



REVIEW

# Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context [version 1; peer review: 2 approved]

Susan G. Wardle , Chris I. Baker 

Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, 20892, USA

**v1** **First published:** 11 Jun 2020, 9(F1000 Faculty Rev):590  
<https://doi.org/10.12688/f1000research.22296.1>  
**Latest published:** 11 Jun 2020, 9(F1000 Faculty Rev):590  
<https://doi.org/10.12688/f1000research.22296.1>

**Abstract**

Object recognition is the ability to identify an object or category based on the combination of visual features observed. It is a remarkable feat of the human brain, given that the patterns of light received by the eye associated with the properties of a given object vary widely with simple changes in viewing angle, ambient lighting, and distance. Furthermore, different exemplars of a specific object category can vary widely in visual appearance, such that successful categorization requires generalization across disparate visual features. In this review, we discuss recent advances in understanding the neural representations underlying object recognition in the human brain. We highlight three current trends in the approach towards this goal within the field of cognitive neuroscience. Firstly, we consider the influence of deep neural networks both as potential models of object vision and in how their representations relate to those in the human brain. Secondly, we review the contribution that time-series neuroimaging methods have made towards understanding the temporal dynamics of object representations beyond their spatial organization within different brain regions. Finally, we argue that an increasing emphasis on the context (both visual and task) within which object recognition occurs has led to a broader conceptualization of what constitutes an object representation for the brain. We conclude by identifying some current challenges facing the experimental pursuit of understanding object recognition and outline some emerging directions that are likely to yield new insight into this complex cognitive process.

**Keywords**

object recognition, human vision, fMRI, MEG, DNN, visual perception

**Open Peer Review**

**Reviewer Status**  

	Invited Reviewers	
	1	2
<b>version 1</b> 11 Jun 2020		

**Faculty Reviews** are written by members of the prestigious **Faculty Opinions Faculty**. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

- 1 **Aude Oliva**, MIT, Cambridge, USA
- 2 **Marius Peelen**, Radboud University, Nijmegen, The Netherlands

Any comments on the article can be found at the end of the article.

**Corresponding author:** Susan G. Wardle ([susan.wardle@nih.gov](mailto:susan.wardle@nih.gov))

**Author roles:** **Wardle SG:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Baker CI:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work is supported by the Intramural Research Program of the National Institute of Mental Health (ZIAMH002909). *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2020 Wardle SG and Baker CI. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The author(s) is/are employees of the US Government and therefore domestic copyright protection in USA does not apply to this work. The work may be protected under the copyright laws of other jurisdictions when used in those jurisdictions.

**How to cite this article:** Wardle SG and Baker CI. **Recent advances in understanding object recognition in the human brain: deep neural networks, temporal dynamics, and context [version 1; peer review: 2 approved]** F1000Research 2020, 9(F1000 Faculty Rev):590 <https://doi.org/10.12688/f1000research.22296.1>

**First published:** 11 Jun 2020, 9(F1000 Faculty Rev):590 <https://doi.org/10.12688/f1000research.22296.1>

## Introduction

Object recognition is one of the classic “problems” of vision<sup>1</sup>. The underlying neural substrate in humans was revealed by classic neuropsychological studies which pointed to selective deficits in visual object recognition following lesions to specific brain regions<sup>2,3</sup>, yet we still do not understand how the brain achieves this remarkable behavior. How is it that we reliably<sup>4</sup> and rapidly<sup>5</sup> recognize objects despite considerable retinal image transformations arising from changes in viewing angle, position, image size, and lighting? Much experimental and computational work has focused on this problem of invariance<sup>4,6–13</sup>. Early neuroimaging studies of object recognition using functional magnetic resonance imaging (fMRI) focused on regions in the lateral occipital and ventral temporal cortex, which were found to respond more strongly to the presentation of objects than to textures or scrambled objects<sup>14,15</sup>. More recently, the application of multivariate analysis techniques has led to broader investigation of the structure of object representations<sup>a</sup> throughout the ventral temporal cortex<sup>16,17</sup> and their temporal dynamics across the whole brain<sup>18,19</sup>. While these representations are assumed to contribute to object recognition behavior, they may also contribute to other tasks. This shift toward object representations has also accompanied a greater focus on revealing how a broad range of different object categories are represented rather than investigating the invariant representation of single objects. Such object categorization involves a similar issue of extrapolation across changes in visual features as invariance, since exemplars (e.g. Great Dane and Chihuahua) of a category (e.g. “dog”) often have significantly different visual features from one another.

The aim of this review is to provide an overview of recent advances in understanding object recognition in the human brain. In this review, we primarily consider contemporary work from the past three years in human cognitive neuroscience, identifying the current trends in the field rather than providing an exhaustive summary. In addition, we focus on the neural basis of visual object recognition in the human brain (for reviews including non-human primate studies, see 20,21) rather than the related topics of computer vision, object memory, and semantic object knowledge. We define visual objects as meaningful conjunctions of visual features<sup>13</sup> and object recognition as the ability to distinguish an object identity or category from all other objects<sup>21</sup>. Face recognition is not covered in this review, as faces are a unique object class that are processed within a specialized network of regions<sup>22,23</sup>.

We identify three current trends in the approach towards understanding object recognition within the field of cognitive neuroscience. Firstly, the rapidly growing popularity of deep neural networks (DNNs) has influenced both the type of analytic approach used and the framework from which the questions are asked. Secondly, the adaptation of multivariate methods

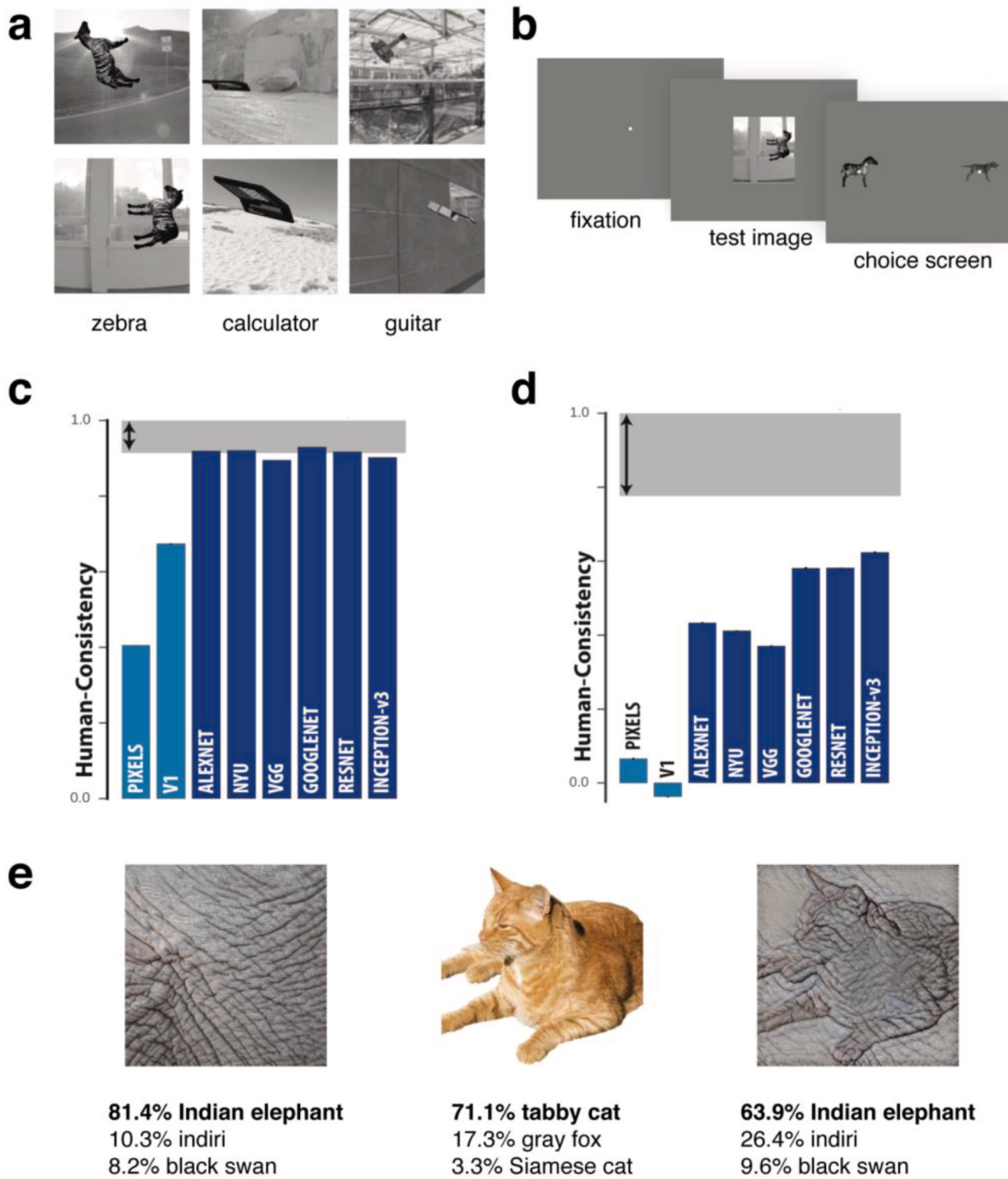
to time-series neuroimaging methods such as magnetoencephalography (MEG) and electroencephalography (EEG) has highlighted the importance of considering the temporal dynamics in the neural processing of object recognition at a resolution not accessible with fMRI. Finally, the field has begun to move away from examining single objects in isolation towards examining objects within more naturalistic contexts including a variety of both task and visual contexts. In the sections below, we examine each of these trends in turn.

## Deep neural networks as models of object vision

DNNs are a class of brain-inspired computer vision algorithms<sup>24–26</sup>. Although there are many variants of the specific network architecture, the term DNN refers to artificial neural networks in which there are multiple (i.e. “deep”) layers in-between the input and output stages<sup>27</sup>. DNNs have risen to prominence within cognitive neuroscience relatively recently given high levels of performance in object classification<sup>28</sup>, in some cases even performing as well as humans<sup>29</sup>. This has led to consideration of the utility of DNNs as potential models of biological vision<sup>26,30</sup>. However, overall performance does not necessarily indicate that the underlying processing is similar to that in the brain. In this section, we highlight several fundamental differences between state-of-the-art DNNs and the brain and consider the potential of DNNs to inform our understanding of human object recognition given these differences.

DNNs have recently achieved human levels of performance in terms of accuracy for image classification<sup>29</sup>. Specifically, this had been achieved for images from the large database ImageNet and not yet for real world images taken in the wild. An interesting question is to what degree the pattern of successful classification and errors made by DNNs mirror those made by humans making perceptual judgments. Several studies have reported both similarities and differences between human behavior and DNNs. For example, while DNNs can capture human shape sensitivity (with stimuli very different to those on which they were trained)<sup>31</sup>, they perform less well than simple categorical models in capturing similarity judgements<sup>32,33</sup> and do not capture human sensitivity to properties such as symmetry<sup>33</sup>. One study that revealed clear differences between human and DNN representations compared the performance of humans, macaque monkeys, and DNNs on an invariant object recognition task<sup>34</sup>. Stimuli were rendered 3D objects of 24 basic-level categories (e.g. zebra, calculator) superimposed on a natural image background at different orientations/viewpoints (Figure 1a). Monkey and human subjects viewed these images and then a binary response screen with two objects in canonical view was shown, and their task was to match the object from the previous stimulus (Figure 1b). Notably, while results for object-level confusion were similar among humans, monkeys, and DNNs (Figure 1c), performance at the image level did not match between domains (Figure 1d). This difference in error patterns suggests that accuracy is not an adequate measure of the similarity between humans and DNNs, as vastly different response patterns can yield comparable accuracy.

<sup>a</sup> We use the term ‘object representations’ here to mean the measured patterns of response in the brain associated with object perception, rather than a specific internal representation.



**Figure 1. Deep neural networks and object recognition.** (a–b) Stimuli and behavioral experimental design used in 34. On each trial, human and monkey observers briefly viewed a synthetic test image of a rotated object placed on a random scene background. They then reported which object had been presented by making a binary choice from one of two objects presented in canonical view on the test screen. (c) Results showed that for object category, humans and several different deep neural networks (DNNs) performed similarly. However, humans made different errors than DNNs at the image level (d). (e) Example images from 35. A DNN was more likely to classify an image based on texture (Indian elephant) than shape (tabby cat), whereas human observers do the reverse. Figures a-d were adapted from Rajalingham *et al.*<sup>34</sup> under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

The observation that humans and DNNs do not show similar patterns of errors at the image level implies that DNNs and humans are not solving the task in the same way or are not relying on the same source of information. A striking demonstration showed that DNNs can be fooled into misclassifying an object by making small changes to the image that are barely perceptible to human observers<sup>34</sup>. The human visual system is also better able to generalize classification across different forms of noise than DNNs<sup>36</sup>. An example of clear divergence in the source of information used by humans compared to DNNs is the demonstration that DNNs may favor texture over shape in classifying objects, with the reverse true for human observers<sup>35</sup>. For example, DNNs such as ResNet-50 trained on ImageNet labelled a picture of a tabby cat rendered with the texture of elephant skin as an “Indian elephant”, whereas human observers would label it as “cat” (Figure 1e). Interestingly, re-training the ResNet-50 architecture to learn a shape-based representation using stylized images in which texture was not predictive of object category led to performance more similar to human observers. Furthermore, there were surprising performance benefits that emerged from the shape-based network, such as greater tolerance to image distortions and better object detection performance.

Beyond comparing network performance with human behavior, recent studies have also compared the representations for objects and scenes within different layers of DNNs to human brain representations measured with fMRI or MEG<sup>37–45</sup>. Generally, these studies have found that lower layers of DNNs correlate more with earlier regions within the visual processing hierarchy and higher layers with later regions such as the ventral temporal cortex<sup>39,45–48</sup>. Similarly, time-resolved neuroimaging methods (see also next section) such as MEG have revealed that lower layers of DNNs correlate with human brain activity earlier in time than higher network layers<sup>37,40,47</sup>. However, substantial differences among the human brain, behavior, and DNN representations are also reported, which show that the relationship among them is complex<sup>38,39,41,44</sup>. For example, for a stimulus set that balanced animacy and appearance, DNNs represented animacy over visual appearance, with the opposite relationship in the ventral temporal cortex<sup>38</sup>. Similarly, despite striking differences in the representational structure of behavior and fMRI responses, they both showed strong correlations with DNN representations<sup>39</sup>. Critically, simply calculating correlations is not sufficient for characterizing the similarity between object representations in the human brain and the representations measured by human behavior or in artificial networks. This is because the correlation among these different representations (i.e. among the brain, behavior, and/or DNNs) can be equal in magnitude but explain different parts of the underlying variance. Fundamental progress will be made when we have better methods of revealing what is *driving* the correlation among representations in DNNs, behavior, and the human brain, where such correlations do exist.

There are several emerging directions that may increase the utility of DNNs for advancing our understanding of human object recognition. It is already clear that the link between

visual object representations in the brain and DNN representations for the same objects is not straightforward<sup>38,39,41</sup>. Most comparisons have been made with existing pre-trained DNNs; however, deeper insights are likely to emerge from training DNNs to test specific predictions<sup>35</sup>, which requires systematically varying the task or stimulus set. The addition of biologically plausible architecture to DNNs such as spike-timing-dependent plasticity and latency coding<sup>49,50</sup> may further facilitate the comparison of DNNs and the human brain. For example, the inclusion of recurrent connections more closely captures the dynamic representation of objects in the human brain<sup>51,52</sup>. Similarly, transforming the input images to DNNs in a manner similar to the perturbations resulting from the optics of the human eye, for example by applying retinal filters<sup>34</sup>, may increase the similarity in the underlying representations between these networks and the brain or behavior. One of the most interesting findings thus far has been that DNNs occasionally spontaneously demonstrate features of visual processing that mirror human perception such as generalization over shape or image distortion<sup>31,35</sup>. Examination of the conditions under which this occurs may be enlightening for understanding how the human brain achieves object recognition under much more varied viewing conditions and tasks than even state-of-the-art DNNs.

### The temporal dynamics in neural object representations

In recent years, the application of multivariate analyses to time-series neuroimaging methods such as MEG and EEG has facilitated new investigation into the temporal dynamics of cognitive processes. Visual object recognition has been one of the main subfields of cognitive neuroscience to first adapt these methods<sup>53</sup>. Object recognition is fast<sup>5</sup>; we can recognize an object in tens of milliseconds. This is much faster than the typical resolution of BOLD fMRI (e.g. 2 seconds); thus, unpacking the temporal evolution of object representations requires alternative neuroimaging methods with millisecond precision. Here we focus on recent work that has revealed the temporal dynamics of object representations in the human brain.

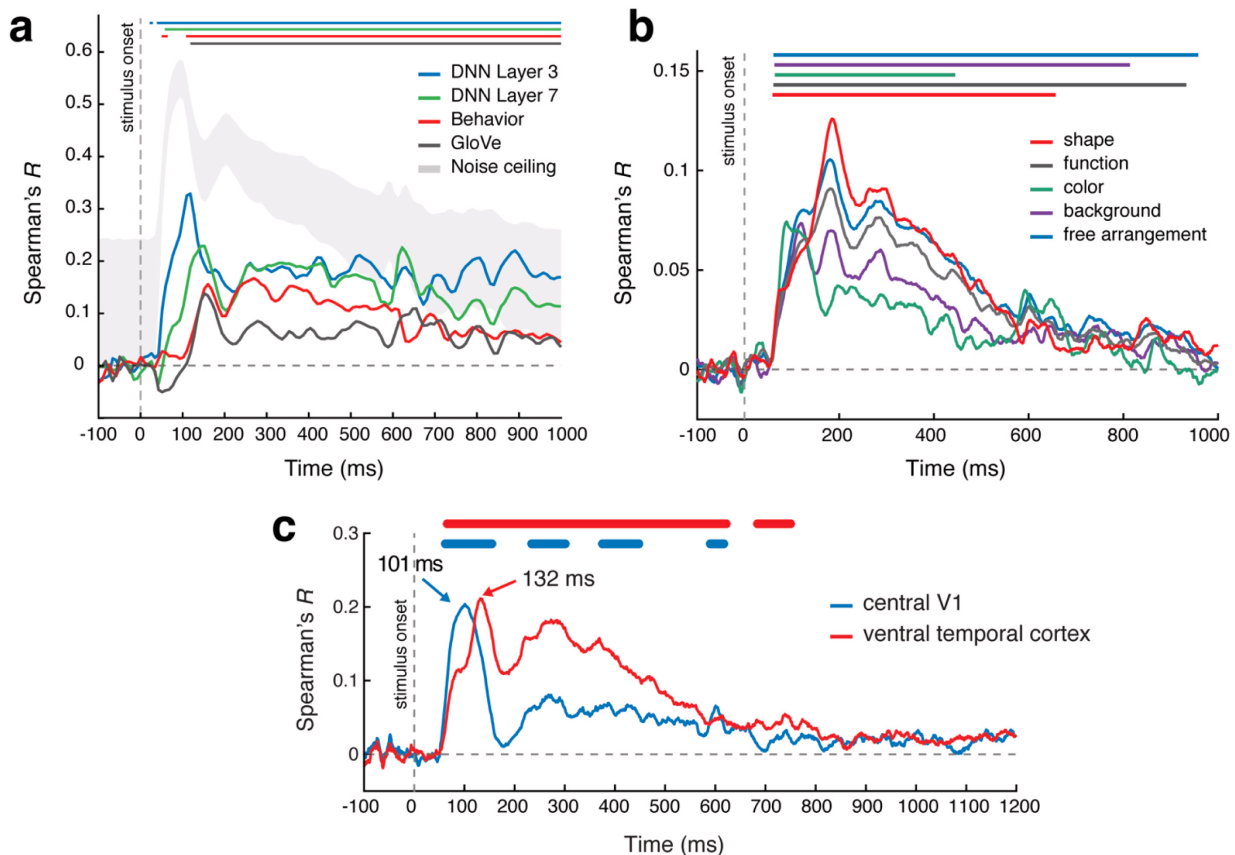
Object representations potentially reflect a number of different properties, which together can be considered to form an “object concept”<sup>54</sup>. For example, an object concept might include its visual features, the conceptual knowledge associated with an object such as its function, and its relationship to other objects. Neuroimaging methods with high temporal resolution offer the potential to examine the time course of the contribution of these different properties to the underlying object representations. MEG decoding studies have revealed that object identity and category can be decoded in under 100 milliseconds following visual stimulus onset<sup>18,19</sup>. The facilitation of objects presented in typical rather than atypical visual field locations occurs around 140 milliseconds<sup>55</sup>, suggestive of a relatively early contribution of expectation based on visual experience. In contrast, contextual facilitation for classifying the animacy of degraded objects in scenes compared to the same objects presented in the absence of scene context occurs relatively late, 320 milliseconds after stimulus onset, suggestive of a feedback mechanism<sup>56</sup>.



The contribution of conceptual information to object representations develops after initial visual processing. The emergence of categorical structure based on animacy and real-world object size occurs around 150 milliseconds<sup>37</sup>. This is consistent with estimates of the lower bound of the formation of conceptual object representations<sup>37</sup>. Using MEG data recorded for two stimulus sets of 84 object concepts, generalization across exemplars emerged ~150 milliseconds after onset. The shared semantic relationships between the objects was assessed with the Global Vectors for Word Representation (GloVe) model<sup>58</sup>, an unsupervised algorithm trained on word co-occurrences. Consistent with the time course of generalization around 150 milliseconds, Figure 2a shows that the correlation with the MEG data for behavioral similarity judgements on the stimuli and the GloVe model of semantic information based on word representations both peaked around this time and later than the correlation with representations of the stimuli from an early layer of a DNN. Similarly, the correlation between dynamic MEG representations of objects

on their natural backgrounds and measures of behavioral similarity based on shape, color, function, background, or free arrangement is all before 200 milliseconds<sup>41</sup> and consists of overlapping representations in time (Figure 2b). For individual object representations, a model that combines a visual feature model (e.g. HMax<sup>59</sup> or AlexNet<sup>28</sup> DNN) with a model of semantic features better predicts neural representations measured with MEG than using visual features alone<sup>60,61</sup>. The contribution of semantic information to object representations has been linked to activity in the perirhinal cortex<sup>62</sup> and anterior temporal cortex<sup>63</sup>. Collectively, these results are indicative of a relatively early role for conceptual information in object representations that follows the initial visual processing.

Another advantage of studying object representations with high temporal resolution is the potential to disentangle the role of feedforward versus feedback processes in their formation. Feedback is theoretically difficult to study empirically and although its role in visual perception has been acknowledged



**Figure 2. Temporal dynamics of object recognition.** (a) The correlation over time between magnetoencephalography (MEG) whole-brain object representations and the representations from several models based on deep neural network (DNN) layers, behavioral similarity judgments, and the Global Vectors for Word Representations (GloVe) model<sup>37</sup>. Note that the lower DNN layer has an earlier peak than the higher layer. (b) Correlation over time between whole-brain MEG object representations and models based on several different visual and conceptual features<sup>41</sup>. (c) Functional magnetic resonance imaging–MEG “fusion” reveals a peak correlation between whole-brain MEG object representations and those in the primary visual cortex (V1) at 101 milliseconds (ms) and ventral temporal cortex at 132 ms<sup>19</sup>. Figure b is adapted from Cichy *et al.*<sup>41</sup> under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

for decades, the advent of recurrent connections in DNNs<sup>52</sup> has reignited interest in attempting to separate the contribution of feedback vs. feedforward processes in object recognition. For example, a computational model incorporating recurrent connections could partially account for occluded object representations measured with MEG, which had a decoding peak much later in time than un-occluded objects<sup>43</sup>. This suggests feedback processes assist in processing objects under more ambiguous viewing conditions such as occlusion. One recent approach towards isolating the contribution of feedback has been to use the rapid serial visual presentation of objects at very brief presentations under the assumption that rapid presentation disrupts feedback processing of the preceding object(s)<sup>64,65</sup>.

One of the challenges the contribution of time-resolved neuroimaging has brought to light is how best to integrate fMRI results with MEG/EEG to elucidate the combined spatial and temporal processing of object recognition. One approach is to use source localization to model the spatial source of the MEG signal in the brain<sup>52</sup>. An alternative method, fMRI-MEG “fusion”, correlates dissimilarity matrices constructed separately from fMRI and MEG data over time (MEG) and regions of interest (fMRI)<sup>19,66</sup>. This approach has been used successfully to demonstrate that whole-brain object representations measured with MEG have a peak correlation earlier in time with the primary visual cortex (V1) and later in time with the ventral temporal cortex<sup>19,66</sup> (Figure 2c). Furthermore, fusion revealed temporal differences in the contribution of task versus object representations across the visual hierarchy<sup>67</sup>. Although these results provide a useful validation of the method, the interpretation of fusion results is not straightforward, particularly because of the substantial differences in the spatial resolution between fMRI and MEG. For example, one pair of studies used an object stimulus set that controlled for shape (e.g. snake and rope) across category in order to examine the influence of perceptual and categorical similarity on object representations. Even though the studies used identical stimuli, the results were different between the two neuroimaging modalities: they found more evidence for categorical similarity with fMRI<sup>68</sup> and perceptual similarity with MEG<sup>69</sup>.

The results reviewed above demonstrate the importance of understanding the temporal dynamics of object recognition. So far, multivariate methods applied to MEG and EEG data with high temporal resolution have yielded new insights into the temporal dynamics of semantic versus visual features in object representations and highlighted a possible role for feedback from higher visual areas in the representation of degraded and occluded objects as well as in task-relevant representations. The development of a new generation of wearable MEG systems based on arrays of optically pumped magnetometers promises further advancement in the measurement of brain activity at a high temporal resolution in more varied contexts<sup>70,71</sup>. Significant progress will be made with further improvements in linking spatial and temporal neuroimaging data.

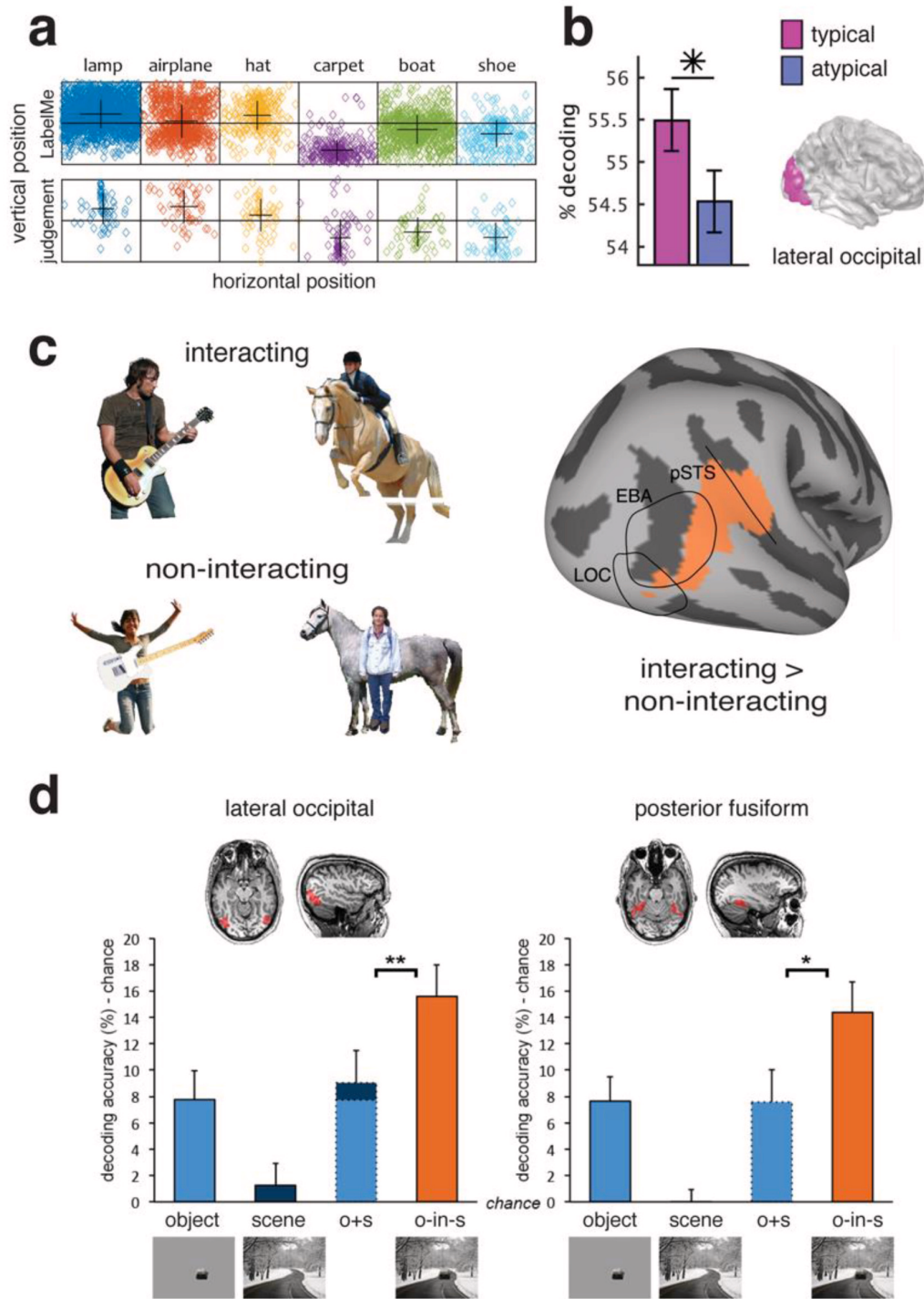
## Contextual effects on object representations

Traditionally, object perception has been studied empirically by presenting single objects in isolation on blank backgrounds<sup>17,72,73</sup>. This approach facilitates studying aspects of object recognition such as viewpoint and position invariance without a contribution from the background; however, it likely over-emphasizes the role of object shape. More recently, the context in which object recognition occurs has been increasingly considered in studies aiming to understand the underlying neural mechanisms. This can be the visual context, such as the placement of an object in a scene (either relevant<sup>39,56</sup> or irrelevant<sup>34</sup>), the action of an agent (e.g. person) involving the object<sup>74</sup>, or even the use of 3D real objects rather than 2D images<sup>75,76</sup>. Or, alternatively, this can be task context, with neural object representations measured as participants perform different tasks on the same object stimuli<sup>77</sup>. An advantage of all of these approaches with broader scope is that they examine object recognition in circumstances that more closely mimic real-world perception. The results we review here suggest that both visual and task context play a significant role in object processing.

### Visual context: interactions with people and scenes

The simplest form of visual context is to present two objects at a time instead of one. In object-selective cortex, the brain activation patterns to two objects are well-predicted by the average responses to the objects presented in isolation<sup>78,79</sup>. More recently, it has been shown that even without the visual context of a detailed scene, the brain representations of objects are affected by expectation driven by context. For example, a fMRI study looked at object pairs taken from scenes (such as a sofa and TV, car and traffic light) presented in their original location versus interchanged locations relative to each other on a blank background<sup>80</sup>. In the object-selective cortex, the mean of the activation patterns for two isolated objects presented centrally was less similar to the activation patterns for the object pairs when they were in their original location compared with the interchanged location, but this was not the case in the early visual cortex. This suggests that the object-selective cortex is sensitive to the expected location of different objects relative to each other.

A related observation is that the location of objects within scenes in the real world is not arbitrary, and objects occur within relatively predictable locations related to their function<sup>81</sup>. In some cases, this produces a statistical regularity in the visual field location (Figure 3c). There is some evidence that object processing is facilitated when this expectation is adhered to and objects occur in their typical retinotopic visual field location (i.e. their position relative to the direction of eye gaze). For example, in the object-selective cortex, objects in their typical visual field location (e.g. hat in upper visual field, shoe in lower visual field) could be decoded at a higher rate from the fMRI activation patterns than when they were in the atypical portion of the visual field (Figure 3d)<sup>82</sup>. Other higher visual areas in ventral temporal cortex did not show such a difference.



**Figure 3. Effect of visual context on the neural representation of objects.** (a) Some objects are associated with a typical visual field position in which they tend to occur<sup>82</sup>. The top row shows object locations from a labelled image database, and the bottom row shows the placement of objects by human observers. (b) In the lateral occipital cortex (LOC), decoding accuracy was higher for objects presented in their typical (e.g. hat in upper visual field) than their atypical (e.g. hat in lower visual field) location. (c) Example stimuli used in 74 of objects in “interacting” and “non-interacting” contexts. A decoding searchlight analysis revealed areas that had higher decoding accuracy for interacting than non-interacting objects. (d) Super-additive decoding accuracy in object-selective lateral occipital and posterior fusiform regions for degraded objects in scenes compared to decoding accuracy for isolated objects or scenes alone<sup>56</sup>. EBA, extrastriate body area; pSTS, posterior superior temporal sulcus Figure d is adapted from Brandman and Peelan<sup>56</sup> under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).



EEG results suggest there is a difference in the representation of objects in typical vs. atypical locations as early as 140 milliseconds after stimulus onset<sup>55</sup>. Overall, the sensitivity of the object-selective cortex to statistical regularities in the location of objects is consistent with the idea of efficient coding in the visual system<sup>83</sup>, which argues that statistical regularities in the environment can be exploited by neural coding in order to conserve the amount of brain resources engaged in representing the complex visual world.

Another consideration in the representation of multiple objects beyond their relative location is their function. An inherent property of objects is their manipulability, and several studies have investigated how this affects their neural representation<sup>74,84,85</sup>. The degree to which interactions with people and scenes mediates object representations is not homogenous across brain regions. For example, one study examined the effect of interaction on object representations using a stimulus set of humans, guitars, and horses<sup>74</sup> (Figure 3a). They measured brain responses to the isolated objects and for object pairs when they were both interacting (e.g. person riding a horse) and not interacting (e.g. person in front of a horse). In some brain regions, the representation of meaningfully interacting objects was not well predicted by the responses to their individual parts, suggesting coding of the object interaction. For example, a decoding searchlight analysis of the fMRI data revealed areas overlapping with the body-selective extrastriate body area (EBA) and posterior superior temporal sulcus (pSTS) that had higher decoding accuracy for interacting than non-interacting objects (Figure 3b).

Beyond simple object pairs, similar logic has also been applied to examine how scene context affects object representations. For example, one study measured BOLD activation patterns to degraded objects (either animate or inanimate) presented both in isolation and within intact scenes<sup>56</sup>. A classifier was trained to distinguish activation patterns associated with animate versus inanimate objects on separate data from intact isolated objects and then tested on the patterns associated with the degraded objects both in isolation and within a scene. In both lateral occipital and posterior fusiform regions, cross decoding accuracy for object animacy was significantly higher for the degraded objects within scenes than that predicted by accuracy for isolated degraded objects and isolated intact scenes (Figure 3c). However, in scene-selective regions, this was not the case, and decoding accuracy was only additive. These results suggest that object representations in object- but not scene-selective regions are enhanced by the presence of relevant visual context.

Collectively, the studies discussed above highlight the importance of considering the visual context in which objects occur. In the next section, we consider the importance of task context.

### Task context: the stability of object representations in visual cortex

Given that objects are both recognizable and actionable things, an important question is how the neural representation of

objects supports behavior. We can make a multitude of judgements about an object, as well as pick them up and use them in action. How do neural object representations change depending on the goal of the observer? In an experimental paradigm, this usually takes the form of keeping the visual stimuli constant and changing the task of the observer. Such changes in task may affect the relevant information and consequently change the distribution of attention. Within the higher visual cortex, where category-selectivity emerges, the majority of results seem to support fairly limited transformation of object representations as a function of task relative to the modulation by object type<sup>66,76,85–89</sup>. However, in the early visual cortex, there may be strong effects of task, potentially reflecting changes in spatial attention. Consistent with these generalizations, an MEG study found that the impact of task (semantic, e.g. classify the object as small or large, or perceptual, e.g. color discrimination) had a relatively late magnitude effect on object representations across the whole-brain MEG signal rather than a qualitative change to the nature of the representation<sup>67</sup>. Furthermore, MEG–fMRI fusion suggested the effect of task increased further up the processing hierarchy. Together, this suggests that other brain regions in addition to higher-level visual cortex have an important role in task modulation<sup>90</sup>. This is in contrast to the effect of visual context reviewed above, in which there was significant modulation of object representations in higher-level visual regions.

Consistent with a locus that is not restricted to the visual cortex, there is considerable evidence for a substantial role of parietal and frontal brain regions in the task modulation of object representations. For example, to address this question, one fMRI study used a stimulus set of 28 objects where semantic category and action associated with the objects were dissociated<sup>86</sup>. Participants performed two tasks on the same stimuli while within the fMRI scanner: rate objects on a four-point scale from very similar to very different for either hand action similarity or category similarity. For example, pictures of a drum and hammer would be similar for action/manipulation similarity, but drum and violin would be more similar for categorical similarity (both musical instruments). An analysis of the similarity of the brain activation patterns for the different objects revealed that in parietal and prefrontal areas, an action model of the stimuli correlated more with the similarity of object activation patterns during the action task, and vice versa for the category task<sup>86</sup>. Frontoparietal areas also showed greater within-task correlations than between-task, but this did not differ for occipitotemporal areas. Physical and perceived shape correlated with representations more in occipitotemporal regions. Consistent with this, there is evidence for a difference in the representational space of how objects are represented in occipitotemporal and posterior parietal regions<sup>91</sup>, with more flexible representations modulated by task in the posterior parietal cortex<sup>88</sup>.

Collectively, these findings suggest that while task context can affect object representations within the brain, these effects tend to be largest at higher stages of the visual hierarchy with strongest effects in the prefrontal and parietal cortex.

## “Beyond” object recognition

Here we have reviewed three current trends in the field of object recognition: the influence of DNNs, temporal dynamics, and the relevance of different forms of context. These trends have focused the field to consider object representations more broadly rather than object recognition *per se*. Such representations are likely critical for object recognition but will also contribute to many other behaviors<sup>92</sup>. To conclude, we briefly consider some current challenges in the pursuit of understanding object representations in the human brain and outline some emerging trends that are likely to help push the field forward.

The first issue we consider is what should “count” as an object representation. A consequence of the relevance of visual and task context reviewed in the section above is that it suggests object representations are broader than the particular conjunction of visual properties that visually define the object. The frequent investigation of the neural representation of isolated objects without context may have over-emphasized the role of shape in the underlying representations of real-world objects. Indeed, shape has been found to be a strong predictor of the similarity of the neural representations for different objects<sup>41</sup>. Similarly, the focus on functional object-selective brain regions, which are localized by contrasts between, for example, isolated objects and scrambled objects<sup>73</sup>, emphasizes the role of brain regions which are sensitive to shape above other object properties. However, there is evidence that other high-level regions such as scene-selective cortex<sup>85</sup> and parietal and prefrontal regions<sup>86,88,90,91</sup> are also engaged in object processing. Similar to the importance of visual and task context in object representations, further consideration of object-specific properties such as the role of color<sup>93,94</sup> and material properties<sup>95</sup> is likely to provide a new perspective on the nature of object representations. Object representations in the human brain are also tied to other features such as conceptual knowledge<sup>54</sup> about their function and relationship to other objects, which are yet to be emulated by DNNs in a way that produces the same flexibility as the human brain.

A second important issue in investigating the nature of object representations is stimulus selection and presentation. In the last decade, there has been concentrated effort to use larger stimulus sets ( $n = \sim 100$ ) of objects in neuroimaging event-related designs in an effort to reveal the inherent organization of object categories in brain representations without imposing stimulus groupings in the experimental design<sup>17</sup>. This is in contrast to blocked stimulus presentation, which is not

desirable for investigating representational structure because of inherent biases in the experimental design arising from grouping stimuli together into blocks. However, a limitation of representational similarity analysis<sup>96</sup> is that it is relative to the stimulus set used in the experiment and even with  $\sim 100$  stimuli there are likely to be inherent biases in the stimulus selection. For example, a stimulus set in which shape is a critical difference between stimuli is likely to emphasize a significant role for shape in the organization of the representational space. One recent approach that has potential to move the field forward is the use of very large stimulus sets. Recent databases of 5,000<sup>97</sup> and 26,000<sup>98</sup> visual object images have potential to reveal new insight that has not been possible using experimenter-selected restricted stimulus sets of  $\sim 100$  images. Additionally, the method used for image selection in the creation of these large stimulus sets is still important in avoiding biases. For example, the THINGS database<sup>98</sup> was created by systematically sampling concrete picturable and nameable nouns from American English in order to avoid any explicit or implicit biases in stimulus selection.

Finally, there has been considerable debate over what degree object representations are reducible to the low- and mid-level visual features that co-vary with category membership<sup>38,68,69,99–103</sup>. However, this question may be ill-posed. By definition, visual object representations must be characterized by visual features to some degree; even though different object images can be matched for some visual features (e.g. spatial frequency), they will always differ on others (e.g. global form).

In summary, progress in understanding object recognition over the last three years has been characterized by the influence of DNNs, inspection of the time course of neural responses in addition to their spatial organization, and a broader conceptualization of what constitutes an object representation that includes the influence of context. A cohesive understanding of the neural basis of object recognition will also require integrating our knowledge of visual object processing with related processes such as object memory<sup>104</sup>, which are typically studied independently. Although DNNs have now reached human levels of performance<sup>28,29</sup> for object categorization under controlled conditions, humans perform this task daily under much more varied conditions and constraints. The continued evolution of the field in terms of sophisticated analytic tools, larger stimulus sets, and the consideration of the context in which object recognition occurs will provide further insight into the human brain’s remarkable flexibility.

## References

- Gauthier I, Tarr MJ: **Visual Object Recognition: Do We (Finally) Know More Now Than We Did?** *Annu Rev Vis Sci.* 2016; **2**: 377–96.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Warrington EK: **Neuropsychological studies of object recognition.** *Philos Trans R Soc Lond B Biol Sci.* 1982; **298**(1089): 15–33.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Humphreys GW, Forde EME: **Hierarchies, similarity, and interactivity in object**

**recognition: “Category-specific” neuropsychological deficits..** *Behav Brain Sci.* 2001; **24**(3): 453–76.  
[PubMed Abstract](#) | [Publisher Full Text](#)

- Biederman I: **Recognition-by-components: A theory of human image understanding.** *Psychol Rev.* 1987; **94**(2): 115–47.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Thorpe S, Fize D, Marlot C: **Speed of processing in the human visual system.**



- Nature*. 1996; **381**(6582): 520–2.  
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Ullman S, Soloviev S: **Computation of pattern invariance in brain-like structures**. *Neural Netw*. 1999; **12**(7–8): 1021–36.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  7. Wallis G, Rolls ET: **Invariant face and object recognition in the visual system**. *Prog Neurobiol*. 1997; **51**(2): 167–94.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  8. Logothetis NK, Pauls J, Poggio T: **Shape representation in the inferior temporal cortex of monkeys**. *Curr Biol*. 1995; **5**(5): 552–63.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  9. Edelman S: **Representation and Recognition in Vision**. MIT Press, 1999.  
[Reference Source](#)
  10. DiCarlo JJ, Cox DD: **Untangling invariant object recognition**. *Trends Cogn Sci*. 2007; **11**(8): 333–41.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  11. Ullman S: **High-level Vision**. MIT Press, 2000.  
[Reference Source](#)
  12. **F** Ward EJ, Isik L, Chun MM: **General Transformations of Object Representations in Human Visual Cortex**. *J Neurosci*. 2018; **38**(40): 8526–37.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
  13. Kravitz DJ, Vinson LD, Baker CI: **How position dependent is visual object recognition?** *Trends Cogn Sci*. 2008; **12**(3): 114–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  14. Malach R, Reppas JB, Benson RR, *et al.*: **Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex**. *Proc Natl Acad Sci U S A*. 1995; **92**(18): 8135–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  15. Grill-Spector K, Kourtzi Z, Kanwisher N: **The lateral occipital complex and its role in object recognition**. *Vision Res*. 2001; **41**(10–11): 1409–22.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  16. **F** Haxby JV, Gobbini MI, Furey ML, *et al.*: **Distributed and overlapping representations of faces and objects in ventral temporal cortex**. *Science*. 2001; **293**(5539): 2425–30.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
  17. Kriegeskorte N, Mur M, Ruff DA, *et al.*: **Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey**. *Neuron*. 2008; **60**(6): 1126–41.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  18. Carlson T, Tovar DA, Alink A, *et al.*: **Representational dynamics of object vision: The first 1000 ms**. *J Vis*. 2013; **13**(10): 1.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  19. Cichy RM, Pantazis D, Oliva A: **Resolving human object recognition in space and time**. *Nat Neurosci*. 2014; **17**(3): 455–62.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  20. Kravitz DJ, Saleem KS, Baker CI, *et al.*: **The ventral visual pathway: An expanded neural framework for the processing of object quality**. *Trends Cogn Sci*. 2013; **17**(1): 26–49.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  21. DiCarlo JJ, Zoccolan D, Rust NC: **How Does the Brain Solve Visual Object Recognition?** *Neuron*. 2012; **73**(3): 415–34.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  22. Kanwisher N, McDermott J, Chun MM: **The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception**. *J Neurosci*. 1997; **17**(11): 4302–11.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  23. Grill-Spector K, Weiner KS, Kay K, *et al.*: **The Functional Neuroanatomy of Human Face Perception**. *Annu Rev Vis Sci*. 2017; **3**: 167–96.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  24. Serre T: **Deep Learning: The Good, the Bad and the Ugly**. *Annu Rev Vis Sci*. 2019; **5**: 399–426.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  25. Kietzmann TC, McClure P, Kriegeskorte N: **Deep neural networks in computational neuroscience**. *Oxford Research Encyclopedia of Neuroscience*. 2019; **10**: 115.  
[Publisher Full Text](#)
  26. **F** Kriegeskorte N: **Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing**. *Annu Rev Vis Sci*. 2015; **1**: 417–46.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
  27. Kriegeskorte N, Golan T: **Neural network models and deep learning**. *Curr Biol*. 2019; **29**(7): R231–R236.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  28. Krizhevsky A, Sutskever I, Hinton GE: **Imagenet classification with deep convolutional neural networks**. *NIPS*. 2012.  
[Publisher Full Text](#)
  29. He K, Zhang X, Ren S, *et al.*: **Deep residual learning for image recognition**. in 2016-December, IEEE, 2016; 770–778.  
[Publisher Full Text](#)
  30. Schrimpf M, Kubilius J, Hong H, *et al.*: **Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?** *bioRxiv*. 2018; 407007.  
[Publisher Full Text](#)
  31. Kubilius J, Bracci S, Op de Beeck HP: **Deep Neural Networks as a Computational Model for Human Shape Sensitivity**. *PLoS Comput Biol*. 2016; **12**(4): e1004896.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  32. Jozwik KM, Kriegeskorte N, Storrs KR, *et al.*: **Deep Convolutional Neural Networks Outperform Feature-Based But Not Categorical Models in Explaining Object Similarity Judgments**. *Front Psychol*. 2017; **8**: 1089.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  33. Pramod RT, Arun SP: **Do Computational Models Differ Systematically From Human Object Perception?** 2016; 1601–1609.  
[Publisher Full Text](#)
  34. **F** Rajalingham R, Issa EB, Bashivan P, *et al.*: **Large-Scale, High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans, Monkeys, and State-of-the-Art Deep Artificial Neural Networks**. *J Neurosci*. 2018; **38**(33): 7255–69.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
  35. Geirhos R, Rubisch P, Michaelis C, *et al.*: **Imagenet-Trained Cnns Are Biased Towards Texture; Increasing Shape Bias Improves Accuracy And Robustness**. [online]. 2019.  
[Reference Source](#)
  36. Geirhos R, *et al.*: **Generalisation in humans and deep neural networks**. *Advances in Neural Information Processing Systems*. 2018-December, 2018; 7538–7550.  
[Reference Source](#)
  37. Bankson BB, Hebart MN, Groen IIA, *et al.*: **The temporal evolution of conceptual object representations revealed through models of behavior, semantics and deep neural networks**. *Neuroimage*. 2018; **178**: 172–82.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  38. **F** Bracci S, Ritchie JB, Kalfas I, *et al.*: **The Ventral Visual Pathway Represents Animal Appearance over Animacy, Unlike Human Behavior and Deep Neural Networks**. *J Neurosci*. 2019; **39**(33): 6513–25.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
  39. King ML, Groen IIA, Steel A, *et al.*: **Similarity judgments and cortical visual responses reflect different properties of object and scene categories in naturalistic images**. *Neuroimage*. 2019; **197**: 368–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  40. **F** Seeliger K, Fritsche M, Güçlü U, *et al.*: **Convolutional neural network-based encoding and decoding of visual object recognition in space and time**. *Neuroimage*. 2018; **180**(Pt A): 253–66.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
  41. **F** Cichy RM, Kriegeskorte N, Jozwik KM, *et al.*: **The spatiotemporal neural dynamics underlying perceived similarity for real-world objects**. *Neuroimage*. 2019; **194**: 12–24.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
  42. Horikawa T, Kamitani Y: **Generic decoding of seen and imagined objects using hierarchical visual features**. *Nat Commun*. 2017; **8**: 15037.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  43. Rajaei K, Mohsenzadeh Y, Ebrahimpour R, *et al.*: **Beyond core object recognition: Recurrent processes account for object recognition under occlusion**. *PLoS Comput Biol*. 2019; **15**(5): e1007001.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  44. Groen II, Greene MR, Baldassano C, *et al.*: **Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior**. *eLife*. 2018; **7**: e32962.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  45. Zeman AA, Ritchie JB, Bracci S, *et al.*: **Orthogonal Representations of Object Shape and Category in Deep Convolutional Neural Networks and Human Visual Cortex**. *Sci Rep*. 2020; **10**(1): 2453.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  46. Khaligh-Razavi SM, Kriegeskorte N: **Deep supervised, but not unsupervised, models may explain IT cortical representation**. *PLoS Comput Biol*. 2014; **10**(11): e1003915.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  47. Cichy RM, Khosla A, Pantazis D, *et al.*: **Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence**. *Sci Rep*. 2016; **6**: 27755.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
  48. Güçlü U, van Gerven MAJ: **Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream**. *J Neurosci*. 2015; **35**(27): 10005–14.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  49. Mozafari M, Ganjtabesh M, Nowzari-Dalini A, *et al.*: **Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks**. *Pattern Recognit*. 2019; **94**: 87–95.  
[Publisher Full Text](#)
  50. **F** Kheradpisheh SR, Ganjtabesh M, Thorpe SJ, *et al.*: **STDP-based spiking deep convolutional neural networks for object recognition**. *Neural Netw*. 2018; **99**:



- 56–67.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
51. **F** Kar K, Kubilius J, Schmidt K, *et al.*: Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat Neurosci.* 2019; **22**(6): 974–83.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
52. Kietzmann TC, Spoerer CJ, Sörensen LKA, *et al.*: Recurrence is required to capture the representational dynamics of the human visual system. *Proc Natl Acad Sci U S A.* 2019; **116**(43): 21854–63.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
53. Contini EW, Wardle SG, Carlson TA: Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia.* 2017; **105**: 165–76.  
[PubMed Abstract](#) | [Publisher Full Text](#)
54. Martin A: GRAPES—Grounding representations in action, perception, and emotion systems: How object properties and categories are represented in the human brain. *Psychon Bull Rev.* 2016; **23**(4): 979–90.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
55. Kaiser D, Moeskops MM, Cichy RM: Typical retinotopic locations impact the time course of object coding. *Neuroimage.* 2018; **176**: 372–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Brandman T, Peelen MV: Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J Neurosci.* 2017; **37**(32): 7700–10.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Khaligh-Razavi SM, Cichy RM, Pantazis D, *et al.*: Tracking the Spatiotemporal Neural Dynamics of Real-world Object Size and Animacy in the Human Brain. *J Cogn Neurosci.* 2018; **30**(11): 1559–76.  
[PubMed Abstract](#) | [Publisher Full Text](#)
58. Pennington J, Socher R, Manning CD: GloVe: Global vectors for word representation. in 2014; 1532–1543.  
[Publisher Full Text](#)
59. Riesenhuber M, Poggio T: Hierarchical models of object recognition in cortex. *Nat Neurosci.* 1999; **2**(11): 1019–25.  
[PubMed Abstract](#) | [Publisher Full Text](#)
60. Clarke A, Devereux BJ, Randall B, *et al.*: Predicting the Time Course of Individual Objects with MEG. *Cereb Cortex.* 2015; **25**(10): 3602–12.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
61. Bruffaerts R, Tyler LK, Shafto M, *et al.*: Perceptual and conceptual processing of visual objects across the adult lifespan. *Sci Rep.* 2019; **9**(1): 13771.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
62. **F** Devereux BJ, Clarke A, Tyler LK: Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Sci Rep.* 2018; **8**(1): 10636.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
63. Chiou R, Lambon Ralph MA: The anterior temporal cortex is a primary semantic source of top-down influences on object recognition. *Cortex.* 2016; **79**: 75–86.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
64. Mohsenzadeh Y, Qin S, Cichy RM, *et al.*: Ultra-Rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *Elife.* 2018; **7**: e36329.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
65. Grootswagers T, Robinson AK, Carlson TA: The representational dynamics of visual objects in rapid serial visual processing streams. *Neuroimage.* 2019; **188**: 668–79.  
[PubMed Abstract](#) | [Publisher Full Text](#)
66. Mohsenzadeh Y, Mullin C, Lahner B, *et al.*: Reliability and Generalizability of Similarity-Based Fusion of MEG and fMRI Data in Human Ventral and Dorsal Visual Streams. *Vision.* 2019; **3**(1): 8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
67. Hebart MN, Bankson BB, Harel A, *et al.*: The representational dynamics of task and object processing in humans. *eLife.* 2018; **7**: e32816.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
68. Proklova D, Kaiser D, Peelen MV: Disentangling Representations of Object Shape and Object Category in Human Visual Cortex: The Animate-Inanimate Distinction. *J Cogn Neurosci.* 2016; **28**(5): 680–92.  
[PubMed Abstract](#) | [Publisher Full Text](#)
69. Proklova D, Kaiser D, Peelen MV: MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *Neuroimage.* 2019; **193**: 167–77.  
[PubMed Abstract](#) | [Publisher Full Text](#)
70. Iivanainen J, Stenroos M, Parkkonen L: Measuring MEG closer to the brain: Performance of on-scalp sensor arrays. *Neuroimage.* 2017; **147**: 542–53.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
71. Boto E, Holmes N, Leggett J, *et al.*: Moving magnetoencephalography towards real-world applications with a wearable system. *Nature.* 2018; **555**(7698): 657–61.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
72. Logothetis NK, Pauls J, Bülthoff HH, *et al.*: View-dependent object recognition by monkeys. *Curr Biol.* 1994; **4**(5): 401–14.  
[PubMed Abstract](#) | [Publisher Full Text](#)
73. Kourtzi Z, Kanwisher N: Cortical Regions Involved in Perceiving Object Shape. *J Neurosci.* 2000; **20**(9): 3310–8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
74. Baldassano C, Beck DM, Fei-Fei L: Human-Object Interactions Are More than the Sum of Their Parts. *Cereb Cortex.* 2017; **27**(3): 2276–88.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
75. **F** Freud E, Macdonald SN, Chen J, *et al.*: Getting a grip on reality: Grasping movements directed to real objects and images rely on dissociable neural representations. *Cortex.* 2018; **98**: 34–48.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Faculty Opinions Recommendation](#)
76. Snow JC, Pettypiece CE, McAdam TD, *et al.*: Bringing the real world into the fMRI scanner: Repetition effects for pictures versus real objects. *Sci Rep.* 2011; **1**: 130.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
77. Harel A, Kravitz DJ, Baker CI: Task context impacts visual object processing differentially across the cortex. *Proc Natl Acad Sci U S A.* 2014; **111**(10): E962–E971.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
78. MacEvoy SP, Epstein RA: Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr Biol.* 2009; **19**(11): 943–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
79. Zoccolan D, Cox DD, DiCarlo JJ: Multiple Object Response Normalization in Monkey Inferotemporal Cortex. *J Neurosci.* 2005; **25**(36): 8150–64.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
80. Kaiser D, Peelen MV: Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *Neuroimage.* 2018; **169**: 334–41.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
81. Kaiser D, Quek GL, Cichy RM, *et al.*: Object Vision in a Structured World. *Trends Cogn Sci.* 2019; **23**(8): 672–85.  
[PubMed Abstract](#) | [Publisher Full Text](#)
82. Kaiser D, Cichy RM: Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J Neurophysiol.* 2018; **120**(2): 848–53.  
[PubMed Abstract](#) | [Publisher Full Text](#)
83. Simoncelli EP: Vision and the statistics of the visual environment. *Curr Opin Neurobiol.* 2003; **13**(2): 144–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
84. Zopf R, Butko M, Woolgar A, *et al.*: Representing the location of manipulable objects in shape-selective occipitotemporal cortex: Beyond retinotopic reference frames? *Cortex.* 2018; **106**: 132–50.  
[PubMed Abstract](#) | [Publisher Full Text](#)
85. Bainbridge WA, Oliva A: Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage.* 2015; **122**: 408–16.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
86. **F** Bracci S, Daniels N, Op de Beeck H: Task Context Overrides Object- and Category-Related Representational Content in the Human Parietal Cortex. *Cerebral Cortex.* 2017; **27**(1): 310–321.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#) | [Faculty Opinions Recommendation](#)
87. Vaziri-Pashkam M, Xu Y: Goal-Directed Visual Processing Differentially Impacts Human Ventral and Dorsal Visual Representations. *J Neurosci.* 2017; **37**(36): 8767–82.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
88. Xu Y, Vaziri-Pashkam M: Task modulation of the 2-pathway characterization of occipitotemporal and posterior parietal visual object representations. *Neuropsychologia.* 2019; **132**: 107140.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
89. Bugatus L, Weiner KS, Grill-Spector K: Task alters category representations in prefrontal but not high-level visual cortex. *Neuroimage.* 2017; **155**: 437–49.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
90. Erez Y, Duncan J: Discrimination of Visual Categories Based on Behavioral Relevance in Widespread Regions of Frontoparietal Cortex. *J Neurosci.* 2015; **35**(36): 12383–93.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
91. Vaziri-Pashkam M, Xu Y: An Information-Driven 2-Pathway Characterization of Occipitotemporal and Posterior Parietal Visual Object Representations. *Cereb Cortex.* 2019; **29**(5): 2034–50.  
[PubMed Abstract](#) | [Publisher Full Text](#)
92. Peelen MV, Downing PE: Category selectivity in human visual cortex: Beyond visual object recognition. *Neuropsychologia.* 2017; **105**: 177–83.  
[PubMed Abstract](#) | [Publisher Full Text](#)
93. Rosenthal I, Ratnasingham S, Haile T, *et al.*: Color statistics of objects, and color tuning of object cortex in macaque monkey. *J Vis.* 2018; **18**(11): 1.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
94. Teichmann L, Grootswagers T, Carlson TA, *et al.*: Seeing versus knowing: The temporal dynamics of real and implied colour processing in the human brain. *Neuroimage.* 2019; **200**: 373–81.  
[PubMed Abstract](#) | [Publisher Full Text](#)

95. Schmid AC, Doerschner K: **Representing stuff in the human brain.** *Curr Opin Behav Sci.* 2019; **30**: 178–85.  
[Publisher Full Text](#)
96. Kriegeskorte N, Kievit RA: **Representational geometry: Integrating cognition, computation, and the brain.** *Trends Cogn Sci.* 2013; **17**(8): 401–12.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
97. Chang N, Pyles JA, Marcus A, *et al.*: **BOLD5000, a public fMRI dataset while viewing 5000 visual images.** *Sci Data.* 2019; **6**(1): 49.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
98. Hebart MN, Dickter AH, Kidder A, *et al.*: **THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images.** *PLoS One.* 2019; **14**(10): e0223792.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
99. Coggan DD, Giannakopoulou A, Ali S, *et al.*: **A data-driven approach to stimulus selection reveals an image-based representation of objects in high-level visual areas.** *Hum Brain Mapp.* 2019; **40**(1): 4716–31.  
[Publisher Full Text](#)
100. Bracci S, Ritchie JB, de Breeck HO: **On the partnership between neural representations of object categories and visual features in the ventral visual pathway.** *Neuropsychologia.* 2017; **105**: 153–64.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
101. Bracci S, Op de Beeck H: **Dissociations and Associations between Shape and Category Representations in the Two Visual Pathways.** *J Neurosci.* 2016; **36**(2): 432–44.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
102. Long B, Yu CP, Konkle T: **Mid-level visual features underlie the high-level categorical organization of the ventral stream.** *Proc Natl Acad Sci U S A.* 2018; **115**(38): E9015–E9024.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
103. Wardle SG, Ritchie JB: **Can object category-selectivity in the ventral visual pathway be explained by sensitivity to low-level image properties?** *J Neurosci.* 2014; **34**(45): 14817–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
104. Brady TF, Konkle T, Alvarez GA, *et al.*: **Visual long-term memory has a massive storage capacity for object details.** *Proc Natl Acad Sci U S A.* 2008; **105**(38): 14325–9.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)



# Open Peer Review

Current Peer Review Status:  

---

## Editorial Note on the Review Process

**Faculty Reviews** are written by members of the prestigious **Faculty Opinions Faculty**. They are commissioned and are peer reviewed before publication to ensure that the final, published version is comprehensive and accessible. The reviewers who approved the final version are listed with their names and affiliations.

---

## The reviewers who approved this article are:

### Version 1

1 **Marius Peelen**

Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

**Competing Interests:** No competing interests were disclosed.

2 **Aude Oliva**

Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

**Competing Interests:** No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**