early in the time course compared to a higher tier category grouping of human versus non-human bodies.

It is plausible that these anomalies in the emergence of hierarchical category structure reflect that category representations are not a strict linear hierarchy, consistent with recent ideas on the importance of recurrent processing such as lateral connections and feedback on feedforward visual object processing (Kravitz et al., 2013). Carlson et al. (2013) suggest these discrepancies may be partially explained by the inclusion of face and body stimuli in the design, as there is evidence for special processing systems specifically tuned for the detection of faces and bodies in both human and primate ventral temporal cortex (Downing et al., 2001; Kanwisher et al., 1997). If some categories are
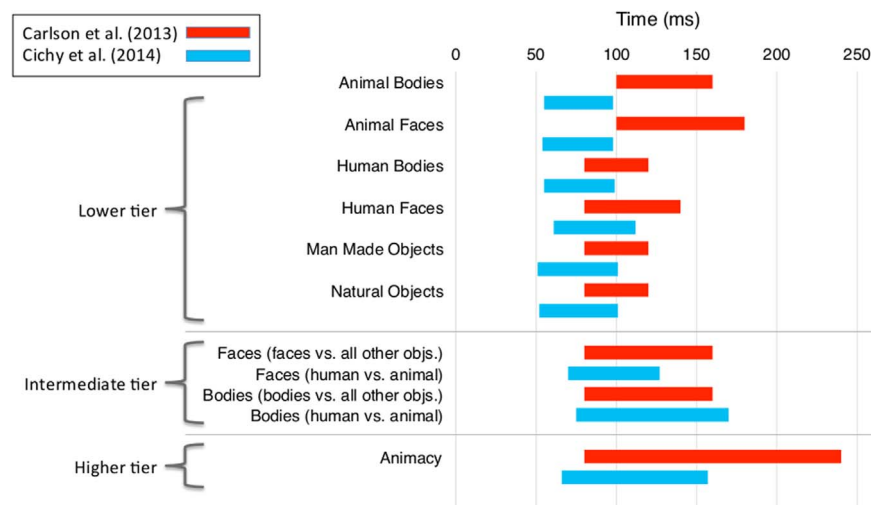
Fig. 5. Onset and peak decoding times for category tiers within the object category hierarchy, as evaluated by Carlson et al. (2013) and Cichy et al. (2014). Each coloured bar represents the time from the onset of significant category decoding to the time of peak decoding (relative to post-stimulus onset) for Carlson et al. (2013) (red) and Cichy et al. (2014) (blue). Note, the intermediate tier categories are evaluated using different stimulus comparisons by Carlson et al. (2013) and Cichy et al. (2014) (specific comparisons are noted in brackets).

processed differently than others, a straightforward linear emergence of category structure is not expected. A further caveat is that these results are specific to the selected stimulus set and the imposed category structure of the experimental design.

2.4. Decoding the time-course of face processing

The object category of faces stands out as having a particularly strong contribution to decoded brain representations measured with EEG (Kaneshiro et al., 2015) and MEG (Van de Nieuwenhuijzen et al., 2013). For example, in the MDS results for large object sets, human faces tend to cluster prominently (Fig. 2). Two EEG decoding studies have examined the temporal dynamics of face representations in detail (Barragan-Jason et al., 2015; Cauchoix et al., 2014). A strong feature of these studies is that they used large natural image stimulus sets containing pictures of faces with natural backgrounds, and faces were not repeated in the experimental design. This is in contrast to the majority of temporal object decoding studies to date, which tend to repeat the same object exemplars multiple times within the experiment to increase signal-to-noise. Both studies had participants complete a go/no-go task and trained a classifier to detect the target versus non-target faces across the EEG time-course (human vs. animal in Cauchoix et al., 2014, and famous vs. unfamiliar faces in Barragan-Jason et al., 2015). In both studies, the time-course of face decoding follows a similar pattern: there is an initial sharp rise in performance of the classifier early in the time-course, shortly followed by a relatively brief plateau or drop in decoding, before a second, more gradual rise in decoding accuracy, with peak decoding occurring relatively late in the time-course (350 ms and 600 ms for humans/animals and familiarity respectively). They suggest that the two (early vs. late) peaks in decoding reflect different levels of processing within the object recognition hierarchy. An initial fast feedforward signal driven by low-level stimulus properties is later modified to incorporate more detailed feedback from higher cortical areas (Barragan-Jason et al., 2015; Cauchoix et al., 2014).

A recent study which combined MEG decoding with fMRI localization provides further insight into different levels of face processing. Vida et al. (2017) evaluated time-series decoding performance of face identity in specific regions of the brain linked to face processing. They used an fMRI localiser to identify two regions that were preferentially activated by faces, over and above activity associated with objects more broadly: right lateral occipital cortex and the right fusiform gyrus. At these sites, and a control site (left V1) they related the time-course of face decoding to three models of face processing: one that represented

low-level, V1-like responses ("image-based" model), a higher-level "identity-based" face model, and a behavioural-rating model where participants rated a subset of the images as to how similar the face identities were on a scale from 1 to 8. Their results showed that firstly, while face identity was decodable at all three regions from ~50–400 ms, the image-based model was a better predictor of the neural data than the identity-based model until ~200 ms, after which its performance dropped below that of the identity-based model at a number of time-points. However, this shift occurred in the face-selective regions only, with the image-based model outperforming others in the V1 control region throughout the time-course.

These results are in line with previous findings (see Section 2.1) that suggest low-level visual features are associated with activity in early visual areas, while high-level conceptual information is represented in cortical regions further down the ventral processing stream. The behavioural-rating model similarly correlated with the neural data significantly in all three regions from ~50–400 ms, however, after controlling for representations in the V1 control region, this significant time-window reduced to 100–250 ms in the face-selective regions. Notably, behavioural ratings were significantly more similar to the identity-based model than the image-based model. Significantly, by using time-series decoding methods, Vida et al. were able to show that face-selective regions process information related to *both* low-level image properties as well as higher-level face identity-based representations, highlighting the potential shortcomings of associating localised regions with single functions. The challenges involved in empirically separating out the contribution of low level properties from more abstract object representations in object decoding studies is discussed in detail in the following section.

2.5. Is a category a category?

Above we reviewed several studies focusing on uncovering the dynamic category structure of object representations. However, when evaluating object representations, it is important to separate effects related to the abstract concept of an object "category" from other potentially confounding features that co-vary with category membership such as low-level visual properties including shape, colour, and luminance contrast (Wardle and Ritchie, 2014). For example, items within the category of tools tend to have handles, resulting in a characteristic long and often slender feature to their appearance (e.g., hammer, scissors, screwdriver, toothbrush). This relatively consistent information within this category could result in decoding based on low-level feature similarities related to shape, such that instead of decoding

the category of "tools" compared to "faces", we may instead be erroneously decoding "long and thin" versus "round". Low-level stimulus properties such as orientation (Cichy et al., 2015; Ramkumar et al., 2013) are readily decodable from whole-brain MEG signals, and perceptual similarity is a strong determinant of the decodability of abstract visual stimuli (Wardle et al., 2016).

Time-varying decoding studies investigating object recognition have attempted to account for potential low-level confounds using a variety of different approaches. In fMRI studies, V1 is often used as a control region to show that unsupervised categorical clustering of object representations (e.g., by animacy) do not emerge in V1, but are present in IT (Cichy et al., 2014; Kriegeskorte et al., 2008). As this degree of spatial localisation is not possible with MEG, time-varying decoding studies often exploit the time-course to demonstrate that models of early visual processing may perform well early on in the time-course but later the contribution from categorical models such as animacy emerges. For example, Carlson et al. (2013) used RSA to evaluate the performance of low-level feature based models of vision on object representations measured with MEG. The models of early visual processing included a shape-based model that compared the image silhouettes (Jaccard, 1901), a colour-based model (CIE), and a hierarchical visual processing model (HMAX) containing layers tuned to process varying levels of stimulus complexity (Riesenhuber and Poggio, 1999; Serre et al., 2007). Notably, these models were able to successfully differentiate human faces from other objects (silhouette model), humans from human bodies (CIE), and man-made objects from other objects (HMAX). While the models were only able to distinguish a maximum of two out of a possible 10 object categories in that particular data set, these results highlight the importance of accounting for the potential influence of low-level stimulus properties on object category decoding.

Although there is evidence that visual similarity accounts for at least some of the representational similarity shared by objects within the same category, it is unlikely to account for all observed category effects. As a control for low-level stimulus properties when decoding object category, Carlson et al. (2011) evaluated decoding accuracy of cars and faces from MEG activation patterns compared to artificially generated car and face textures which preserved local image statistics but removed recognisable form. If classification of object category is heavily based on low-level image statistics, it is expected that the classifier would find it difficult to discriminate between objects and their matched texture images, which share low-level image properties (e.g., classifying 'car' vs. 'car-texture'). However, object categories were able to be accurately decoded from their texture counterparts in the whole-brain MEG activation patterns, and this distinction emerged earlier in the time course than information differentiating the two object categories of cars and faces (Carlson et al., 2011). These results are consistent with a transition from decoding based on V1-like image properties to IT-like object category over the MEG decoding time course (Cichy et al., 2014).

The contribution of semantic meaning to category representations was explored in a study by Clarke et al. (2014), which compared the performance of both low-level visual processing and feature-based semantic models for their 302 object stimuli from 11 categories. They found that a model which combined both category-specific semantic information and low-level visual features best accounted for variability in neural object representations later in the MEG time course, over and above what could be achieved by using the HMAX model alone (Clarke et al., 2014). These results suggest there is a contribution of abstract category membership (as indexed by semantic similarity) to the brain representation. Similarly, an EEG study evaluating event-related potentials (ERPs) associated with object representations aimed to provide evidence for the involvement of conceptual category membership by examining decoding of the categories 'animals' versus 'tools' across three modalities of visual pictures, spoken words and written words (Simanova et al., 2010). However, classifier performance was

much higher for visual pictures than for spoken or written words, complicating the critical cross-classification analysis across modality.

Similar to the frequently-used approach of assessing the potential contribution of low-level properties to object representations with the HMAX model (Carlson et al., 2013; Clarke et al., 2014; Isik et al., 2014; Kriegeskorte et al., 2008), recently Cichy et al. (2016) compared MEG object decoding performance with the performance of deep neural networks (DDNs). DNNs are a powerful form of computer vision model, comprised of multiple non-linear processing layers with the ability to learn tasks such as object categorization (see Kriegeskorte, 2015). The recent advancement of these models has propelled computer modelling of object recognition to a level that is beginning to rival human object categorisation performance (He et al., 2015). By comparing the decoding performance of the DNN model to MEG and fMRI brain data, Cichy et al. (2016) showed a hierarchy of both spatially and temporally ordered processes that correlated with the DDN processing layers in an ordered fashion; deeper layers were associated with later brain activity in higher cortical processing regions. Furthermore, by analysing activity throughout the entire brain, they identified involvement of not only the ventral, but also the dorsal pathway in object recognition. A key issue for the future is to understand to what degree deep-neural networks mimic the object recognition processes in the human brain, and in what respects the two systems are different.

The above approaches are data-driven, and aim to show that categorical structure that emerges in the neuroimaging data cannot be completely explained by sensitivity to lower level visual properties confounded with object category. A more compelling line of evidence against reducing category representations entirely to low-level properties involves experimental manipulations designed to separate the two factors. A recent MEG/fMRI study (Kaiser et al., 2016a) sought to specifically identify object category representations independent of shared visual properties by purposefully selecting visually similar stimuli that belonged to the semantic categories of body parts and clothing (e.g., gloves vs. hands, shirts vs. torsos). With this stimulus set, visual similarity is balanced across category membership. While their fMRI results showed overlapping spatial representations for both shape and category information, the MEG data showed a specific, comparatively late time window within which category-selective information was present (Fig. 6b) and in contrast to this, shape dependent responses (Fig. 6a) were decoded relatively early on, from 90 ms post-stimulus onset, and was sustained throughout the time-course.

Object invariance is both an interesting theoretical question for investigation and an experimental manipulation that minimises the contribution of low-level features. Object invariance is the ability to successfully recognise objects despite high variability in their appearance, for instance, due to differences in viewpoint and size. Carlson et al. (2011) showed that object category information for faces and cars could be decoded from the neural data despite changes in retinal location of the image; evidence for a position-invariant object representation. They tested a classifier on information from a novel retinal location that was not included in the original training set, and were able to successfully recover information about the object category as early as 135 ms, suggesting that position-invariant category information is present early in the neural signal. Furthermore, this was not simply due to low-level image statistics, as category information could not be decoded from abstract textures generated to preserve the local low-level image statistics of the object images (Carlson et al., 2011).

Invariant object decoding was further investigated by Isik et al. (2014). Isik et al. showed position invariant decoding across three stimulus locations for a set of six object exemplars, while also evaluating size invariance. Using three different stimulus sizes for each exemplar, they showed that size-invariant information was present in the signal from around 125 ms post-stimulus onset, earlier in the time-course than position-invariant information which emerged around 150 ms for their stimulus set. Furthermore, by comparing the decoding
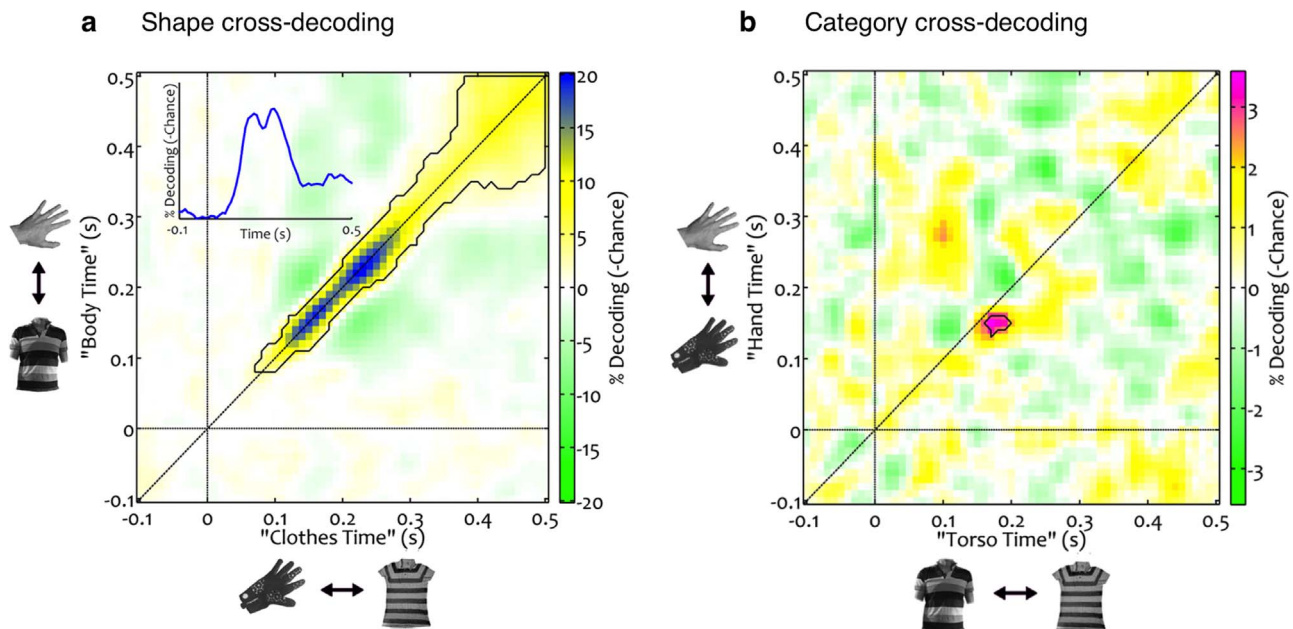
Fig. 6. Temporal cross decoding of object category (body parts vs. clothing). The black outlined areas in both plots indicate when cross-decoding of (a) shape or (b) object category was significantly above chance. Images reprinted from Kaiser et al. (2016a) with permission from ©The American Physiological Society.

latencies for the different train-test conditions, they found that comparisons which required a greater transformation of the data from the train to the test condition (e.g., for small, middle and large sized stimuli, train small/test large requires a bigger transformation than train middle/test large) resulted in later decoding onsets.

Successful object decoding across changes in retinal size or position suggests access to more abstract object representations, which is likely to reflect higher-level visual processing. Isik et al. (2014) compared their results to the different levels of the computational HMAX model (Riesenhuber and Poggio, 1999; Serre et al., 2007), which includes multiple stages representing early V1 processing (layer C1), mid-ventral processing (later C2), and higher level IT-like processing (layer C3). Comparison of the experimental results with the response of the HMAX model for their six object exemplars revealed that only the IT layer of HMAX could successfully decode object exemplars across all transformations of size and position. Isik et al. suggest their observed relationship between the degree of image transformation and decoding time (i.e., that decoding occurs later for larger visual transformations across object size and position) is consistent with a feedforward processing hierarchy supporting invariant object representations. In the next section we consider the link between object category representations and human behaviour.

3. Linking dynamic object category representations to human behaviour

One of the critical limitations of decoding approaches in neuroscience is that successful decoding of a stimulus property or experimental condition does not entail that this information is explicitly used by the brain for behaviour (cf. de-Wit et al., 2016; Ritchie et al., in press). An emerging goal is to find ways to link measured brain activation patterns to behaviour (Mack and Palmeri, 2011; Nosofsky and Palmeri, 1997; Philiastides and Sajda, 2006). In this section we review two recent approaches to linking the decoded dynamic object representations from time-series neuroimaging data with human behaviour. In the first approach, the goal is to link reaction times (RTs) for object categorisation to the brain's evolving representation of the objects in multidimensional space (Section 3.1). The second involves uncovering the effects of attention on decodable object representations (Section 3.2).

3.1. Categorisation reaction times

One fruitful approach to evaluate the relationship between brain and behaviour has been to link human RTs for categorisation to the multidimensional representation of object exemplars in neuroimaging data (Ritchie and Carlson, 2016). Human categorisation behaviour can be conceptualised as a process of evidence accumulation leading to a decision. Behavioural RTs, such as those measured by a simple button press to categorise a stimulus (e.g., fish as 'animate', boot as 'inanimate'), can be thought of as a proxy for the point at which a decision has been made. Carlson et al. (2014) tested this idea by combining previously collected fMRI data from human IT (Kriegeskorte et al., 2008) with separately collected behavioural RTs for categorising the same 92 object images as either 'animate' or 'inanimate'. Carlson et al. (2014) linked the decision boundary from the fMRI data (boundary is that used by a classifier to categorise stimuli by animacy) to RTs for human observers to complete the same categorisation task. They showed that the distance to the classifier decision boundary predicted reaction times for animacy categorisation. Specifically, objects represented further from the animacy decision boundary were associated with faster behavioural categorisation RTs (i.e., more easily categorised by human observers) than objects represented closer to the boundary (Fig. 7a).

Ritchie et al. (2015) expanded on this idea by studying the dynamic coupling between emerging object representations and behaviour using MEG decoding. Their findings replicated the relationship between representational distance and RTs found by Carlson et al. (2014) for fMRI, as a larger distance from the classifier boundary was associated with a faster reaction time for animacy categorisation. The key advantage of examining this relationship with MEG is that its emergence following stimulus onset can be tracked over time, which cannot be determined from the fMRI data. Ritchie et al. (2015) found that as decoding accuracy increased following stimulus onset (indicating greater separability of individual object representations with more processing time), the correlation between representational distance and RTs similarly increased. Importantly, this suggests that the strength of the association between representations and behaviour appears to follow decoding performance over time (Ritchie et al., 2015). The significance of this result is that it implies that the brain "reads out" information at the optimal time for making a categorical
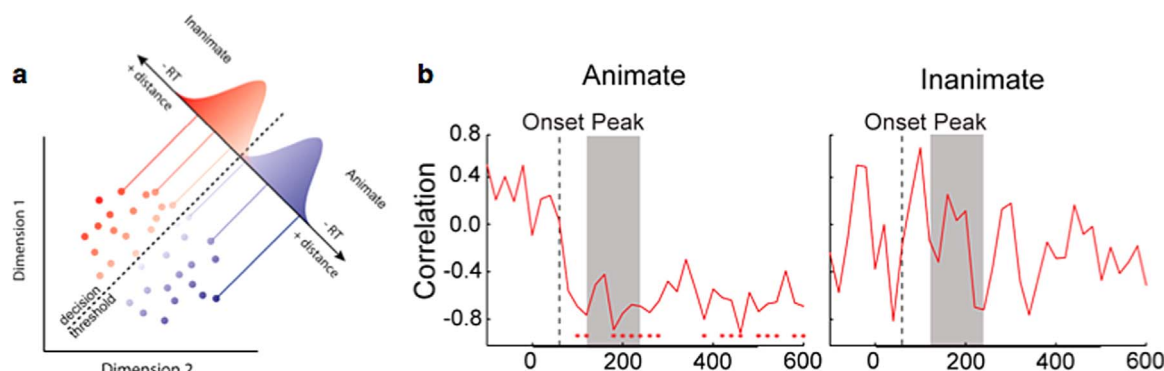
Fig. 7. Linking behavioural RTs for object animacy categorisation to dynamic object representations measured with MEG. (a) Diagram illustrating how representational distance from the classifier category boundary relates to RTs. (b) Left panel shows the correlation between RTs and the MEG decoding time-course for animate objects, while the right panel shows the same for inanimate objects. Grey bar in both plots indicates the window of peak decoding. Red stars along the x-axis indicate time-points at which the correlation between representational distance and reaction time is significant. Images from Ritchie et al. (2015) reprinted with permission from the author.
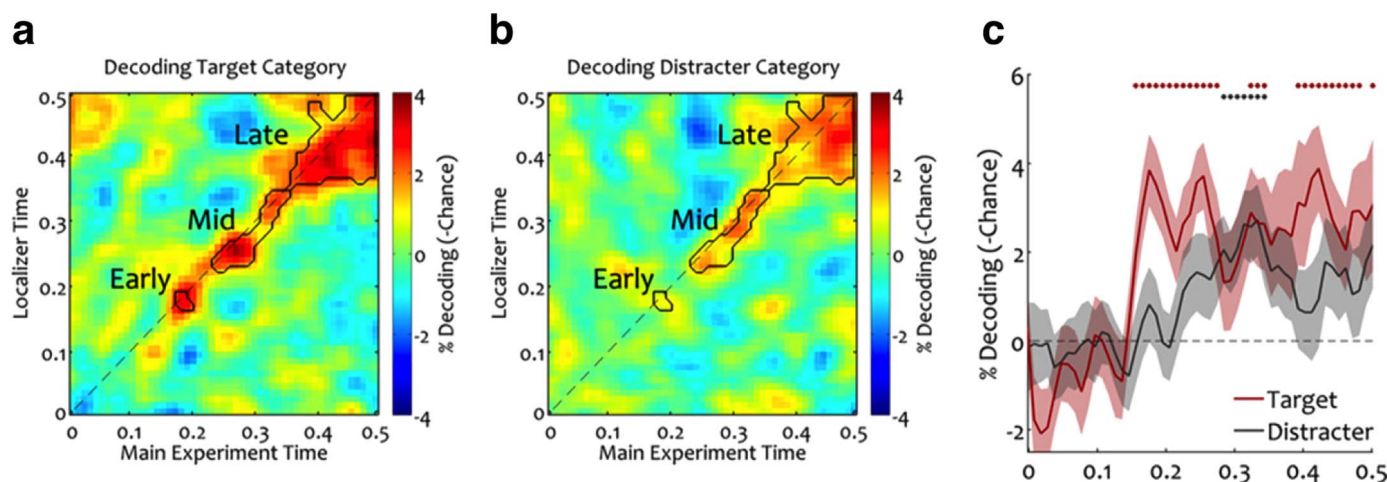

Fig. 8. MEG decoding of attended versus unattended targets. Temporal cross-decoding heat-maps of (a) target category and (b) unattended distractor category (training on isolated objects, testing on objects embedded in cluttered natural scenes in both instances); (c) Time-course of decoding performance for the target and distractor conditions separately. Dots indicate above-chance decoding performance. Images reprinted with permission of Society for Neuroscience, from Kaiser et al. (2016b); permission conveyed through Copyright Clearance Center, Inc.

decision. This highlights the utility of time-series decoding methods, as this result could not be accessed by examining time-averaged activation patterns as in fMRI.

Interestingly, in both the fMRI and MEG studies, the relationship between representational distance and RTs appeared to be driven by animate rather than inanimate objects (Carlson et al., 2014, Richie et al., 2015). This is shown in Fig. 7b by comparing the left (animate) and right (inanimate) RT correlation panels. For animate object decoding, a negative correlation is observed with RTs around the time of peak decoding (indicating that faster RTs were associated with greater distances from the decision boundary for animacy). However, this relationship is not found with inanimate objects. The authors suggest this asymmetry may be because inanimacy is a negatively defined category (i.e., 'not animate').

3.2. Attention

Attention has strong modulatory effects on neural processing (Gandhi et al., 1999; Kastner and Ungerleider, 2000; Moran and Desimone, 1985). A recent MEG study by Kaiser et al. (2016b) evaluated the effects of top-down attention on object category representations. Participants performed a category search task by detecting the presence of either a car or a person embedded in cluttered natural scenes. This design allows comparison of the neural signal associated with the attended versus unattended object category. The classifier was trained on data from viewing car and people exemplars that were

separate to the experimental data set and presented as segmented objects on a plain background. The classifier was then tested on the data for when the objects were embedded in complex natural scenes as either a target or a distractor.

The results demonstrate a substantial effect of attention on object representations. Under attended conditions, object category could be decoded rapidly in under 200 ms even though the objects were in highly cluttered scenes (Fig. 8c). However, unattended object categories embedded in complex scenes were decoded much later (Fig. 8c), demonstrating a strong top-down modulation of attention on visual processing. Attended objects were decoded more easily than unattended objects across all early (180 – 220 ms), mid (230 – 340 ms), and late (350 – 500 ms) time windows (Fig. 8a, b). The greater temporal generalisation of cross-decoding during the late time window indicates that the representational similarity shared by isolated objects and objects in scenes is sustained for longer in the late processing stages, which is particularly evident for attended objects (Kaiser et al., 2016b). These results suggest that top-down attention rapidly modulates object category representations, facilitating rapid target detection in natural scenes (Thorpe et al., 1996). By using temporal decoding, this study was able to reveal the time course of category-specific attentional enhancement, extending previous EEG findings which instead focused on comparing evoked potentials to target presence versus absence (Codispoti et al., 2006; Thorpe et al., 1996).

4. Summary and conclusions

Above we have reviewed how time-series decoding methods have expanded our understanding of the neural processes underlying human object recognition. Object recognition remains a significant area of interest within the field of vision research and the recent implementation of time-series decoding methods has allowed researchers to unpack some of the neural processes that underlie *how* the brain achieves this highly complex and multifaceted task. Complementing other neuroimaging techniques, such as fMRI, which provide static snapshots of the visual object processing hierarchy, time-series decoding has unveiled a system of dynamically changing processes that extend throughout the brain. This new perspective brings a host of interesting avenues for further exploration. For example, at what point in time do high-level constructs of attention and consciousness modulate the dynamics of object processing, and what is the effect of this modulation on perception? How do space- and feature-based attention processes differentially influence the processing hierarchy? There remain unanswered questions about the processes involved in visual working memory and how these representations in the mind's eye compare to the perception of a physical stimulus; are they fundamentally different or is visual working memory a process of revisiting points within these dynamic representations? Finally, it would be of particular interest to know how these various modulations of object representations differentially influence behaviour, further strengthening our understanding of the connection between brain and behaviour. Time-series decoding methods have greatly enriched our knowledge of neural object recognition processes, and show great potential for further advances in this and many other domains.

References

Altmann, C.F., Bülthoff, H.H., Kourtzi, Z., 2003. Perceptual organization of local elements into global shapes in the human visual cortex. Curr. Biol. 13 (4), 342–349. http://dx.doi.org/10.1016/S0960-9822(03)00052-6.

Averbeck, B.B., Latham, P.E., Pouget, A., 2006. Neural correlations, population coding and computation. Nat. Rev. Neurosci. 7 (5), 358–366.

Barragan-Jason, G., Cauchoix, M., Barbeau, E.J., 2015. The neural speed of familiar face recognition. Neuropsychologia 75, 390–401. http://dx.doi.org/10.1016/j.neuropsychologia.2015.06.017.

Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2002. Parallel visual motion processing streams for manipulable objects and human movements. Neuron 34 (1), 149–159.

Biederman, I., 1987. Recognition-by-components: a theory of human image understanding. Psychol. Rev. 94 (2), 115.

Carlson, Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011. High temporal resolution decoding of object position and category. J. Vis. 11 (10), 9. http://dx.doi.org/10.1167/11.10.9.

Carlson, Schrater, P., He, S., 2003. Patterns of activity in the categorical representations of objects. J. Cogn. Neurosci. 15 (5), 704–717. http://dx.doi.org/10.1162/jocn.2003.15.5.704.

Carlson, T.A., Ritchie, J.B., Kriegeskorte, N., Durvasula, S., Ma, J., 2014. Reaction time for object categorization is predicted by representational distance. J. Cogn. Neurosci. 26 (1), 132–142. http://dx.doi.org/10.1162/jocn_a_00476.

Carlson, Tovar, Alink, A D.A., Kriegeskorte, N., 2013. Representational dynamics of object vision: the first 1000 ms. J. Vis. 13 (10). http://dx.doi.org/10.1167/13.10.1.

Cauchoix, M., Barragan-Jason, G., Serre, T., Barbeau, E.J., 2014. The neural dynamics of face detection in the wild revealed by MVPA. J. Neurosci. 34 (3), 846–854. http://dx.doi.org/10.1523/JNEUROSCI.3030-13.2014.

Chait, M., Poeppel, D., Cheveigné, A. de, Simon, J.Z., 2007. Processing asymmetry of transitions between order and disorder in human auditory cortex. J. Neurosci. 27 (19), 5207–5214. http://dx.doi.org/10.1523/JNEUROSCI.0318-07.2007.

Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. Nat. Neurosci. 2 (10), 913–919.

Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., Oliva, A., 2016. Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Sci. Rep. 6. http://dx.doi.org/

10.1038/srep27755.

Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17 (3), 455–462. http://dx.doi.org/10.1038/nn.3635.

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. http://dx.doi.org/10.1016/j.neuroimage.2015.07.011.

Clarke, A., Devereux, B.J., Randall, B., Tyler, L.K., 2014. Predicting the time course of individual objects with MEG. Cereb. Cortex. http://dx.doi.org/10.1093/cercor/bhu203.

Clarke, A., Tyler, L.K., 2014. Object-specific semantic coding in human perirhinal cortex. J. Neurosci. 34 (14), 4766–4775. http://dx.doi.org/10.1523/JNEUROSCI.2828-13.2014.

Codispoti, M., Ferrari, V., Junghöfer, M., Schupp, H.T., 2006. The categorization of natural scenes: brain attention networks revealed by dense sensor ERPs. NeuroImage 32 (2), 583–591. http://dx.doi.org/10.1016/j.neuroimage.2006.04.180.

Connolly, A.C., Guntupalli, J.S., Gors, J., Hanke, M., Halchenko, Y.O., Wu, Y.-C., Haxby, J.V., 2012. The representation of biological classes in the human brain. J. Neurosci. 32 (8), 2608–2618. http://dx.doi.org/10.1523/JNEUROSCI.5547-11.2012.

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 (2 Pt 1), 261–270.

de-Wit, L., Alexander, D., Ekroll, V., Wagemans, J., 2016. Is neuroimaging measuring information in the brain? Psychonomic Bull. Rev. 23 (5), 1415–1428. http://dx.doi.org/10.3758/s13423-016-1002-0.

DiCarlo, J.J., Cox, D.D., 2007. Untangling invariant object recognition. Trends Cogn. Sci. 11 (8), 333–341.

Downing, P., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. Science 293 (5539), 2470–2473.

Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. Nature 392 (6676), 598–601.

Gandhi, S.P., Heeger, D.J., Boynton, G.M., 1999. Spatial attention affects brain activity in human primary visual cortex. Proc. Natl. Acad. Sci. 96 (6), 3314–3319. http://dx.doi.org/10.1073/pnas.96.6.3314.

Goddard, E., Carlson, T.A., Dermody, N., Woolgar, A., 2016. Representational dynamics of object recognition: feedforward and feedback information flows. NeuroImage 128, 385–397. http://dx.doi.org/10.1016/j.neuroimage.2016.01.006.

Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. Vis. Res. 41 (10–11), 1409–1422.

Grill-Spector, K., Malach, R., 2004. The human visual cortex. Annu. Rev. Neurosci. 27 (1), 649–677. http://dx.doi.org/10.1146/annurev.neuro.27.070203.144220.

Grootswagers, T., Wardle, S.G., Carlson, T.A., 2016. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time-series neuroimaging data. J. Cogn. Neurosci.. http://dx.doi.org/10.1162/jocn_a_01068.

Haxby, J.V., Connolly, A.C., Guntupalli, J.S., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37 (1), 435–456. http://dx.doi.org/10.1146/annurev-neuro-062012-170325.

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 (5539), 2425–2430. http://dx.doi.org/10.1126/science.1063736.

Haynes, J.-D., 2015. A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. Neuron 87 (2), 257–270.

Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. 7 (7), 523–534. http://dx.doi.org/10.1038/nrn1931.

He, K., Zhang, X., Ren, S., Sun, J., 2015.Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Computer Vision (ICCV), 2015 IEEE International Conference on* pp. 1026–1034. IEEE.

Hung, C.P., Kreiman, G., Poggio, T., DiCarlo, J.J., 2005. Fast readout of object identity from macaque inferior temporal cortex. Science 310 (5749), 863–866.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. Proc. Natl. Acad. Sci. 96 (16), 9379–9384. http://dx.doi.org/10.1073/pnas.96.16.9379.

Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T., 2014. The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. 111 (1), 91–102. http://dx.doi.org/10.1152/jn.00394.2013.

Jaccard, P., 1901. Etude comparative de la distribution florale dans une portion des Alpes et du Jura. Impr. Corbaz.

Kaiser, D., Azzalini, D.C., Peelen, M.V., 2016a. Shape-independent object category responses revealed by MEG and fMRI decoding. J. Neurophysiol. 115 (4), 2246–2250. http://dx.doi.org/10.1152/jn.01074.2015.

Kaiser, D., Oosterhof, N.N., Peelen, M.V., 2016b. The neural dynamics of attentional selection in natural scenes. J. Neurosci. 36 (41), 10522–10528. http://dx.doi.org/10.1523/JNEUROSCI.1385-16.2016.

Kaneshiro, B., Guimaraes, M.P., Kim, H.-S., Norcia, A.M., Suppes, P., 2015. A representational similarity analysis of the dynamics of object processing using single-trial EEG classification. PLOS One 10 (8), e0135697. http://dx.doi.org/10.1371/journal.pone.0135697.

Kanwisher, N., Mcdermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kastner, S., Ungerleider, L.G., 2000. Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci. 23 (1), 315–341. http://dx.doi.org/10.1146/annurev.neuro.23.1.315.

Kiani, R., Esteky, H., Mirpour, K., Tanaka, K., 2007. Object category structure in

response patterns of neuronal population in monkey inferior temporal cortex. J. Neurophysiol. 97 (6), 4296–4309. http://dx.doi.org/10.1152/jn.00024.2007.

King, J.-R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. 18 (4), 203–210. http://dx.doi.org/10.1016/j.tics.2014.01.002.

Kravitz, D.J., Saleem, K.S., Baker, C.I., Ungerleider, L.G., Mishkin, M., 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. Trends Cogn. Sci. 17 (1), 26–49. http://dx.doi.org/10.1016/j.tics.2012.10.011.

Kriegeskorte, N., 2008. Representational similarity analysis–connecting the branches of systems neuroscience. Front. Syst. Neurosci.. http://dx.doi.org/10.3389/neuro.06.004.2008.

Kriegeskorte, N., 2015. Deep neural networks: a new framework for modelling biological vision and brain information processing. bioRxiv, 29876. http://dx.doi.org/10.1101/029876.

Kriegeskorte, N., Kievit, R.A., 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cogn. Sci. 17 (8), 401–412. http://dx.doi.org/10.1016/j.tics.2013.06.007.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Bandettini, P.A., 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron 60 (6), 1126–1141. http://dx.doi.org/10.1016/j.neuron.2008.10.043.

Liu, H., Agam, Y., Madsen, J.R., Kreiman, G., 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 62 (2), 281–290. http://dx.doi.org/10.1016/j.neuron.2009.02.025.

Mack, M.L., Palmeri, T.J., 2011. The timing of visual object Categorization. Front. Psychol. 2. http://dx.doi.org/10.3389/fpsyg.2011.00165.

Malach, R., Levy, I., Hasson, U., 2002. The topography of high-order human object areas. Trends Cogn. Sci. 6 (4), 176–184.

Martin, A., Wiggs, C. L., Ungerleider, L. G., Haxby, J. V., 1996. Neural correlates of category-specific knowledge. Retrieved from http://www.nature.com/nature/journal/v379/n6566/abs/379649a0.html.

Mervis, C.B., Rosch, E., 1981. Categorization of natural objects. Annu. Rev. Psychol. 32 (1), 89–115.

Moran, J., Desimone, R., 1985. Selective attention gates visual processing in the extrastriate cortex. Front. Cogn. Neurosci. 229, 342–345.

Nosofsky, R.M., Palmeri, T.J., 1997. An exemplar-based random walk model of speeded classification. Psychol. Rev. 104 (2), 266.

Nowak, L.G., Bullier, J., 1997. The Timing of Information Transfer in the Visual System. In: Rockland, K.S., Kaas, J.H., Peters, A. (Eds.), Extrastriate Cortex in Primates. Springer, US, 205–241. http://dx.doi.org/10.1007/978-1-4757-9625-4_5.

O'Toole, A.J., Jiang, F., Abdi, H., Haxby, J.V., 2005. Partially distributed representations of objects and faces in ventral temporal cortex. J. Cogn. Neurosci. 17 (4), 580–590.

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

Philiastides, M.G., Sajda, P., 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb. Cortex 16 (4), 509–518.

Ramkumar, P., Jas, M., Pannasch, S., Hari, R., Parkkonen, L., 2013. Feature-specific information processing precedes concerted activation in human visual cortex. J. Neurosci. 33 (18), 7691–7699. http://dx.doi.org/10.1523/JNEUROSCI.3905-12.2013.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nat. Neurosci. 2 (11), 1019–1025. http://dx.doi.org/10.1038/14819.

Ritchie, J.B., Carlson, T.A., 2016. Neural Decoding and 'Inner' Psychophysics: A Distance-to-Bound Approach for Linking Mind, Brain, and Behavior. *Front. Neurosci. 10: 190. Doi:* http://dx.doi.org/10.3389/Fnins.

Ritchie, J.B., Kaplan, D.M., Klein, C., in press. Decoding the brain: Neural representation and the limits of multivariate pattern analysis in cognitive neuroscience. British Journal for the Philosophy of Science, http://dx.doi.org/10.1101/127233.

Ritchie, J.B., Tovar, D.A., Carlson, T.A., 2015. Emerging object representations in the visual system predict reaction times for categorization. PLoS Comput Biol. 11 (6), e1004316. http://dx.doi.org/10.1371/journal.pcbi.1004316.

Schmolesky, M.T., Wang, Y., Hanes, D.P., Thompson, K.G., Leutgeb, S., Schall, J.D., Leventhal, A.G., 1998. Signal timing across the macaque visual system. J. Neurophysiol. 79 (6), 3272–3278.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., Poggio, T., 2007. A quantitative theory of immediate visual recognition. In: T.D, Paul Cisek, J.F.K. (Eds.), Progress in Brain Research 165. Elsevier, 33–56. http://dx.doi.org/10.1016/S0079-6123(06)65004-8.

Sha, L., Haxby, J.V., Abdi, H., Guntupalli, J.S., Oosterhof, N.N., Halchenko, Y.O., Connolly, A.C., 2015. The animacy continuum in the human ventral vision pathway. J. Cogn. Neurosci. 27 (4), 665–678. http://dx.doi.org/10.1162/jocn_a_00733.

Simanova, I., Gerven, M., van, Oostenveld, R., Hagoort, P., 2010. Identifying object categories from event-related EEG: toward decoding of conceptual representations. PLOS One 5 (12), e14465. http://dx.doi.org/10.1371/journal.pone.0014465.

Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381 (6582), 520–522.

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

Van Essen, D.C., Anderson, C.H., Felleman, D.J., 1992. Information processing in the primate visual system: an integrated systems perspective. Science 255 (5043), 419.

Vida, M.D., Nestor, A., Plaut, D.C., Behrmann, M., 2017. Spatiotemporal dynamics of similarity-based neural representations of facial identity. Proc. Natl. Acad. Sci. 114 (2), 388–393.

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. http://dx.doi.org/10.1016/j.neuroimage.2016.02.019.

Wardle, S.G., Ritchie, J.B., 2014. Can object category-selectivity in the ventral visual pathway be explained by sensitivity to low-level image properties? J.Neurosci. 34 (45), 14817–14819. http://dx.doi.org/10.1523/JNEUROSCI.3566-14.2014.