ANNUAL REVIEWS

*Annual Review of Vision Science*

# Visual Representations: Insights from Neural Decoding

Amanda K. Robinson,[1] Genevieve L. Quek,[2] and Thomas A. Carlson[3]

[1] Queensland Brain Institute, The University of Queensland, Brisbane, Australia; email: amanda.robinson@uqconnect.edu.au

[2] The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia; email: g.quek@westernsydney.edu.au

[3] School of Psychology, University of Sydney, Sydney, Australia; email: thomas.carlson@sydney.edu.au

Annu. Rev. Vis. Sci. 2023. 9:313–35

First published as a Review in Advance on March 8, 2023

The *Annual Review of Vision Science* is online at vision.annualreviews.org

https://doi.org/10.1146/annurev-vision-100120-025301

## Keywords

decoding, multivariate pattern analysis, object representations, neural dynamics, representational structure, internal representations

## Abstract

Patterns of brain activity contain meaningful information about the perceived world. Recent decades have welcomed a new era in neural analyses, with computational techniques from machine learning applied to neural data to decode information represented in the brain. In this article, we review how decoding approaches have advanced our understanding of visual representations and discuss efforts to characterize both the complexity and the behavioral relevance of these representations. We outline the current consensus regarding the spatiotemporal structure of visual representations and review recent findings that suggest that visual representations are at once robust to perturbations, yet sensitive to different mental states. Beyond representations of the physical world, recent decoding work has shone a light on how the brain instantiates internally generated states, for example, during imagery and prediction. Going forward, decoding has remarkable potential to assess the functional relevance of visual representations for human behavior, reveal how representations change across development and during aging, and uncover their presentation in various mental disorders.

# DECODING NEURAL RESPONSES

## What Are Neural Representations?

**Neural representations:** neural activity patterns that encode information about external input (e.g., visual features or objects) or internally generated states (e.g., preparatory templates)

**Multivariate pattern analysis (MVPA):** also known as decoding; method for quantifying the discriminability of neural activity patterns corresponding to different conditions, with separable patterns indicating condition-related representations

The study of neural representations is predicated on the concepts that distributed, population-level neural activity plays an important role in the functioning of the brain, and that neural activation patterns can be characterized by their relationship to both perception and mental states. In this way, neural representations are understood to be patterns of brain activity that encode information about both the internal and the external world—a neural code for the features of perceptual input, as well as internally generated cognitive states. An important aspect of representations is that they can be used by the brain to guide behavior; indeed, some have argued that behavioral relevance is a nonnegotiable criterion of using the term representation (Baker et al. 2022). In this review, however, we use the term in a broader sense to mean a neural pattern associated with a given state (see the section titled Linking Neural Representations to Behavior).

Over the past several decades, characteristic neural patterns of varying levels of abstraction have been reported in the literature, arising in response to high-level visual categories such as faces and objects (Haxby et al. 2001) and basic visual features such as orientation (Kamitani & Tong 2005) and color (Brouwer & Heeger 2009), as well as in the context of higher cognitive states such as preparatory attention (Gayet & Peelen 2022). These representations are at once both reliable and complex: Representational formats can vary according to the mental state or process under inspection, the neural imaging method used, the brain region considered, the amount of processing time devoted to a stimulus, or the task performed (Bracci & Op de Beeck 2022). Thus, neural representations are necessarily specific to modality, time, space, and context.

Although a host of methods exist to measure and interpret neural representations, including repetition suppression (Barron et al. 2016, Grill-Spector & Malach 2001) and adaptation (Webster 2011), multivariate pattern analysis (MVPA) methods that use decoding and encoding models to correlate neural patterns with conditions have enjoyed wide uptake in recent years (for a brief comparison, see the sidebar titled Encoding Models). In this review, we focus on the utility of multivariate decoding as a sensitive yet flexible method for contrasting different neural representations and characterize how neural decoding methods have advanced our knowledge of visual representations in the brain. In particular, we highlight the recent uptake of decoding methods for studying the temporal dynamics of visual processing and the organization or structure of information representation in the brain and point to the transition of neural decoding from a method used to index core representations to a technique now being used to index transient state spaces, giving insight into mechanisms of attention, prediction, and imagery.

## ENCODING MODELS

Taking multivariate analyses a step further, newer work has focused on predicting neural activity from stimuli using encoding models. Encoding and decoding are complementary techniques, but encoding explicitly models how information is represented in patterns of neural activity (Naselaris et al. 2011). An extension of encoding models, forward encoding, is used to reconstruct neural activity for stimuli never used to train the model (Brouwer & Heeger 2009). Such methods can be powerful for characterizing the neural coding of continuous stimulus attributes (e.g., orientation). For complex scenes and objects, techniques have to be applied to extract continuous or linear features from the images (Kay et al. 2008) or linearize the stimulus space (Gifford et al. 2022), which limits interpretability about the representations. We see enormous potential in encoding models, with their increased explanatory power, but think that decoding still has its place for now, particularly for classes of stimuli with no clear linear relationships that evoke abstract representations, such as objects.

## What Is Neural Decoding?

Whereas it was once standard practice to quantify conditional differences at the level of individual voxels or electrodes, the mass-univariate approach has now been firmly eclipsed by multivariate analyses of brain activity, which characterize neural representations by relating perceptual and cognitive conditions to their associated neural patterns of activity (Hebart & Baker 2018, Kragel et al. 2018). Broadly described, neural decoding methods aim to dissociate patterns of neural responses across different conditions or labels: When the patterns for two conditions are separable (e.g., for horizontally and vertically oriented gabors), the neural response is considered to contain information about the stimulus dimension under inspection (i.e., an orientation representation). The unifying element of multivariate decoding methods is their assessment of information jointly represented across individual voxels or sensors, achieved by considering the spatial or temporal distribution of activation from neuroimaging methods such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG) (**Figure 1**). The much-discussed advantage of this pattern-based approach over univariate methods is its increased sensitivity: Whereas the strength of activation on any individual neuron, sensor, or voxel may be indistinguishable for two conditions, the pattern reflected in jointly considered sensors may give near-perfect separability.

**Time-resolved decoding:** a variant that captures representational dynamics by inspecting the dissociability of neural activity patterns at sequential time points in MEG or EEG data

Broadly speaking, multivariate decoding methods aim to discriminate population-level neural activity between conditions or stimuli of interest, even if mean activity levels do not vary. In practice, this involves training machine learning algorithms (e.g., linear discriminant analysis, support vector machines) to associate patterns of neural activation across voxels or sensors with labels that capture the manipulation of interest (e.g., experimental conditions, stimulus classes). The classifier is then tested on held-out neural data that it has never encountered before, with above-chance classification accuracy taken as evidence that the neural patterns encode the dimension along which the stimuli or conditions vary (see the sidebar titled How High Is High Enough?). As the accessibility of these powerful methods has improved with the proliferation of open source toolboxes (Bode et al. 2019, Hebart et al. 2015, Oosterhof et al. 2016), multivariate decoding has begun to burgeon out of basic vision science and into other areas such as developmental research (Ashton et al. 2022). Rather than providing an in-depth tutorial on variations in the mechanistic implementation of multivariate decoding (extensively covered elsewhere; see Grootswagers et al. 2017, Pereira et al. 2009), we focus on how such methods have provided insight into visual perception at many different levels, from low-level visual features, to conceptual representations of objects, all the way through to internally generated representations evoked by imagery.

## Spatial and Temporal Dynamics

Within visual cognitive neuroscience, the multivariate framework has had a substantial impact on understanding both the spatial and temporal dynamics of neural representations. In the years since Haxby and colleagues' (2001) seminal paper, which used pattern classification to investigate face and object representations in the ventral temporal cortex, fMRI research has witnessed a shift away from asking which brain region is activated by a particular visual stimulus and toward questions about how a given stimulus is encoded at a population level. With the majority of fMRI studies of high-level vision published in the past decade now including MVPA of some type, the language of interpreting neuroimaging data has evolved such that the dominant parlance is now one of information and representation, rather than the historic notion of activation (Hebart & Baker 2018, Kragel et al. 2018, Kriegeskorte et al. 2006).

Although the origins of multivariate methods can be found in fMRI research, neural decoding methods have since taken root in the temporal domain, with a proliferation of studies that use time-resolved decoding (e.g., Carlson et al. 2011, Cichy et al. 2014, Dijkstra et al. 2018,

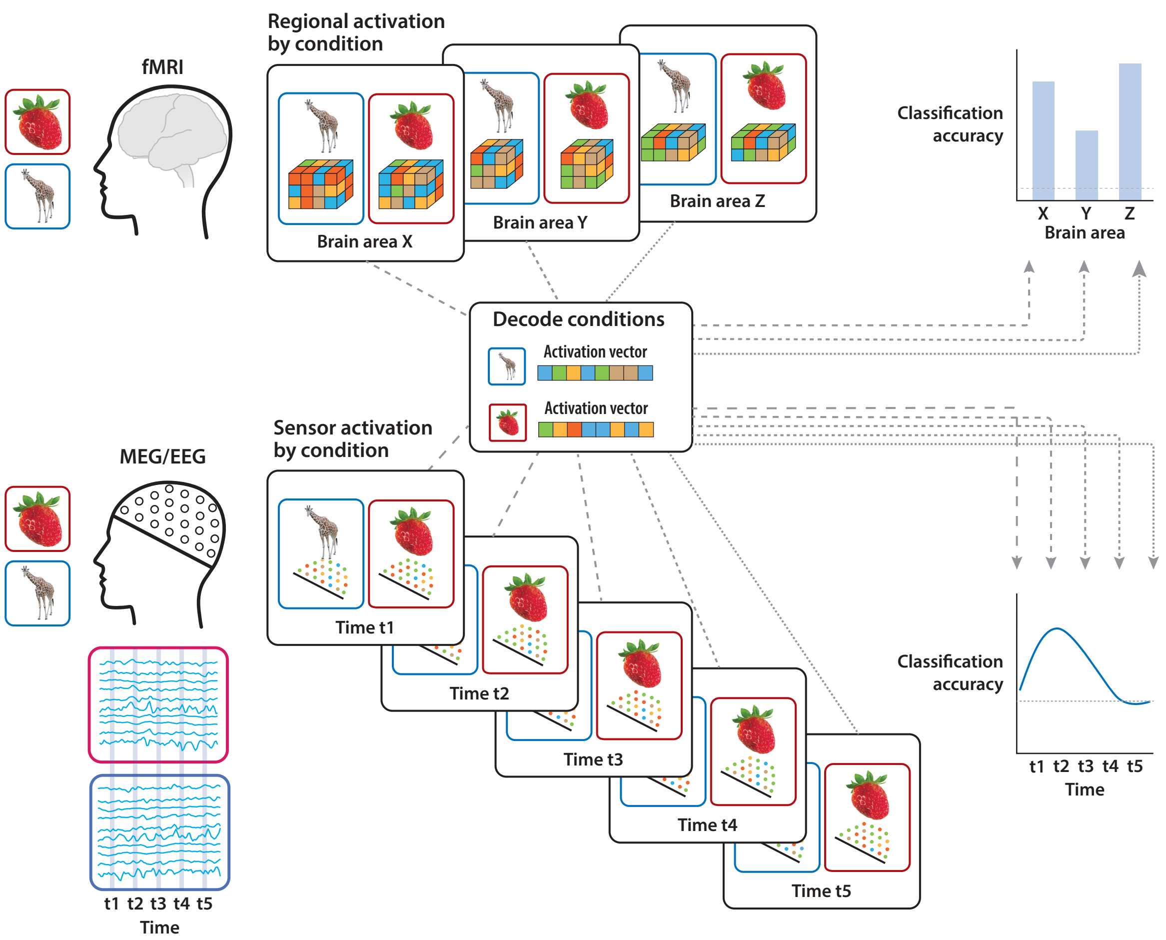

**Figure 1**

Multivariate decoding relies on discriminating patterns of activity evoked by one condition (e.g., an image of a strawberry) from those evoked by another condition (e.g., an image of a giraffe). The activation patterns for discrimination can take many forms, for example, activation across voxels (fMRI), channels/time (MEG, EEG), sensor positions (single unit recordings), or even behavioral metrics. Activation patterns are vectorized to form features for classification. A classifier is trained to discriminate the neural patterns across conditions using multiple trials per condition and tested on held-out data. If the classifier reliably performs above chance for this novel data, then information in the neural signal is understood to discriminate between the conditions, providing an index of the neural representation of the dimension along which the conditions vary. Abbreviations: EEG, electroencephalography; fMRI, functional magnetic resonance imaging; MEG, magnetoencephalography.

Hebart et al. 2018, Kaiser et al. 2016, Philiastides & Sajda 2006). In the context of continuous data (e.g., MEG or EEG), brain states corresponding to different conditions or stimuli are studied as dynamic, unfolding processes, with a classification algorithm implemented at each individual time point or time window to yield a decoding accuracy time series (Grootswagers et al. 2017). This is an important development of multivariate methods—equally as important as knowing where information is represented in the brain is understanding the time course over which the representations emerge and decay. Time-resolved decoding has provided unprecedented insights into the temporal dynamics of visual processing, for example, by revealing that representations associated with different levels of category abstraction are staged in time (Contini et al. 2017).

## HOW HIGH IS HIGH ENOUGH?

A common criticism leveled at decoding studies concerns low classification accuracy values. In time-resolved decoding, for instance, it is common to achieve reliable above-chance decoding with classification of approximately 60%, relative to the chance level of 50%. We and others argue that reliability is key to interpretation over value; decoding accuracy is not an effect size (Carlson et al. 2020, Hebart & Baker 2018). Classification accuracy is considerably influenced by data processing (filtering, artifact correction or rejection, trial averaging), the number of classification features (voxels, electrodes, time points), feature selection methods (e.g., choosing time points or regions guided by literature, applying principal component analysis, t-values), and classification methods (classifier choice, cross-validation schemes), as well as factors such as the underlying effect size and the number of trials per class (Grootswagers et al. 2017, Hebart & Baker 2018). Importantly, most of these factors do not seem to influence the statistical reliability of the effects, just the magnitude (Grootswagers et al. 2017). Of course, higher versus lower decoding magnitude can reflect the strength of the neural signal or noise, but reliable above-chance decoding is indicative of condition-relevant information contained in the neural signal, regardless of decoding magnitude.

Adding the temporal dimension has also given rise to more complex and informative variations of the method—for example, temporal generalization, in which a classifier trained on the spatial distribution of activation at one particular timepoint is tested on a range of different timepoints (King & Dehaene 2014, Stokes et al. 2013). This form of cross-decoding is relevant for revealing the temporal stability of neural activation patterns and speaks to the stationarity (or lack thereof) of various representational spaces (King & Dehaene 2014). However, neural decoding for continuous data suffers from the same challenge common to all time-series analyses—namely, multiple comparisons. How should decoding accuracy be evaluated against chance when there are 1,000+ accuracies to inspect? Current proponents of multivariate methods applied to MEG and EEG data favor Bayesian methods that quantify cumulative evidence, rather than frequentist corrections (Teichmann et al. 2021).

## MULTIFACETED VISUAL REPRESENTATIONS IN THE BRAIN

Multivariate decoding methods have been instrumental in moving the needle on understanding the functional architecture of the visual system. A great deal has been learned about the spatiotemporal characteristics of brain processes supporting the extraction of low-, mid-, and high-level properties of visual input. In this section, we describe insights from decoding studies that characterize visual perception as a set of increasingly complex representations that span from features through to concepts, ultimately focused on deriving meaning from visual stimuli.

### Visual Features

Representations of basic low-level units of visual information (e.g., orientation, position, spatial frequency) form the basis for all subsequent processing and are thus the foundation of visual perception. As a sensitive framework through which to investigate fine-tuned representations, decoding has confirmed and progressed our knowledge of how the visual system implements the first computations to result in perception. For example, fMRI decoding has shown that early visual cortex represents fundamental visual features such as orientation (Haynes & Rees 2005, Kamitani & Tong 2005), color (Brouwer & Heeger 2009), and motion direction (Kamitani & Tong 2006), as well as conjunctions of color and motion direction (Seymour et al. 2009). Different feature representations have distinct but overlapping temporal profiles. The earliest

**Temporal generalization:** assesses the stability of neural representations over time by quantifying how well models trained at one time point generalize to others

**Cross-decoding:** a variant of MVPA that examines the generalizability of representations between stimuli or states by training and testing classification models on different conditions

EEG and MEG responses, beginning at 50–70 ms after stimulus onset, reflect representations of stimulus position (Blom et al. 2020, Carlson et al. 2011, Robinson et al. 2021), spatial frequency (Ramkumar et al. 2013, Robinson et al. 2017), orientation (Cichy et al. 2015, Moerel et al. 2022b), and color (Rosenthal et al. 2021, Teichmann et al. 2020). While it is difficult to directly contrast time course information across studies employing different methods, comparisons of different features within the same study can shine a light on the differential dynamics of feature processing. For example, one study showed that shape decoding (evident from 60 ms onward) preceded color decoding (evident from 70 ms), whereas shape–color congruency was not evident until as late as 200 ms after stimulus onset (Teichmann et al. 2020). These results highlight the distinction between low-level feature coding and higher-level knowledge-based feature interactions.

## Objects and Categories

A major contribution of decoding methods applied to neuroimaging data has been to unlock the neural code within brain regions already identified via univariate methods as specialized for processing different object categories. In the field of face perception, for example, univariate fMRI studies had already revealed a network of regions in occipitotemporal cortex with face-selective properties (i.e., stronger responses to face versus nonface stimuli) (Kanwisher et al. 1997, McCarthy et al. 1997, Puce et al. 1996). Other studies established further organizational principles in visual cortex, showing modular regions that respond more strongly to specific categories of visual stimuli, such as bodies (Downing et al. 2001) and scenes (Epstein & Kanwisher 1998); retinotopic organization within object-selective areas (Levy et al. 2001); and a coarse scale organization based on object size across ventral temporal cortex (Konkle & Oliva 2012). Multivariate approaches to fMRI data have drilled further into these striking findings, revealing overlapping representations of various high-level categories such that the very same face-selective regions noted above have been shown to also contain relevant information about other object categories (Haxby et al. 2001). In this way, neural decoding studies have led to the development of a more tempered or graded set of linguistic terms than the older notions of selectivity or domain specificity encouraged.

In the temporal domain, decoding techniques applied to MEG and EEG data have characterized the emergence of increasingly abstract categorical representations as processing proceeds from early visual cortex to high-level ventral visual areas (Cichy et al. 2014, 2016). Whereas image-related differences in the low-level features of different objects are decodable as soon as 60 ms after stimulus onset (Carlson et al. 2011), higher-level information about basic categories (e.g., dog, boat) and overarching categories (animate or inanimate) arises comparatively later in time (Carlson et al. 2013, Cichy et al. 2014, Grootswagers et al. 2019a), more than 100 ms after image presentation. Notably, representations of different category levels do not arise and decay in nonoverlapping windows, but instead appear to cascade through time, highlighting the multifaceted nature of object recognition processes (Grootswagers et al. 2019a).

## Features Versus Objects: An Untenable Distinction

The overlapping temporal profiles of different category labels characterize the complexity of the brain processes involved in extracting high-level meaning from visual images. Indeed, since all objects are a conjunction of features, and since object categories comprise exemplars with similar featural profiles (e.g., bananas tend to be curved, leaves tend to be green), it is necessarily the case that object categories differ from one another on low- and mid-level features, as well as in terms of high-level abstract meaning. As can be seen in **Figure 2**, some of the categories of a large

**Figure 2**

Images from a large stimulus set show consistent visual features for some categories. For example, the category mean of human face is very representative of a human face, and the mean animate images have face-like attributes. Figure reproduced from Grootswagers & Robinson (2021) (CC BY 4.0).

and prolific stimulus set (Kriegeskorte et al. 2008b) can be distinguished by consistent features even when the images within a category are averaged (Grootswagers & Robinson 2021). The inherent contribution of visual features to category representations is exemplified in the computer vision literature, where deep neural networks can reliably label objects based purely on featural covariance reflected in the training image set (e.g., He et al. 2015).

That high-level information about semantic categories is necessarily confounded with perceptual or featural differences gives rise to interpretational challenges in decoding studies. Such is the sensitivity of these methods that even small differences between stimuli or conditions can drive classifier performance (see the sidebar titled Decoding Traps for New Players). Does above-chance decoding reflect true appreciation of the high-level distinction between categories, or does

## DECODING TRAPS FOR NEW PLAYERS

Decoding is a very sensitive measure for distinguishing between neural responses associated with different conditions, but it also has its downsides. Decoding can be useful in distinguishing subtle differences in object-related representations (e.g., cat versus dog) and the relationships among many different objects. For this reason, decoding results have enhanced knowledge about how representations are structured in the brain. Yet decoding can be considered a kind of black box in terms of interpretation (Carlson et al. 2018). A decodable contrast means that there is a difference in the patterns of activity across conditions, but it does not tell you what drives the difference. Moreover, machine learning approaches can pick up on any information that distinguishes between classes, so it is extremely important to remove potential confounds. For example, eye movements can contribute to the decodability of stimulus position from MEG signals (Quax et al. 2019). As is true of any type of analysis, very careful experimental design and interpretation are essential in decoding analyses.

it merely arise due to the brain's sensitivity to low-level featural differences between the categories under inspection, such as image contrast or luminance (Harrison 2022)? Elucidating the relationships among image features, category-level representations, and conceptual meaning has thus been a major challenge in object decoding studies over the past decade, with several tactics proposed to tackle the issue.

One approach to disentangling the contribution of image features and semantic meaning to the neural response to objects relies on eliminating the low-level featural differences between the to-be-decoded categories to the greatest extent possible (Bracci et al. 2017). For example, studies have carefully matched the perceptual similarity of stimuli belonging to different semantic categories by orthogonalizing object shape and category (e.g., a human hand and a glove have a near-identical shape). Using this approach, Bracci & Op de Beeck (2016) found a dissociation between shape and category within the ventral and dorsal visual streams, with a progression from shape to category information along the visual hierarchy. A further study found that ventral occipitotemporal cortex responses reflected object appearance rather than category (Bracci et al. 2019). Cross-decoding methods have extended this approach even further, showing that where intermediate visual representations emphasize shape information (independent of category), later representations encode information about object category that generalizes across different object shapes (Kaiser et al. 2016). In the field of face processing, mechanisms underlying face detection have been elucidated by studying face pareidolia, the perception of illusory faces in objects (Taubert et al. 2020). Illusory face stimuli have features more similar to objects yet show striking face-like neural responses in the brain at early stages of processing, indicating the existence of a rapid yet imprecise mechanism for face detection (Wardle et al. 2020). Together, such work shows that categorical representations are partially, but not entirely, influenced by underlying features common to the category.

Elsewhere, others have tried to dissociate perceptual and conceptual aspects of object category by examining the neural response to stimuli that preserve object features while disrupting recognizability (i.e., high-level category information). So-called texform stimuli achieve this goal by obscuring the basic category labels of objects while maintaining the mid-level visual features (e.g., shape or curvature). Studies using these control stimuli, first introduced by Long et al. (2017), have shown that key organizing principles in object-selective cortex (e.g., animacy and real world size) are at least partly accounted for by mid-level featural differences between objects along dimensions such as curvature. Similar effects have been reported in the time domain, where texforms evoke neural responses containing information about both animacy and real-world size during the same time windows as real objects do (Wang et al. 2022), although it appears that these representations are more susceptible to masking effects than real objects are (Grootswagers et al. 2019b).

A final approach to tackling low-level confounds between object categories relies on increasing featural variation both within and across categories. Image sets such as THINGS (Hebart et al. 2019) and ECOSET (Mehrer et al. 2021) contain thousands of highly variable natural images. Increasing the variation within each category serves to dampen the low-level confounds that might otherwise dominate measurements of neural representations (Grootswagers & Robinson 2021). Studying object-specific neural responses with such large stimulus sets has the advantage of more closely mimicking natural vision, as well as allowing more fine-grained analyses of visual features, categories, and semantics (Chang et al. 2019, Grootswagers et al. 2022, Hebart et al. 2023). For example, models of image statistics and object category can be compared with the neural data to assess how much each model accounts for the variance in neural information (Grootswagers et al. 2019a, Moerel et al. 2022a). Work in this space is undoubtedly the future in characterizing the relationship among image features, categories, conceptual representations, and ultimately perceptual experience.

# REPRESENTATIONAL GEOMETRY

## From “What” to “How”

Beyond straightforward demonstrations of distinguishable brain states, multivariate methods are now also in wide use as a means for characterizing the organizational structure of visual representations—the representational geometry of vision. This endeavor has deep roots; indeed, understanding the format of how the brain encodes information has been a core challenge in cognitive neuroscience for decades (see the sidebar titled Representational Geometry: New Solutions to Old Ideas). In this case, the underlying premise is that characterizing the relationships between different perceptual and conceptual conditions serves to elucidate how the brain represents information along different dimensions—that is, the structure of informational representation in a brain area or brain state. When considered in terms of their relationships with one another, neural representations give rise to a mapping or geometry in which certain types of perceptual or cognitive information are made explicit, and other types of information are abstracted away.

For example, a hypothetical fMRI experiment concerned with how shape and color are encoded in brain areas X, Y, and Z (**Figure 3*a***) might find that neural responses to stimuli group by shape in brain area X, group by color in area Y, and have no observable organization in area Z, even though all three areas have equivalent univariate activation. Thus, brain areas X and Y explicitly code for some stimulus properties (shape and color, respectively) while abstracting away from others (color and shape, respectively). From this, we can deduce that area X represents shape, area Y represents color, and area Z is not critical for either. A key idea in this case is that the format of representation should enable information to be “read out” by a biologically plausible process (for further discussion, see the section titled Linking Neural Representations to Behavior). Extending the example above, we can add a decision boundary (**Figure 3*a***) that divides the representation in area X into wider and taller shapes and the representation in area Y into warmer and cooler colors. Understanding how information is structured in a representation thus informs both what kind of information is being represented and how this information might be accessed or read out (Ritchie & Carlson 2016).

In reality, of course, it is unlikely that the brain utilizes such a low-dimensional format as in this example. That is, individual brain regions, brain states, or behavioral metrics likely describe more than one single dimension concurrently (Kriegeskorte & Kievit 2013). **Figure 3*b*** characterizes this complexity clearly; in the figure, observer ratings of the emotional content of images

### REPRESENTATIONAL GEOMETRY: NEW SOLUTIONS TO OLD IDEAS

The question of how objects are represented in the brain has been pondered since at least the 1970s. Shepard & Chipman (1970) showed that the shape similarity judgments from active viewing closely resembled those from visual imagery. This seminal paper suggested that both judgments were guided by the same internal representation of shape. Interestingly, one of the earliest MVPA studies was similarly motivated. Edelman (1998) found a correspondence between the brain’s internal representation of shape in ventral temporal cortex and behavioral judgments of shape. While a handful of later studies took an interest in this approach (Hanson et al. 2004, O’Toole et al. 2005), most MVPA work in the 2000s focused on what could be decoded from brain recordings, as opposed to how MVPA could be used to measure representational structure. Prominent publications shifted the focus from what could be decoded to studying how information is structured in brain representations (Kiani et al. 2007, Kriegeskorte et al. 2008b). This work developed an accessible framework (Kriegeskorte et al. 2008a) and ignited a body of research studying representational structure in the brain.

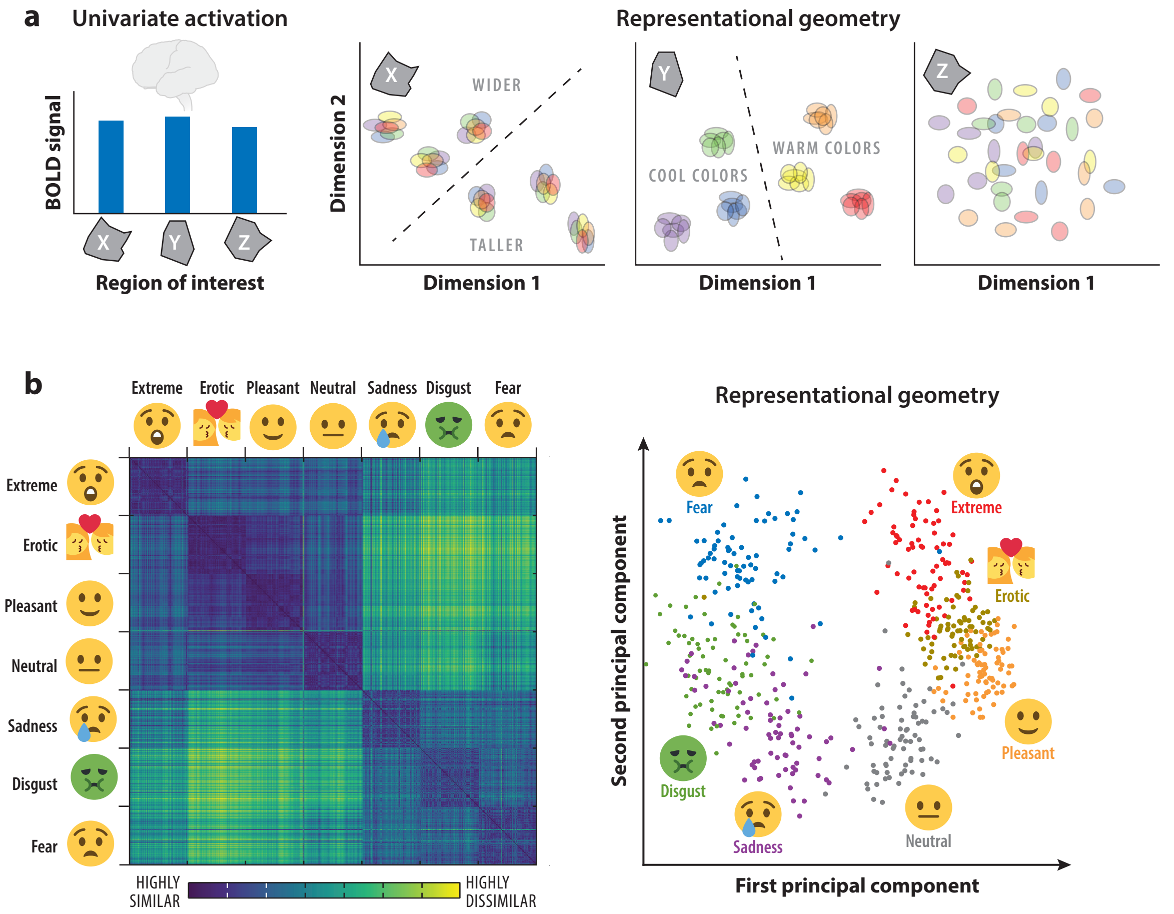

**Figure 3**

The format of representations in the brain can elucidate neural coding principles. (*a*) In a hypothetical example, we can imagine neural responses to visual stimuli that vary across two feature dimensions (e.g., color and shape) in three different brain regions X, Y, and Z. All three brain regions respond to the stimuli but differentially code the feature dimensions. Region X codes for shape; region Y codes for color; and region Z codes for something else about the stimuli, neither color nor shape. (*b*) Representational structure for emotion-related images as measured by behavior. The representational dissimilarity matrix (*left*) shows clustering of different aspects related to emotion. The representational space (*right*) shows that the representations are formatted along dimensions that roughly correspond to emotional valence and the strength of emotion. Abbreviation: BOLD, blood-oxygen-level-dependent.

(Grootswagers et al. 2020) capture information about various aspects important for emotion categorization in the brain—such that a representational mapping built from these ratings clearly reflects two dimensions of organization roughly corresponding to the valence and strength of the emotional content depicted. This shift toward examining the fine-grained structure of neural representations has shone a light on key organizational principles of object vision. Representational similarity analysis (RSA) (Kriegeskorte et al. 2008a) has emerged as the method of choice for quantitatively studying representational geometry by indexing the degree to which pairwise combinations of neural patterns for a set of stimuli or conditions resemble one another (i.e., their similarity) (for comprehensive reviews of RSA, see Kriegeskorte & Kievit 2013, Kriegeskorte et al. 2008a). Similarity between conditions or stimuli is readily quantified through a variety of methods (e.g., correlation, univariate differences); however, the most common method is to use decoding

**Representational similarity analysis:** a computational approach to modeling the structure of (neural) representations by quantifying item similarity to index (dis)similarity, with higher decoding accuracy indexing higher dissimilarity. The resulting representational similarity structure elicited in a given brain region or time window can then be compared to different theoretical models of stimulus relationships (e.g., based on similarity of stimulus features or category), effectively disentangling the distinctions and relationships between different image dimensions.

## Beyond Animacy: Curvature, Agency, and Capacity for Movement

One domain in which the representational structure framework has made a significant impact concerns the high-level categorical distinction between animate and inanimate objects. With animacy already long understood to be an important organizing principle for the visual system, RSA methods have refined our understanding of how objects within these overarching categories are represented within the visual system (Kriegeskorte et al. 2008b). For example, a large body of work has used RSA to quantify the degree to which mid-level featural differences between animate and inanimate objects (e.g., in curvature) account for the discriminability of the neural responses to these categories (Grootswagers et al. 2019b, Long et al. 2018, Wang et al. 2022). At the same time, however, the representational framework has also provided insights into how covarying conceptual attributes of objects contribute to the animate–inanimate dichotomy. For example, distinct face and body representations appear to be major contributors to the animacy division (Grill-Spector & Weiner 2014, Ritchie et al. 2021). Some have suggested that animacy can be better described as a continuum (Sha et al. 2015) that incorporates other organizing principles such as capacity for agency (Thorat et al. 2019) and humanness (Contini et al. 2021, Ritchie et al. 2021). More generally, RSA methods have also shown that movement-related representations account for variance in object representations over and above the variance accounted for by aliveness (Shatek et al. 2022). Perhaps owing to their complex and multifaceted nature, animacy representations appear to be widely distributed throughout visual cortex, with MEG–fMRI fusion showing that information about objects' animacy and real-world size is evident from 150 ms after image presentation and associated with representations in parahippocampal cortex (Khaligh-Razavi et al. 2018). Together, this work highlights how investigating fine-grained relationships between stimuli has shone new light on some of the most important organizing principles of the visual system.

## Representational Fusion

One important advantage of the RSA framework is that the representational geometries it gives rise to are abstracted away from specific measurement units (e.g., blood-oxygen-level-dependent activation, amplitude) and thus directly comparable across neuroimaging modalities, behavioral metrics, brain regions, samples, people, and even species (Kriegeskorte & Kievit 2013, Mur et al. 2013). Among the most exciting extensions in the neurocognitive decoding toolkit, this type of representational fusion has enabled increasingly complex overlay between modalities in vision science, marrying object representations between fMRI and MEG or EEG data to recover the spatiotemporal extent of visual representations (Cichy et al. 2014, 2016). This approach has also advanced our understanding of more abstract, conceptual representations of visual categories, for example, showing representational alignment between the organization of object representations in inferior temporal cortex and word usage patterns in human speech evaluated by natural language processing models (Carlson et al. 2014). More recently, an elegant study incorporating encoding models (see the sidebar titled Encoding Models) showed that visual and linguistic semantic representations align with one another at the border of the human visual cortex (Popham et al. 2021). These studies highlight the utility of studying fine-grained representational structure to tap into the richness of conceptual representations within the brain and the capacity of multivariate methods to uncover so-far unobserved organizational principles within the visual system.

## LINKING NEURAL REPRESENTATIONS TO BEHAVIOR

Decades of neuroimaging research has pursued the neural correlates of human behaviors, yet determining whether patterns of activity have a causal role in behavior has been a significant challenge. Decoding approaches have proven to be a fruitful framework for linking neural representations to behavior in many ways. Exciting techniques have been used to assess whether neural responses are necessary for behavior, for example, by comparing classifier accuracy with behavioral accuracy (Moshel et al. 2022), correlating trial-by-trial reaction times with classifier confidence or distance to the boundary (Ritchie & Carlson 2016, Ritchie et al. 2015), and analyses of neural representations during behavioral errors (Robinson et al. 2022, Williams et al. 2007). RSA, in particular, provides a useful way to test the supposition that conditions that evoke more similar representations in one domain (e.g., responses in lateral occipital cortex) should also be more similar in other domains (e.g., reaction time on object classification tasks). Comparisons of behaviorally derived models and neural data serve to identify candidate neural substrates underlying certain behaviors. For example, Wardle et al. (2016) showed that intermediate visual responses (approximately 100–300 ms post stimulus onset) could almost entirely be explained by human judgments of perceptual similarity, a result that aligns with the idea that intermediate visual representations encode the perceptual form of the stimulus.

At the same time, another important takeaway from studies using a decoding framework to link brain and behavior is that not all decoded information is relevant for behavior (Grootswagers et al. 2018). Indeed, it seems likely that the strict correspondence between neural and measures of perceptual dissimilarity should depend on the relevant processing stage required for the classification task in question. Searchlight fMRI has shown that animacy representations are evident throughout the entire ventral stream but that correlations between the distance to classifier boundary and reaction time are only evident for higher areas along the ventral stream (Grootswagers et al. 2018). Other work has failed to find any correlation between neural representations and behavior. This raises an interesting question: If there is information in the brain, why can we not access it? In one EEG study, for instance, Moshel et al. (2022) found that neural responses could distinguish between very realistic computer-generated fake faces and real faces approximately 170 ms after the faces were presented, but behaviorally, participants could not distinguish which faces were real. Perhaps in such situations, more sustained processing is required before representations become relevant for behavior.

## ROBUST BUT PERMEABLE VISUAL REPRESENTATIONS

The increasing focus on representation over activation has brought with it questions about the stability or reliability of visual representations. On the one hand, we know that meaningful visual representations are reliably elicited despite varying circumstances—for example, in response to different exemplars of the same stimulus class, in the context of highly degraded stimuli, or in the presence of an orthogonal task. On the other hand, however, there can be no doubt that higher-order cognitive states are capable of modulating, and even generating, decodable representations in the brain—as is the case in visual imagery. In this section, we review findings that contrast the robust nature of visual representations with their capacity for modulation by top-down factors such as attention, expectation, and prediction.

### Visual Representations Are Remarkably Robust

Decoding studies have provided striking insight into the robust nature of visual representations. For example, work in the temporal domain has shown that relevant information about stimulus features and categories persists in the brain well after the stimulus has disappeared (e.g., approximately 500 ms; Contini et al. 2017) and is evident even when visual processing is disrupted via masking (Grootswagers et al. 2019a, King & Wyart 2021, Robinson et al. 2019). Results from these rapid image sequence designs indicate that multiple successive visual events can be decoded in the same time window, suggesting that multiple representations can be online concurrently. This is consistent with findings that visual representations of objects remain decodable even when visual competition is high, for example, when stimuli are presented simultaneously (Grootswagers et al. 2021, Macevoy & Epstein 2009); when the stimulus presentation rate is very high, e.g., 60 Hz (Grootswagers et al. 2019b); or when attention is diverted away from the stimulus (Grootswagers et al. 2021, Macevoy & Epstein 2009, Moerel et al. 2022b).

Relatedly, featural and categorical representations elicited by object or scene images appear to be largely robust to the participant's task (Grootswagers et al. 2019a, Harel et al. 2014, Hebart et al. 2018, Shatek et al. 2022) such that most decoding work that examines object representations does so only in the context of orthogonal task designs (Cichy et al. 2014, Kriegeskorte et al. 2008b). That said, the task does exert an influence on representations, particularly in higher visual areas and later time periods (Harel et al. 2014, Hebart et al. 2018), suggesting that current goals have a role to play during recurrent or feedback-related processing. Together, the findings that object representations are reliably evident despite compromised viewing conditions and under a variety of task circumstances suggest that the feedforward sweep of information through the visual system is quite robust (DiCarlo & Cox 2007).

## Modulating Visual Representations

While representations of images and categories are undeniably robust, they are nonetheless subject to modulation by a host of top-down factors. Traditional analysis methods have focused on detailing the neural correlates of different cognitive states, revealing the extent to which neural activity is enhanced or reduced under different conditions (e.g., attended, unattended). In contrast, decoding (and encoding; see the sidebar titled Encoding Models) assesses how the representational format of information encoded in the brain changes under different mental states—reflected as subtle changes in the discriminability of conditions or stimuli. In doing so, decoding can offer complementary insights into top-down influences on vision by elucidating the underlying mechanisms of these effects. In this section, we review some of the most intriguing multivariate findings in the areas of visual selective attention, imagery, and prediction and discuss how these methods are shaping our understanding of top-down influences on vision.

**Selective attention.** In the domain of selective attention—the process by which a stimulus is selected for subsequent processing among competing distractors—multivariate decoding has been instrumental in revealing how attending to a visual stimulus influences the quality of its representation, extending and complementing the many decades of research showing that attention enhances neural responses at early stages of processing in the extrastriate cortex (Desimone & Duncan 1995, Kastner et al. 1998, Mangun 1995). Time-resolved decoding studies using object stimuli in rapid presentation designs have provided evidence for multiple attentional stages. First, core visual representations are readily observed in response to visual stimuli from approximately 70 ms onward, arising regardless of task relevance and when presented for very brief durations (Grootswagers et al. 2019a, King & Wyart 2021, Marti & Dehaene 2017, Mohsenzadeh et al. 2018, Robinson et al. 2019). Such core representations seem to reflect an automatic or preattentive stage of visual processing, insofar as the discriminability of the neural responses during this period is not improved by attention, even if the amplitude of the neural response is increased (Hillyard & Anllo-Vento 1998). Second, in contrast, representational content during later stages of visual processing (e.g., 150–250 ms post stimulus presentation) seems to be highly sensitive to attentional influence, such that object representations become more distinct (i.e., discriminable) when observers monitor the visual sequence for potential targets (Grootswagers et al. 2021, Marti & Dehaene 2017, Moerel et al. 2022b). Finally, targets elicit more prolonged representations relative to distractors; target selection itself is associated with a later, discrete stage of processing, as indexed by above-chance decoding from 300 ms that is not evident in distractors (Marti & Dehaene 2017).

Computational methods such as decoding have a particular advantage in elucidating top-down effects on visual processing thanks to their ability to examine the representational format of multiple stimuli presented concurrently. For example, studies have used decoding to understand how attending to one of two overlaid objects influences the neural representations of each: While both object identities appear to be represented in the neural response to the display, the representation of the attended object becomes more distinct from 100 ms when participants direct their attention to it (Grootswagers et al. 2021). Furthermore, these attention-related refinements of the representational code appear sooner in time for the smaller item at the fovea than for the larger item that extended further into the periphery, suggesting a difference in attentional effects on local and global neural representations. Similar effects have been reported using fMRI, where directing attention to one of two overlaid objects improves the discriminability of that category from others in occipitotemporal regions of the brain (Keller et al. 2022). Importantly, in this study, category representations were most facilitated when residual correlations between category-selective regions of the ventral temporal cortex and higher-order regions were highest, suggesting that attentional enhancement of visual information is instigated by top-down sharing of information (Keller et al. 2022). In accordance with this, using MEG, Goddard and colleagues (2022) found that stimulus coding in frontal brain regions preceded attentional effects in the occipital cortex and that these effects occurred earlier for spatial than for feature-based attention. These findings extend previous work by showing that both attended and unattended stimuli are represented in visual responses, and that top-down attention enhances the basic visual representations of stimuli, rather than involving a fundamental change in how stimuli are represented during attentive tasks.

A technique called cross-decoding has been particularly useful for elucidating how mechanisms underlying high-level vision are implemented during other cognitive processes. For example, Gayet & Peelen (2022) investigated preparatory attention mechanisms in a visual search paradigm in which participants were cued to search for a melon or a box. Of key interest was the neural response to target-absent displays, where a cross-decoding approach showed that neural activation patterns within object selective cortex, but not early visual cortex, corresponded specifically to the target object that observers were holding in mind. Furthermore, they found that preparatory object templates were size specific, such that search representations for objects at further distances (i.e., smaller expected retinal size) did not generalize to search when the object was expected to be closer and thus larger. These findings provide a model of how efficient visual search might proceed in the real world, where the observer might construct an internal representation of the search item as it is likely to look in this context (i.e., smaller or larger depending on where it is likely to be relative to the viewer).

**Mental imagery.** As a top-down process, visual imagery is an intriguing space in which both decoding and encoding models have made substantive progress (see the sidebar titled Encoding Models). In classic retro-cue imagery paradigms, participants see two candidate images in successive intervals, followed by a postcue that indicates which of these two images they should imagine in a subsequent test phase (Dijkstra et al. 2017a, Harrison & Tong 2009). Applied to the imagined stimulus, neural decoding offers the unique possibility to assess similarities in the neural instantiation of real and imagined stimuli, in effect revealing the mechanistic overlap in imagery and perception. The most compelling analyses in these designs are those in which a classifier is trained on the neural response evoked by (real) visual images and tested on the neural data during the period in which participants must imagine the same images (and vice versa). Classification performance for this cross-decoding directly assesses the degree to which representations that underlie perceiving and imagining something overlap. Considerable evidence for such shared mechanisms between mental imagery and perception now exists in both the spatial and temporal domain (Breedlove et al. 2020; Dijkstra et al. 2017b, 2018, 2020; Naselaris et al. 2015; Robinson et al. 2021). Imagery representations are position specific (Robinson et al. 2021), object specific (Dijkstra et al. 2018), and distinctive for complex scenes (Breedlove et al. 2020). Additionally, there is greater similarity between visual and imagined representations in higher-level visual brain regions than in lower-level ones (Breedlove et al. 2020, Ragni et al. 2020) and for later temporal processes than for earlier ones (Dijkstra et al. 2018, Robinson et al. 2021). Furthermore, time-resolved decoding has shown that the dynamics of imagery differ from that of visual perception, with increased variability in the timing of imagery (Dijkstra et al. 2018), to the extent that imagery representations cannot always be decoded using time-resolved methods that assume similar timing from trial to trial (Shatek et al. 2019). Theories of mental imagery have progressed from this imagery-decoding literature, and the current status of the field depicts imagery as a reversal of perception (Dijkstra et al. 2020), relying on generative feedback from high-level brain regions (Breedlove et al. 2020). These proposed mechanisms of imagery have raised the possibility that there is inherent interference between bottom-up perception and top-down imagery that can account for individual differences in the quality of imagery (Sulfaro et al. 2022).

**Prediction and expectation.** Another high-level influence on visual processing is prediction, which serves to guide the interpretation of incoming sensory input (Summerfield & de Lange 2014)—for example, cue-based (Posner et al. 1980) or context-based (Bar 2004) expectations about what is (or is not) likely to occur can influence representations of visual stimuli. Decoding has been instrumental in elucidating the neural basis of how predictions influence our perception of the world, showing that the earliest stages of perception are subject to influence by what we expect to see. Remarkably, in another example of internally generated representations, early visual cortex has been shown to reflect feature-specific patterns of expected (but omitted) stimuli (Ekman et al. 2017, Kok et al. 2014). Relatedly, expected stimuli can be decoded even prior to stimulus presentation (Blom et al. 2020, Kok et al. 2017). In a different line of work, expectations engendered by scene information have been shown to facilitate decoding of degraded object stimuli whose identity is hard to perceive in isolation but readily apparent when presented in the context (Brandman & Peelen 2017). Such expectation-based disambiguation of object representations has subsequently been shown to be causally related to feedback connections from scene-selective to object-selective regions of visual cortex (Wischnewski & Peelen 2021). Similarly, representations of occluded objects seem to necessarily rely on our knowledge and expectations, requiring an internal reconstruction of image features that are not visible due to an occluder. That object decoding is both reduced and delayed under partial occlusion suggests that these representations may rely on recurrent processes (Rajaei et al. 2019). These results highlight that prediction plays a role in neural processing at different time scales and add to the growing body of evidence that contextual expectations enhance representations in the visual system.

Elsewhere, decoding approaches have also been valuable in revealing the visual system's sensitivity to the statistical regularities of real-world environments (for a review, see Kaiser et al. 2019). Basic perceptual analysis of objects can be affected both by their typical absolute location in space (e.g., planes in the upper visual field, shoes in the lower visual field) (de Haas et al. 2016, Kaiser & Cichy 2018) and by their typical positioning with respect to other objects (e.g., cup over saucer versus saucer over cup) (Kaiser & Peelen 2018). This line of research suggests that learned associations between objects' identities and their typical spatial positions can (interactively) influence visual processing of high-level stimuli (Quek & Peelen 2020), a potential mechanism for reducing visual competition to support efficient object representation in the face of limited cortical resources (Kaiser et al. 2019).

### Representational Stability: Dependent on Feedforward, Feedback, and Recurrent Processing?

Prior to sensitive pattern-based methods, it was difficult to deduce the quality of representational content in the visual system under conditions of stimulus competition or the absence of stimuli. Now we have abundant evidence that neural representations are reliably elicited under many different circumstances, indicating a robust feedforward sweep of information yielding consistent pattern-based responses (DiCarlo & Cox 2007). Representational modulation, by contrast, appears to involve recurrent processing within the visual system (Rajaei et al. 2019) and feedback processing from high-level frontoparietal regions that enhance or instantiate visual representations in a top-down manner (Dijkstra et al. 2017b, Keller et al. 2022). A converging line of research shows that perception-like stimulus representations can be induced in the absence of that stimulus under different circumstances, for example, in expectancy (Blom et al. 2020; Kok et al. 2014, 2017), working memory (Albers et al. 2013), attentional preparation (Gayet & Peelen 2022), imagery (Dijkstra et al. 2018, 2020), and occlusion (Teichmann et al. 2022). Efforts to understand the instantiations and changes in representational content associated with different internal states constitute an emerging field with great potential to elucidate how visual processing produces perception.

## NEW DIRECTIONS

Although most work to date has focused on neural decoding in healthy adults, decoding approaches have enormous potential to provide new insights into visual processing in other populations. For example, understanding how visual stimuli are represented during development can lead to greater understanding of the architecture of the visual system. EEG decoding in particular is useful as it has higher tolerance to the movement typically exhibited by children (Ashton et al. 2022). Recent research in this emerging field has shown that time-resolved neural responses to visual objects in infants are distinct from those in adults, suggesting a reorganization of representational structure during development (Bayet et al. 2020). Other work has shown that the distinction of faces from houses is similar in 6–11-year-olds and adults, although face inversion decoding is much more robust in adults, suggesting that visual experience shapes neural representations even in later stages of development (Mares et al. 2020).

At the same time, neural decoding also lends itself well to single-participant analyses and can therefore provide insights into how visual representations are altered in cases of visual deficits (i.e., case studies). Various work has found neural correlates of representational differences in disorders such as prosopagnosia, a deficit in face recognition (Rivolta et al. 2014), and visual object agnosia, a deficit in object recognition (Haigh et al. 2018). There is almost endless possibility in this space, and future work will likely tackle in greater detail how representations vary during development, after brain injury, during healthy aging, and in clinical populations, among other circumstances.

## CONCLUSION

The past several decades have revealed the vast utility of multivariate analyses of neural responses for the study of visual processing. Great strides have been made in understanding and modeling neural responses in the brain by investigating the brain regions and dynamics associated with visual representations (Cichy et al. 2014), the way that these representations generalize over time (King & Dehaene 2014), and how they change with different tasks (Hebart et al. 2018). Decoding approaches have improved our understanding about the spatial and temporal extent of visual processing related to both low-level image features and high-level categorical understanding and have shone a light on how these representations relate to behavior. Furthermore, we know now that representations can be evoked within the visual system in the absence of visual input, for example, during imagery, yielding insights into the processes and purposes underlying these internally generated representations. The sensitivity of decoding analyses is their greatest power, and in combination with clever paradigms, multivariate methods have the potential to continue to elucidate how neural populations function for visual perception.

## SUMMARY POINTS

1. Multivariate decoding is an information-based framework that assesses the similarity and discriminability of neural activation patterns associated with perceptual inputs or cognitive states.
2. Decoding methods have advanced our understanding of the spatiotemporal characteristics of visual processing, revealing how representational content varies across different stages of processing and regions of visual cortex.
3. Representations of objects' perceptual and conceptual features show overlapping temporal profiles; the covarying nature of these features presents a challenge for disentangling the unique contribution of each to visual representations elicited by object images.
4. Decoding has provided new insight into both the stability and permeability of visual representations. However, robust visual representations are subject to modulation by aspects like attention and expectation, and can even be generated by top-down mechanisms in the absence of perceptual input (e.g., visual imagery, search templates).
5. Great strides have been made in relating neural representations to behavior; the importance of this link for interpreting representations remains a point of contention in the literature.
6. Fine-grained assessment of the representational overlap in the neural responses to different stimuli gives rise to representational geometries that reveal latent principles of the ways in which information is encoded in the brain.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

This work was supported by an Australia Research Council (ARC) Discovery Early Researcher Award to A.K.R (DE200101159) and an ARC Discovery Project to T.A.C (DP200101787). We thank J. Brendan Ritchie, Surya Gayet, Anthony Harris, Dragan Rangelov, and Alexander Sulfaro for helpful comments.

## LITERATURE CITED

Albers AM, Kok P, Toni I, Dijkerman HC, de Lange FP. 2013. Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* 23(15):1427–31

Ashton K, Zinszer BD, Cichy RM, Nelson CA, Aslin RN, Bayet L. 2022. Time-resolved multivariate pattern analysis of infant EEG data: a practical tutorial. *Dev. Cogn. Neurosci.* 54:101094

Baker B, Lansdell B, Kording KP. 2022. Three aspects of representation in neuroscience. *Trends Cogn. Sci.* 26(11):942–58

Bar M. 2004. Visual objects in context. *Nat. Rev. Neurosci.* 5(8):617–29

Barron HC, Garvert MM, Behrens TEJ. 2016. Repetition suppression: a means to index neural representations using BOLD? *Philos. Trans. R. Soc. B* 371(1705):20150355

Bayet L, Zinszer BD, Reilly E, Cataldo JK, Pruitt Z, et al. 2020. Temporal dynamics of visual representations in the infant brain. *Dev. Cogn. Neurosci.* 45:100860

Blom T, Feuerriegel D, Johnson P, Bode S, Hogendoorn H. 2020. Predictions drive neural representations of visual events ahead of incoming sensory information. *PNAS* 117(13):7510–15

Bode S, Feuerriegel D, Bennett D, Alday PM. 2019. The Decision Decoding ToolBOX (DDTBOX)—a multivariate pattern analysis toolbox for event-related potentials. *Neuroinformatics* 17(1):27–42

Bracci S, Op de Beeck HP. 2016. Dissociations and associations between shape and category representations in the two visual pathways. *J. Neurosci.* 36(2):432–44

Bracci S, Op de Beeck H. 2022. Understanding human object vision: a picture is worth a thousand representations. *Annu. Rev. Psychol.* 74:113–35

Bracci S, Ritchie JB, Kalfas I, Op de Beeck HP. 2019. The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *J. Neurosci.* 39(33):6513–25

Bracci S, Ritchie JB, Op de Beeck H. 2017. On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia* 105:153–64

Brandman T, Peelen MV. 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *J. Neurosci.* 37(32):7700–10

Breedlove JL, St-Yves G, Olman CA, Naselaris T. 2020. Generative feedback explains distinct brain activity codes for seen and mental images. *Curr. Biol.* 30(12):2211–24.e6

Brouwer GJ, Heeger DJ. 2009. Decoding and reconstructing color from responses in human visual cortex. *J. Neurosci.* 29(44):13992–93

Carlson T, Goddard E, Kaplan DM, Klein C, Ritchie JB. 2018. Ghosts in machine learning for cognitive neuroscience: moving from data to theory. *NeuroImage* 180(Pt A):88–100

Carlson TA, Grootswagers T, Robinson AK. 2020. An introduction to time-resolved decoding analysis for M/EEG. In *The Cognitive Neurosciences*, ed. D Poeppel, GR Mangun, MS Gazzaniga, pp. 679–90. Cambridge, MA: MIT Press. 6th ed.

Carlson TA, Hogendoorn H, Kanai R, Mesik J, Turret J. 2011. High temporal resolution decoding of object position and category. *J. Vis.* 11(10):9

Carlson TA, Simmons RA, Kriegeskorte N, Slevc LR. 2014. The emergence of semantic meaning in the ventral temporal pathway. *J. Cogn. Neurosci.* 26(1):120–31

Carlson TA, Tovar DA, Alink A, Kriegeskorte N. 2013. Representational dynamics of object vision: the first 1000 ms. *J. Vis.* 13(10):1

Chang N, Pyles JA, Marcus A, Gupta A, Tarr MJ, Aminoff EM. 2019. BOLD5000, a public fMRI dataset while viewing 5000 visual images. *Sci. Data* 6:49

Cichy RM, Pantazis D, Oliva A. 2014. Resolving human object recognition in space and time. *Nat. Neurosci.* 17(3):455–62

Cichy RM, Pantazis D, Oliva A. 2016. Similarity-based fusion of MEG and fMRI reveals spatio-temporal dynamics in human cortex during visual object recognition. *Cereb. Cortex.* 26(8):3563–79

Cichy RM, Ramirez FM, Pantazis D. 2015. Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *NeuroImage* 121:193–204

Contini EW, Goddard E, Wardle SG. 2021. Reaction times predict dynamic brain representations measured with MEG for only some object categorisation tasks. *Neuropsychologia* 151:107687

Contini EW, Wardle SG, Carlson TA. 2017. Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia* 105:165–76

de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, et al. 2016. Perception and processing of faces in the human brain is tuned to typical feature locations. *J. Neurosci.* 36(36):9289–302

Desimone R, Duncan J. 1995. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18:193–222

DiCarlo JJ, Cox DD. 2007. Untangling invariant object recognition. *Trends Cogn. Sci.* 11(8):333–41

Dijkstra N, Ambrogioni L, Vidaurre D, van Gerven M. 2020. Neural dynamics of perceptual inference and its reversal during imagery. *eLife* 9:e53588

Dijkstra N, Bosch SE, van Gerven MAJ. 2017a. Vividness of visual imagery depends on the neural overlap with perception in visual areas. *J. Neurosci.* 37(5):1367–73

Dijkstra N, Mostert P, de Lange FP, Bosch S, van Gerven MAJ. 2018. Differential temporal dynamics during visual imagery and perception. *eLife* 7:e33904

Dijkstra N, Zeidman P, Ondobaka S, van Gerven MAJ, Friston K. 2017b. Distinct top-down and bottom-up brain connectivity during visual perception and imagery. *Sci. Rep.* 7:5677

Downing PE, Jiang Y, Shuman M, Kanwisher N. 2001. A cortical area selective for visual processing of the human body. *Science* 293(5539):2470–73

Edelman S. 1998. Representation is representation of similarities. *Behav. Brain Sci.* 21(4):449–67; discussion 467–98

Ekman M, Kok P, De Lange FP. 2017. Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* 8:15276

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

Gayet S, Peelen MV. 2022. Preparatory attention incorporates contextual expectations. *Curr. Biol.* 32(3):687–92.e6

Gifford AT, Dwivedi K, Roig G, Cichy RM. 2022. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage* 264:119754

Goddard E, Carlson TA, Woolgar A. 2022. Spatial and feature-selective attention have distinct, interacting effects on population-level tuning. *J. Cogn. Neurosci.* 34(2):290–312

Grill-Spector K, Malach R. 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol.* 107(1–3):293–321

Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15(8):536–48

Grootswagers T, Cichy RM, Carlson TA. 2018. Finding decodable information that can be read out in behaviour. *NeuroImage* 179:252–62

Grootswagers T, Kennedy BL, Most SB, Carlson TA. 2020. Neural signatures of dynamic emotion constructs in the human brain. *Neuropsychologia* 145:106535

Grootswagers T, Robinson AK. 2021. Overfitting the literature to one set of stimuli and data. *Front. Hum. Neurosci.* 15:682661

Grootswagers T, Robinson AK, Carlson TA. 2019a. The representational dynamics of visual objects in rapid serial visual processing streams. *NeuroImage*. 188:668–79

Grootswagers T, Robinson AK, Shatek SM, Carlson TA. 2019b. Untangling featural and conceptual object representations. *NeuroImage* 202:116083

Grootswagers T, Robinson AK, Shatek SM, Carlson TA. 2021. The neural dynamics underlying prioritisation of task-relevant information. *Neurons Behav. Data Anal. Theory.* 5(1). **https://doi.org/10.51628/001c.21174**

Grootswagers T, Wardle SG, Carlson TA. 2017. Decoding dynamic brain patterns from evoked responses: a tutorial on multivariate pattern analysis applied to time series neuroimaging data. *J. Cogn. Neurosci.* 29(4):677–97

Grootswagers T, Zhou I, Robinson AK, Hebart MN, Carlson TA. 2022. Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Sci. Data.* 9(1):3

Haigh S, Robinson A, Grover P, Behrmann M. 2018. Differentiation of types of visual agnosia using EEG. *Vision* 2(4):44

Hanson SJ, Matsuka T, Haxby JV. 2004. Combinatorial codes in ventral temporal lobe for object recognition: Haxby 2001 revisited: Is there a "face" area? *NeuroImage* 23(1):156–66

Harel A, Kravitz DJ, Baker CI. 2014. Task context impacts visual object processing differentially across the cortex. *PNAS* 111(10):E962–71

Harrison SA, Tong F. 2009. Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458(7238):632–35

Harrison WJ. 2022. Luminance and contrast of images in the THINGS database. *Perception* 51(4):244–62

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293(5539):2425–30

Haynes J-D, Rees G. 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8(5):686–91

He K, Zhang X, Ren S, Sun J. 2015. Deep residual learning for image recognition. arXiv:1512.03385v1 [cs.CV]

Hebart MN, Baker CI. 2018. Deconstructing multivariate decoding for the study of brain function. *NeuroImage*. 180(Pt A):4–18

Hebart MN, Bankson BB, Harel A, Baker CI, Cichy RM. 2018. The representational dynamics of task and object processing in humans. *eLife* 7:e32816

Hebart MN, Contier O, Teichmann L, Rockter AH, Zheng CY, et al. 2023. THINGS-data: a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior. bioRxiv 2022.07.22.501123. **https://doi.org/10.1101/2022.07.22.501123**

Hebart MN, Dickter AH, Kidder A, Kwok WY, Corriveau A, et al. 2019. THINGS: a database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE* 14(10):e0223792

Hebart MN, Görgen K, Haynes J-D. 2015. The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Front. Neuroinform.* 8:88

Hillyard SA, Anllo-Vento L. 1998. Event-related brain potentials in the study of visual selective attention. *PNAS* 95(3):781–87

Kaiser D, Azzalini DC, Peelen MV. 2016. Shape-independent object category responses revealed by MEG and fMRI decoding. *J. Neurophysiol.* 115(4):2246–50

Kaiser D, Cichy RM. 2018. Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J. Neurophysiol.* 120(2):848–53

Kaiser D, Peelen MV. 2018. Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*. 169:334–41

Kaiser D, Quek GL, Cichy RM, Peelen MV. 2019. Object vision in a structured world. *Trends Cogn. Sci.* 23:672–85

Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8(5):679–85

Kamitani Y, Tong F. 2006. Decoding seen and attended motion directions from activity in the human visual cortex. *Curr. Biol.* 16(11):1096–102

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17(11):4302–11

Kastner S, De Weerd P, Desimone R, Ungerleider LG. 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282(5386):108–11

Kay KN, Naselaris T, Prenger RJ, Gallant JL. 2008. Identifying natural images from human brain activity. *Nature* 452(7185):352–55

Keller AS, Jagadeesh AV, Bugatus L, Williams LM, Grill-Spector K. 2022. Attention enhances category representations across the brain with strengthened residual correlations to ventral temporal cortex. *NeuroImage* 249:118900

Khaligh-Razavi S-M, Cichy RM, Pantazis D, Oliva A. 2018. Tracking the spatiotemporal neural dynamics of real-world object size and animacy in the human brain. *J. Cogn. Neurosci.* 30(11):1559–76

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):4296–309

King J-R, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. *Trends Cogn. Sci.* 18(4):203–10

King J-R, Wyart V. 2021. The human brain encodes a chronicle of visual events at each instant of time through the multiplexing of traveling waves. *J. Neurosci.* 41(34):7224–33

Kok P, Failing MF, de Lange FP. 2014. Prior expectations evoke stimulus templates in the primary visual cortex. *J. Cogn. Neurosci.* 26(7):1546–54

Kok P, Mostert P, de Lange FP. 2017. Prior expectations induce prestimulus sensory templates. *PNAS* 114(39):10473–78

Konkle T, Oliva A. 2012. A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74(6):1114–24

Kragel PA, Koban L, Barrett LF, Wager TD. 2018. Representation, pattern information, and brain signatures: from neurons to neuroimaging. *Neuron* 99(2):257–73

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *PNAS* 103(10):3863–68

Kriegeskorte N, Kievit RA. 2013. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* 17(8):401–12

Kriegeskorte N, Mur M, Bandettini PA. 2008a. Representational similarity analysis—connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2:4

Kriegeskorte N, Mur M, Ruff DA, Kiani R, Bodurka J, et al. 2008b. Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron* 60(6):1126–41

Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001. Center-periphery organization of human object areas. *Nat. Neurosci.* 4(5):533–39

Long B, Störmer VS, Alvarez GA. 2017. Mid-level perceptual features contain early cues to animacy. *J. Vis.* 17:20

Long B, Yu C-P, Konkle T. 2018. Mid-level visual features underlie the high-level categorical organization of the ventral stream. *PNAS* 115:E9015–24

Macevoy SP, Epstein RA. 2009. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* 19(11):943–47

Mangun GR. 1995. Neural mechanisms of visual selective attention. *Psychophysiology* 32(1):4–18

Mares I, Ewing L, Farran EK, Smith FW, Smith ML. 2020. Developmental changes in the processing of faces as revealed by EEG decoding. *NeuroImage* 211:116660

Marti S, Dehaene S. 2017. Discrete and continuous mechanisms of temporal selection in rapid visual streams. *Nat. Commun.* 8:1955

McCarthy G, Puce A, Gore JC, Allison T. 1997. Face-specific processing in the human fusiform gyrus. *J. Cogn. Neurosci.* 9(5):605–10

Mehrer J, Spoerer CJ, Jones EC, Kriegeskorte N, Kietzmann TC. 2021. An ecologically motivated image dataset for deep learning yields better models of human vision. *PNAS* 118(8):e2011417118

Moerel D, Grootswagers T, Robinson A, Engeler P, Holcombe AO, Carlson TA. 2022a. Rotation-tolerant representations elucidate the time-course of high-level object processing. PsyArXiv. **https://doi.org/10.31234/osf.io/wp73u**

Moerel D, Grootswagers T, Robinson AK, Shatek SM, Woolgar A, et al. 2022b. The time-course of feature-based attention effects dissociated from temporal expectation and target-related processes. *Sci. Rep.* 12:6968

Mohsenzadeh Y, Qin S, Cichy RM, Pantazis D. 2018. Ultra-rapid serial visual presentation reveals dynamics of feedforward and feedback processes in the ventral visual pathway. *eLife* 7:e36329

Moshel ML, Robinson AK, Carlson TA, Grootswagers T. 2022. Are you for real? Decoding realistic AI-generated faces from neural activity. *Vis. Res.* 199:108079

Mur M, Meys M, Bodurka J, Goebel R, Bandettini P, Kriegeskorte N. 2013. Human object-similarity judgments reflect and transcend the primate-IT object representation. *Front. Psychol.* 4:128

Naselaris T, Kay KN, Nishimoto S, Gallant JL. 2011. Encoding and decoding in fMRI. *NeuroImage* 56(2):400–10

Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL. 2015. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* 105:215–28

Oosterhof NN, Connolly AC, Haxby JV. 2016. CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Front. Neuroinform.* 10:27

O'Toole AJ, Jiang F, Abdi H, Haxby JV. 2005. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17(4):580–90

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45(1 Suppl):S199–209

Philiastides MG, Sajda P. 2006. Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb. Cortex* 16(4):509–18

Popham SF, Huth AG, Bilenko NY, Deniz F, Gao JS, et al. 2021. Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nat. Neurosci.* 24(11):1628–36

Posner MI, Snyder CRR, Davidson BJ. 1980. Attention and the detection of signals. *J. Exp. Psychol. Gen.* 109(2):160–74

Puce A, Allison T, Asgari M, Gore JC, McCarthy G. 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* 16(16):5205–15

Quax SC, Dijkstra N, van Staveren MJ, Bosch SE, van Gerven MAJ. 2019. Eye movements explain decodability during perception and cued attention in MEG. *NeuroImage* 195:444–53

Quek GL, Peelen MV. 2020. Contextual and spatial associations between objects interactively modulate visual processing. *Cereb. Cortex* 30(12):6391–404

Ragni F, Tucciarelli R, Andersson P, Lingnau A. 2020. Decoding stimulus identity in occipital, parietal and inferotemporal cortices during visual mental imagery. *Cortex* 127:371–87

Rajaei K, Mohsenzadeh Y, Ebrahimpour R, Khaligh-Razavi S-M. 2019. Beyond core object recognition: recurrent processes account for object recognition under occlusion. *PLOS Comput. Biol.* 15(5):e1007001

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):7691–99

Ritchie JB, Carlson TA. 2016. Neural decoding and "inner" psychophysics: a distance-to-bound approach for linking mind, brain, and behavior. *Front. Neurosci.* 10:190

Ritchie JB, Tovar DA, Carlson TA. 2015. Emerging object representations in the visual system predict reaction times for categorization. *PLOS Comput. Biol.* 11(6):e1004316

Ritchie JB, Zeman AA, Bosmans J, Sun S, Verhaegen K, Op de Beeck HP. 2021. Untangling the animacy organization of occipitotemporal cortex. *J. Neurosci.* 41(33):7103–19

Rivolta D, Woolgar A, Palermo R, Butko M, Schmalzl L, Williams MA. 2014. Multi-voxel pattern analysis (MVPA) reveals abnormal fMRI activity in both the "core" and "extended" face network in congenital prosopagnosia. *Front. Hum. Neurosci.* 8:925

Robinson AK, Grootswagers T, Carlson TA. 2019. The influence of image masking on object representations during rapid serial visual presentation. *NeuroImage*. 197:224–31

Robinson AK, Grootswagers T, Shatek SM, Gerboni J, Holcombe A, Carlson TA. 2021. Overlapping neural representations for the position of visible and imagined objects. *Neurons Behav. Data Anal. Theory.* 4(1):1–28

Robinson AK, Rich AN, Woolgar A. 2022. Linking the brain with behavior: the neural dynamics of success and failure in goal-directed behavior. *J. Cogn. Neurosci.* 34(4):639–54

Robinson AK, Venkatesh P, Boring MJ, Tarr MJ, Grover P, Behrmann M. 2017. Very high density EEG elucidates spatiotemporal aspects of early visual processing. *Sci. Rep.* 7:16248

Rosenthal IA, Singh SR, Hermann KL, Pantazis D, Conway BR. 2021. Color space geometry uncovered with magnetoencephalography. *Curr. Biol.* 31(3):515–26.e5

Seymour K, Clifford CWG, Logothetis NK, Bartels A. 2009. The coding of color, motion, and their conjunction in the human visual cortex. *Curr. Biol.* 19(3):177–83

Sha L, Haxby JV, Abdi H, Guntupalli JS, Oosterhof NN, et al. 2015. The animacy continuum in the human ventral vision pathway. *J. Cogn. Neurosci.* 27(4):665–78

Shatek SM, Grootswagers T, Robinson AK, Carlson TA. 2019. Decoding images in the mind's eye: the temporal dynamics of visual imagery. *Vision* 3(4):53

Shatek SM, Robinson AK, Grootswagers T, Carlson TA. 2022. Capacity for movement is an organisational principle in object representations. *NeuroImage* 261:119517

Shepard RN, Chipman S. 1970. Second-order isomorphism of internal representations: shapes of states. *Cogn. Psychol.* 1(1):1–17

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J. 2013. Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78(2):364–75

Sulfaro AA, Robinson AK, Carlson TA. 2022. Perception as a hierarchical competition: a model that differentiates imagined, veridical, and hallucinated percepts. bioRxiv 2022.09.02.506121. **https://doi.org/10.1101/2022.09.02.506121**

Summerfield C, de Lange FP. 2014. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* 15(11):745–56

Taubert J, Wardle SG, Ungerleider LG. 2020. What does a "face cell" want? *Prog. Neurobiol.* 195:101880

Teichmann L, Moerel D, Baker C, Grootswagers T. 2021. An empirically-driven guide on using Bayes Factors for M/EEG decoding. *Aperture Neuro*. **https://doi.org/10.52294/ApertureNeuro.2022.2.MAOC6465**

Teichmann L, Moerel D, Rich AN, Baker CI. 2022. The nature of neural object representations during dynamic occlusion. *Cortex* 153:66–86

Teichmann L, Quek GL, Robinson AK, Grootswagers T, Carlson TA, Rich AN. 2020. The influence of object-colour knowledge on emerging object representations in the brain. *J. Neurosci.* 40(35):6779–89

Thorat S, Proklova D, Peelen MV. 2019. The nature of the animacy organization in human ventral temporal cortex. *eLife* 8:e47142

Wang R, Janini D, Konkle T. 2022. Mid-level feature differences support early animacy and object size distinctions: evidence from electroencephalography decoding. *J. Cogn. Neurosci.* 34(9):1670–80

Wardle SG, Kriegeskorte N, Grootswagers T, Khaligh-Razavi SM, Carlson TA. 2016. Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *NeuroImage* 132:59–70

Wardle SG, Taubert J, Teichmann L, Baker CI. 2020. Rapid and dynamic processing of face pareidolia in the human brain. *Nat. Commun.* 11:4518

Webster MA. 2011. Adaptation and visual coding. *J. Vis.* 11(5):3

Williams MA, Dang S, Kanwisher NG. 2007. Only some spatial patterns of fMRI response are read out in task performance. *Nat. Neurosci.* 10(6):685–86

Wischnewski M, Peelen MV. 2021. Causal neural mechanisms of context-based object recognition. *eLife* 10:e69736