What's there, distinctly, when and where?

Marieke Mur & Nikolaus Kriegeskorte

A study shows the transience of early visual representations (while the stimulus is still on) and the persistence of higher representations (outlasting the stimulus) as various categorical distinctions emerge at staggered latencies. Rather than slavishly following the stimulus, representations interact through recurrent signals to infer what's there.

When you open your eyes to an image, a wave of activity sweeps through your brain. From lower to higher visual areas, the image is represented and re-represented at increasing levels of abstraction. At the same time, signals feed back from higher to lower areas. This dynamic process gives rise to your interpretation of the image, imposing categorical boundaries that are in the eye—or, more precisely, in the brain—of the beholder, and emphasizing what the image means to you. For each object-category distinction ('what'), researchers would like to understand when (how long after stimulus onset) it becomes explicitly represented where in the brain (in which visual area). The temporal and spatial aspects of human object recognition have so far been studied separately. In this issue of *Nature Neuroscience*, Cichy *et al.*[1] devised a clever way to relate the two and determine what distinctions between pairs of objects[2] emerge when and where in the brain.

The first cortical stage for incoming visual information is the primary visual cortex (V1), which is located at the back of the brain in the occipital lobe. From there, information travels forward along the ventral visual stream, which culminates in the inferior temporal (IT) cortex. At each stage, an image is represented by the activity pattern across the area's population of neurons. Progressing along the hierarchy, image representations show increasing selectivity for categories and increasing tolerance to changes in position, view, lighting and other so-called 'accidental' properties[3].

Visual object recognition is a rapid process. For example, humans can detect the presence of an animal in a visual scene at latencies of 120 ms (ref. 4). Recent studies showed that object identity and category membership can be decoded from human brain activity less than 100 ms after stimulus onset[5–7]. Fast feedforward processing can account for a component of the object recognition process[8]. However, local recurrent computations and delayed feedback from higher to lower areas are also important[9]. Recurrent processing might explain why category information at higher levels of abstraction appears to emerge later than category information more closely related to the visual input[6].

Human brain activity is commonly measured using functional magnetic resonance imaging (fMRI) or magnetoencephalography (MEG). Of the two, fMRI has better spatial resolution. It gives us largely independent information about the activity for each brain area. It can even be used to characterize the activity pattern representing a stimulus within an area and to decode some of the represented information[10,11]. However, it reflects the temporally sluggish hemodynamic response and therefore gives us little information about the dynamics of processing within a single perceptual act, which might take only a few hundred milliseconds. MEG, conversely, has a temporal resolution in the millisecond range, but poorer spatial precision. It measures the small magnetic fields generated by neuronal activity, using an array of sensors placed around the head.

The representation of an image is thought to rely on the pattern of activity across all of the neurons in a visual area—a neuronal 'population code'. Recent work has used multivariate pattern-information analyses, including decoders, which attempt to read out the contents of the representation[6,10–13]. Moreover, recent studies have employed rich sets of stimuli and characterized how well each distinction between two objects is reflected in an area's activity patterns[2,13–15].

Cichy *et al.*[1] investigated the representation of 92 object images from a range of categories, measuring the brain activity pattern elicited by each image, first with MEG and later with fMRI, in the same 15 subjects. They used representational similarity analysis (RSA)[2] to investigate the representational similarity space for each brain region (based on the fMRI data) and latency after stimulus onset (based on the MEG data). RSA characterizes a representation by a matrix containing a number for each pair of images that specifies how distinct the two images are in the representation.

Cichy *et al.*[1] then connected space and time in an innovative way. They correlated the pattern of representational distinctions estimated from the fMRI data for V1 and IT with the pattern of distinctions from the MEG data (reflecting the entire ventral stream) at each point in time after stimulus onset. This enabled them to estimate a time course, at high-temporal resolution, for the emergence of the V1 pattern of object distinctions and a separate time course for the emergence of the IT pattern of distinctions.

As expected, the V1 representation emerged rapidly and earlier than the IT representation. The V1 representation was quite transient, peaking after 100 ms, and then decaying rapidly even while the 500-ms stimulus was still on. The IT representation emerged slightly later, peaking around 130 ms, and was more persistent. Whereas individual images became distinctly represented rapidly, different categorical divisions became distinct a little later, as has been shown previously[6]. The feedforward and recurrent processing appears to require more time to distinguish more abstract categories, such as animate and inanimate objects, where the members of each category can differ substantially from each other in visual appearance.

Overall, this study reminds us that the dynamics of visual processing is much more complex than a stimulus-evoked feedforward wave of

Marieke Mur & Nikolaus Kriegeskorte are at the Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, UK.
e-mail: nikolaus.kriegeskorte@mrc-cbu.cam.ac.uk

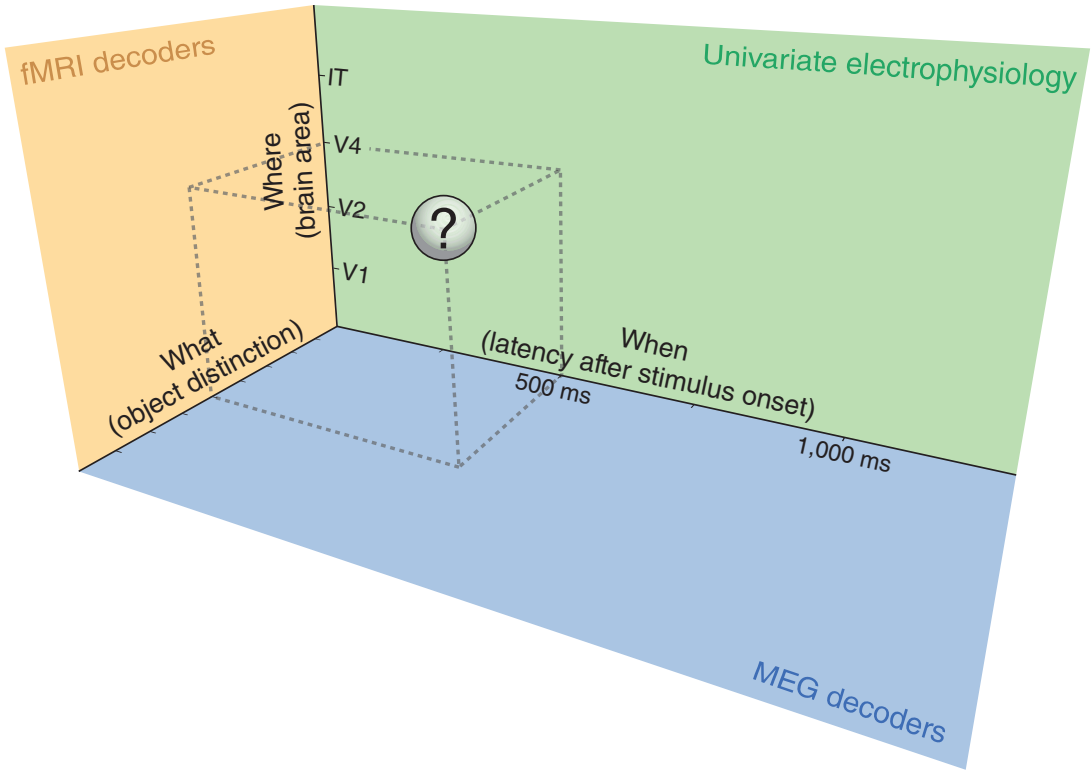

Figure 1 The what, when and where of perceptual processing in the brain. We can summarize what is represented in a brain area by the activity-pattern representational dissimilarity of each pair of images (or each pair of categories). Ideally, we would like to measure all pairwise representational dissimilarities ('what') for each brain area ('where') and latency after stimulus onset ('when'). Combined with pattern decoders, fMRI and MEG can each reveal a kind of projection of the three-dimensional array that fills the space onto the left (orange) and bottom (blue) bounding planes, respectively. The back bounding plane (green) can be characterized by univariate electrophysiology (for example, MEG with source localization). Cichy *et al.*[1] combined representational similarity analyses of MEG and fMRI data to reveal clues to the content of the space for the human brain. Invasive electrode array recordings in multiple areas, combined with pattern decoders, come closest to determining the three-dimensional contents of the space. However, this technique cannot generally be used in human studies.

activity. The emergence of the representation in V1 within 100 ms leaves time for recurrent computations, suggesting that current feedforward models of V1 capture only part of the picture. The rapid decay in V1 of the object distinctions highlights the point that vision, even at this early stage, is not enslaved to the stimulus. Instead, it follows its own rhythm, a rhythm that is perhaps reflected in the frequency of eye movements (about five per second) as we visually explore a complex scene. More speculatively, the representational decay in V1 might also be related to predictive coding. Each area might highlight what is novel and suppress information that has already been incorporated into higher-level representations. The more persistent IT representation might serve the purpose of creating a representation of a scene that is more stable across time and might integrate information over multiple fixations.

The approach of Cichy *et al.*[1] implicitly assumes that there is a single representational similarity space in V1 and another one in IT (as characterized with fMRI), and that any differential emergence of different distinctions over time (MEG) results from the mixture of these representational spaces. This simplifying assumption might be a useful first approximation, enabling us to infer more detailed information about the relationship of the spatial and temporal aspects of the representational dynamics. Ideally, however, we would like to be able to directly measure detailed patterns of activity with simultaneously high spatial and temporal resolution in multiple areas. This would enable us to fill in the three-dimensional space of what distinctions are represented when and where in the brain (**Fig. 1**). In primates, invasive electrode array recordings in multiple areas come closest to providing this information through direct measurement. For human studies, however, the approach used by Cichy *et al.*[1] is likely to be useful in many domains, ranging from perception to decision making and perhaps even to motor control. The study might also inspire new model-based approaches to spatiotemporal analyses of representational similarity.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Cichy, R.M., Pantazis, D. & Oliva, A. *Nat. Neurosci.* **17**, 455–462 (2014).
2. Kriegeskorte, N. & Kievit, R. *Trends Cogn. Sci.* **17**, 401–412 (2013).
3. Rust, N.C. & DiCarlo, J.J. *J. Neurosci.* **30**, 12978–12995 (2010).
4. Kirchner, H. & Thorpe, S.J. *Vision Res.* **46**, 1762–1776 (2006).
5. Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J. & Turret, J. *J. Vis.* **11**, 9 (2011).
6. Carlson, T., Tovar, D.A., Alink, A. & Kriegeskorte, N. *J. Vis.* **13**, 1–19 (2013).
7. Isik, L., Meyers, E.M., Leibo, J.Z. & Poggio, T. *J. Neurophysiol.* **111**, 91–102 (2014).
8. DiCarlo, J.J. & Cox, D.D. *Trends Cogn. Sci.* **11**, 333–341 (2007).
9. Lamme, V.A.F. & Roelfsema, P.R. *Trends Neurosci.* **23**, 571–579 (2000).
10. Haxby, J.V. *et al. Science* **293**, 2425–2430 (2001).
11. Kamitani, Y. & Tong, F. *Nat. Neurosci.* **8**, 679–685 (2005).
12. Hung, C.P., Kreiman, G., Poggio, T. & DiCarlo, J.J. *Science* **310**, 863–866 (2005).
13. Kriegeskorte, N. *et al. Neuron* **60**, 1126–1141 (2008).
14. Kiani, R., Esteky, H., Mirpour, K. & Tanaka, K. *J. Neurophysiol.* **97**, 4296–4309 (2007).
15. Aguirre, G.K. *Neuroimage* **35**, 1480–1494 (2007).