
An Information-Based Approach to Consciousness: Mental State Decoding

John-Dylan Haynes

The debate on the neural correlates of visual consciousness often focuses on the question of which additional processing has to happen for a visual representation to enter consciousness. However, a related question that has only rarely been addressed is which brain regions directly encode specific contents of consciousness. The search for these core neural correlates of contents of consciousness (NCCCs) requires establishing a mapping between sensory experiences and population measures of brain activity in specific brain regions. One approach for establishing this mapping is multivariate decoding. Using this technique, several properties of NCCCs have been investigated. Masking studies have revealed that information about sensory stimuli can be decoded from the primary visual cortex, even if the stimuli cannot be consciously identified by a subject. This suggests that information that does not reach awareness can be encapsulated in early visual stages of processing. Visual imagery representations and veridical perception share similar neural representations in higher-level visual regions, suggesting that these regions are directly related to the encoding of conscious visual experience. But population signals in these higher-level visual regions cannot be the sole carriers of visual experiences because they are invariant to low-level visual features. We found no evidence for increased encoding of sensory information in the prefrontal cortex when a stimulus reaches awareness. In general, we found no role of the prefrontal cortex in encoding sensory experiences at all. However, the improved discrimination of sensory information during perceptual learning could be explained by an improved read-out by the prefrontal cortex. One possible implication is that prefrontal cortical regions do not participate in the encoding of sensory features per se. Instead they may be relevant in making decisions about sensory features, without exhibiting a re-representation of sensory information.

Keywords

Imagery | Masking | Multivariate decoding | Neural correlate of consciousness | Sensory information

1 Introduction

Neural theories of visual consciousness frequently focus on the question of what is needed for a visual stimulus to enter consciousness. A common notion is that representations and processes in sensory regions of the brain can operate outside of conscious perception, and that some “extra property of processing” has to come on top in order to let these representations enter conscious experience (e.g., [Dehaene & Naccache 2001](#)). This extra processing property can range from neural synchronization of neurons encoding the stimulus ([Engel & Singer 2001](#)), recurrent and feedback processing ([Lamme 2006](#), [this collection](#); [Pascual-](#)

[Leone & Walsh 2001](#); [Singer this collection](#)), to participation in a global coherent process, known as neuronal workspace theories ([Baars 2002](#); [Dehaene & Naccache 2001](#)). Discussion of the neural correlates of consciousness (NCC) has often focused on this extra ingredient needed to bring a stimulus representation into consciousness. However, a related, but somewhat different question has often been neglected: Which neural representations (can) precisely participate in encoding the various dimensions of conscious experience? For this it is not enough to establish a correlation between conscious perception and neural signals.

Author

[John-Dylan Haynes](#)

haynes@bccn-berlin.de

Charité – Universitätsmedizin
Berlin, Germany

Commentator

[Caspar M. Schwiedrzik](#)

cschwiedrz@rockefeller.edu

The Rockefeller University
New York, NY, U.S.A.

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

Glossary

Neural encoding	The representation of a sensory feature in a population of neurons.
Mental state decoding	Inferring the representational content of a mental state from a brain activation pattern, typically using multivariate pattern classification.
Neural correlate of a content of consciousness	The brain signal that encodes a specific aspect of conscious experience. The brain signal is the carrier, the specific aspect of consciousness constitutes the phenomenal representational content carried (in short, its “phenomenal content”).
Multivariate pattern classification	A mathematical procedure for identifying patterns of brain activity, the labels of which have been previously learned.
Mapping	The assignment of brain activation patterns to the representational content of mental states.
Low-level visual features	Simple dimensions of visual experience that are encoded in early visual brain regions (e.g., contrast, orientation). If consciously represented, they may constitute corresponding simple forms of phenomenal content.
High-level visual features	More complex dimensions of visual experience that are encoded in downstream visual brain regions (e.g., object identity) and that are to some degree independent of the low-level features by which they are defined.

That would yield a far too large set of candidate brain regions, including, say, signal patterns in the retina that also correlate with conscious perception. Instead, it would be desirable to identify which neural representations most closely encode specific contents of consciousness and can be used to explain dimensions of conscious perception under as many different conditions as possible, and down to the level of single contents. This article will focus on how to identify such core neural correlates of contents of consciousness (NCCCs; [Chalmers 2000](#); [Block 2007](#); [Koch 2004](#)).

It is desirable that studies of visual awareness take NCCCs into account because specific theories of visual awareness make specific predictions regarding the encoding and distribution of sensory information (e.g., [Dehaene & Naccache 2001](#); see also [Baars 2002](#)). In the following, I will first outline the more standard techniques for identifying NCCCs, along with their shortcomings. The next step proposes to use multivariate decoding techniques (reviewed e.g., in [Haynes & Rees 2006](#)) as a tool to identify NCCCs. Decoding can serve as an empirical technique that can establish which brain regions bear most information about specific contents of visual experience. This is an important first step towards establish-

ing a more rigid mapping between visual phenomenal states and content-encoding brain signals. Then, several examples will be presented where multivariate decoding of visual experiences can help inform specific questions regarding NCCCs, such as whether information in V1 participates in visual awareness, whether imagery and perception share the same underlying neural codes, or whether the prefrontal cortex contains any dynamic NCCCs for coding specific dimensions of conscious experience.

2 Why content matters: Binocular rivalry and the multiple levels of conscious experience

In 1996 [Nikos Logothetis & David Leopold](#) published a landmark study on the neural mechanisms of visual awareness. They presented their participating monkeys with a very elaborate visual stimulus display, which allowed them to show one image to one eye and another image to the other eye. For example, the left eye might be stimulated with a line pattern tilted to the left, and the right eye might be stimulated with a line pattern tilted to the right. In such cases, where conflicting input is presented to



Figure 1: Binocular rivalry and levels of perception. (a) Two conflicting stimuli, one presented to the left and one to the right eye, lead to a perceptual alternation between phases where the input of either the left or right eye is consciously seen. In monkey single-cell electrophysiology, this perceptual alternation has a correlate in higher-level visual regions of the temporal lobe, but activity in earlier visual regions shows only small changes in activity patterns. Presumably, signals in the temporal cortex encode the complex figural properties of the stimuli, such as the left being a sunburst pattern and the right being an image of a monkey face. However, due to the invariance of brain responses in higher-level visual regions to low-level features, this cannot explain the perceptual difference between the rivalry of the left and of the right sunburst pattern shown in (b), where the central circle has changed colour but the entire shape remains similar (monkey illustration by Chris Huh, Wikimedia Commons).

the two eyes, human participants don't experience a fusion between the two images. Instead, conscious visual perception alternates between phases where one of the eyes' inputs become visible and phases where the other eye's input are seen. Perception waxes and wanes more or less randomly between two perceptual experiences—despite constant stimulation. Similarly, the monkeys that were exposed to these binocular rivalry stimuli indicated behaviorally by pressing levers that their perception alternated between the inputs to the two eyes.

In parallel, [Leopold & Logothetis \(1996\)](#) investigated what happened to the firing patterns of single neurons in the monkeys' brains. Their setup allowed them to not just look at one location in the brain, but to assess neural correlates in several visual brain regions. They found only a small percentage of single neurons in early visual cortex (V1/V2) whose firing pattern was modulated by the stimulus that was currently dominant. In contrast, in a higher-level visual area—V4—they found that many more cells changed their firing rates with changes in perception. This establishes a clear dissociation between early visual areas where neural signals seem not to correlate with awareness and high visual areas where they do. In a follow-up experiment, [Sheinberg & Logothetis](#)

(1997) investigated the involvement of even higher visual regions in the temporal cortex in binocular rivalry. Because cells in these regions preferentially respond to more elaborate visual features, they used complex shapes and images, such as, for example, an abstract sunburst pattern or a picture of a monkey face ([Figure 1a](#)). They found that in the superior temporal sulcus and in the inferior temporal cortex, a large percentage of cells modulated their firing rate with perceptual dominance. Taken together, these studies seem to suggest that visual awareness affects signals only at late stages of the visual system.

But what does it mean exactly that visual awareness only affects late stages of visual processing? Does it mean that high level visual areas contain all the neural correlates of contents consciousness (NCCCs), in a way similar to a CD encoding the contents of a piece of music? If the signals in these high visual areas are really responsible for encoding all contents of visual experiences then any aspect of conscious perception that changes during binocular rivalry should be explainable by changes in signals in these higher-level brain regions. There are reasons to believe that this cannot be the case. Consider the two images as shown in [Figure 1a](#). At one instant the monkey might

consciously see the face image. This percept would be encoded in activity patterns in the higher visual areas. In the next instant the monkey might see a sunburst pattern, and this experience would also be encoded in the higher visual cortex. At first sight this seems reasonable. Higher-level visual areas are specialized for complex visual information and object features (Sáry et al. 1993). So cells that have a preference for faces might respond during dominance of the face image, and cells with a preference for sunburst patterns might respond during the dominance of that pattern. But there is one difficulty in this interpretation. The images have a high-level interpretation as complex shapes, but they are also composed of a multitude of minute visual features, edges, surfaces, colours, etc. During rivalry, our perception does not only change according to the abstract interpretation, with respect to abstract, high-level interpretation, but also in terms of the minute, fine-grained details of visual experience (see Figure 1b).

This poses a problem because responses in higher-level visual areas are invariant with respect to low-level features (Sáry et al. 1993). Cells in higher-level visual areas in the inferior temporal cortex respond selectively to specific object features in an invariant pattern (Figure 2). A cell specialized for detecting, say, a circle, will respond to this circle irrespective of the low-level features by which it is defined (here brightness, contrast, and colour contrast). This means that such a cell disregards the low-level features and does not convey information about them any more. While cells in high-level visual areas might be able to explain why we see a face one moment and a sunburst pattern the next, they cannot explain why the sunburst pattern is yellow instead of red, or why it is one specific visual pixel collection out of the many possible that would be seen as a sunburst pattern. Thus, visual experience is a multilevel phenomenon, and a theory of the neural correlates of visual awareness will have to be able to explain all the levels of our experience, not just one. This clearly shows the importance of a content-based approach to visual consciousness.

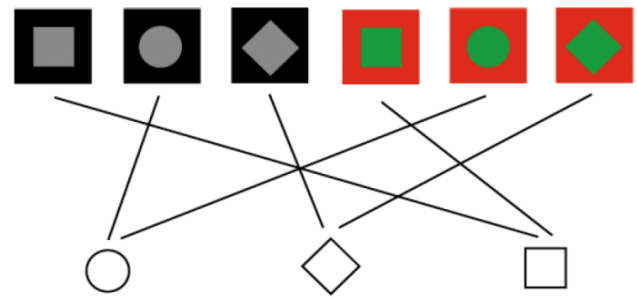


Figure 2: Invariance of single-cell responses in higher-level visual areas. Responses in low-level visual areas (top) are tuned to low-level features such as colour or luminance. In contrast, responses in higher-level object-selective regions (bottom) are largely invariant with respect to these low-level features (Sáry et al. 1993).

3 Approaches to content-selectivity in human neuroimaging and their problems

The limited resolution of current neuroimaging techniques poses a substantial problem for the investigation of the encoding of contents in the human brain, and thus for studies on NCCCs in the human brain. The most important format in which information is coded in the brain is the cortical column (Fujita et al. 1992). Cortical columns consist of small groupings of cells with similar tuning properties, clustered together at a scale of around half a millimetre. Even functional magnetic resonance imaging (fMRI) does not routinely have a sufficient resolution to selectively study the activation of individual cortical columns (but see e.g., Yacoub et al. 2008 for recent progress). For this reason, most research into perceptual contents has relied on experimental “tricks” that allow the tracking of contents indirectly.

In *frequency tagging*, a visual stimulus is tagged with a specific and unique flicker frequency. This then allows for tracing of the processing of this stimulus by searching for brain signals that exhibit the same flicker frequency. This approach has been used to study binocular rivalry, but in quite a different way to that undertaken by Leopold & Logothetis (1996). Tononi et al. (1998) tagged the inputs of the two eyes with different frequencies. They found that the currently dominant percept was accom-

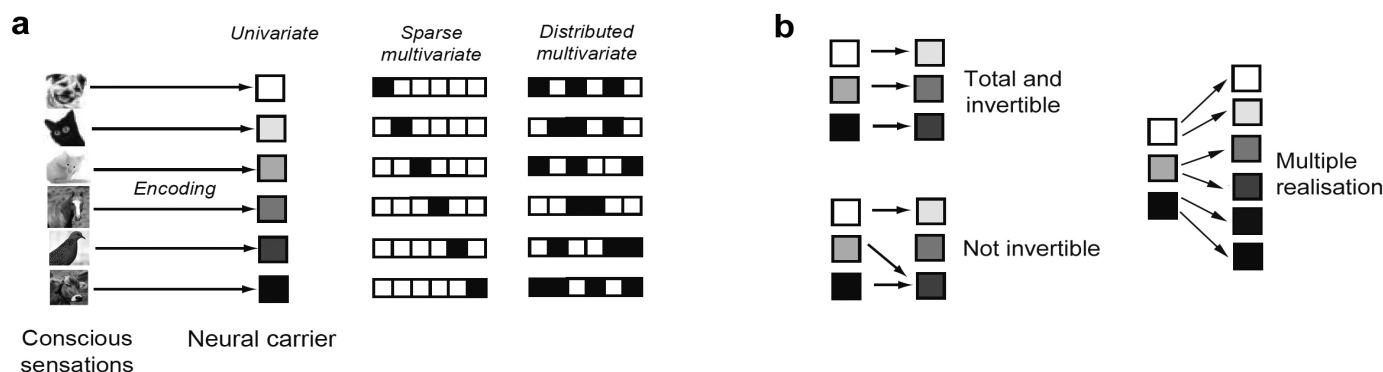


Figure 3: Principles of mapping between mental states and brain states (see text; adapted from Haynes 2009 with additional images from Wikipedia).

panied by wide-spread increases in Magnetoencephalography (MEG)-signals at the tagged frequency across multiple brain regions, mostly in the early visual and temporal cortex. This is a very powerful approach and it reveals how wide-spread the effects are when a stimulus reaches visual awareness. However, it is not always clear whether these findings indicate that the corresponding perceptual features of the stimuli, in this case the orientation of line elements, really are distributed throughout the brain. The key problem is that the feature that is traced (the frequency) is not the main feature that is perceptually relevant (orientation). One could imagine, say, that activity in higher-level brain regions that exhibits the frequency of the dominant stimulus might not be involved in coding the sensory content, but instead in detecting the presence of a change in the visual image, irrespective of what the corresponding feature is. The frequency-tagging approach does not allow for distinguishing between these alternatives.

Another approach to tracking content-selective processing is to use stimuli that are known to activate specific *content-selective brain regions* (Tong et al. 1998; Rees et al. 2000). For example, in a study on binocular rivalry, Tong et al. (1998) used faces and houses as rivalry stimuli. These stimuli are known to activate different brain regions, the fusiform face area (FFA) and the parahippocampal place area (PPA). They found that activity in a content-selective region increased when the corresponding stimulus became perceptually dominant. This goes further than the frequency-tag-

ging approach in that it allows for drawing the plausible conclusion that awareness leads to increased activity in content-selective regions. However, this approach again suffers from several problems. First, it only allows us to address the hypothesis related to very specific stimuli (typically faces and houses) and to very specific brain regions. Because the approach relies on the existence of macroscopic content-selective regions, it would not be possible to test whether, say, the prefrontal cortex, receives sensory information when a stimulus reaches awareness. A further problem is that the high selectivity of FFA and PPA has long been questioned (Haxby et al. 2001).

4 Mapping and decoding

It seems a different, more direct, and generic approach is necessary in order to identify the neural correlates of the contents of consciousness. It may help to start in the simplest possible way. If we want to explain the occurrence of a conscious experience E_1 with the occurrence of a brain state B_1 then—roughly speaking—the experience and the brain state should always happen together. If we want to explain N experiences $E_{1..N}$ with brain states, we will need N different brain states $B_{1..N}$ in order to encode the different experiences. If brain data from a specific area only adopts one of five states every time a participant has one of ten experiences it is impossible to explain the experiences through the different brain states. Ultimately this boils down to a *mapping* problem (Haynes 2009; Figure 3).

A set of conscious sensations, here visual percepts of six different animals, can be encoded in a neural carrier in multiple ways. Three principles are illustrated in [Figure 3a](#). One hypothetical way to code these six animals would be to use a single neuron and to encode the objects by the firing rates of this neuron. One would assign one specific firing rate to each of the different animals, say 1Hz to the dog, 2Hz to the cat and 3Hz to the mouse, etc. This approach is also referred to as a *univariate* code, because it uses only one single parameter of neural activity. It has the advantage of requiring only a single neuron. In *principle* it is possible to encode many different objects with a single neuron. The idea would be very similar to a telephone number, if one thinks of different numbers corresponding to different firing rates. In theory it would be possible to encode every single telephone in the world in this way, provided that the firing rates could be established very precisely and reliably. The disadvantage with this approach—even if firing rates could be established with great precision—is that it can only handle *exclusive* thoughts, i.e. it has no way of dealing with a superposition of different animals, say a cat together with a dog.

A different approach is not to use a single neuron to encode different thoughts, but instead to use a set of neurons to encode a set of thoughts. This population-based approach is also termed “multivariate”. One way to encode thoughts about six different animals would be to assign one specific neuron to the occurrence of each thought. Neuron one, say, might fire when a person thinks about a dog; neuron two would fire when they were thinking about a cat, etc. Here the firing rate is irrelevant; only a threshold is needed, such that one has a way of deciding when a neuron is “active” or “not active”. This specific coding scheme is variably termed “sparse code”, “labelled line code”, “cardinal cell code” or “grandmother cell code” (see e.g., [Quiroga et al. 2008](#)). It has the advantage of being able to handle arbitrary superpositions and combinations of thoughts, say thoughts about a meeting of a dog, a cat, and a mouse. A disadvantage is that a different neuron is needed for the encoding of each new entity. N

neurons can only encode N different thoughts. Given that the average human brain comprises 86 billion neurons ([Azevedo et al. 2009](#)) this might not seem too big a problem. A different way to use a population of neurons to encode a set of thoughts would be a *distributed* multivariate code. Here, each mental state is associated with a single activation pattern in the neural population, but now arbitrary combinations of neurons are possible for the encoding of each single thought. This allows for the encoding of 2^N thoughts with N neurons, if each neuron is only considered to be “on” or “off”.

There are various examples of these different types of codes. The encoding of *intensity* follows a univariate code: The difference between a brighter and a darker image is encoded in a higher versus a lower firing rate of the corresponding neurons in the visual cortex (see e.g., [Haynes 2009](#)). However, to date, I am not aware of any example where different higher-level interpretations of stimuli are coded in a univariate format. There are many examples of labelled line codes. The retinotopic location within the visual field is encoded in a sparse, labelled line format (e.g., [Serenio et al. 1995](#)). One position in the visual field is coded by one set of neurons in the early visual cortex; another position is encoded by a different set of neurons. If two objects appear in the visual field simultaneously, then both of the corresponding sets of neurons become active simultaneously. A similar coding principle is observed for auditory pitch, where different pitches are coded in different cells in the form of a tonotopic map ([Formisano et al. 2003](#)). The somatosensory and motor homunculi are also examples of labelled line codes, each position in the brain corresponding to one specific position in the body ([Penfield & Rasmussen 1950](#)). A distributed multivariate code is, for example, used to code different objects ([Haxby et al. 2001](#)) or different emotions ([Anders et al. 2011](#)).

When identifying the mapping between brain states and mental states one is generally interested in identifying which specific population of neurons is a suitable candidate for explaining a particular class of visual experiences. For this it is possible to formulate a number of

constraints (Haynes 2009). First, the mapping needs to assign one brain state to each mental state in which we are interested. In other words, the mapping has to be total (Figure 3b). This should be easy—it just means that we can assign one measured brain state to each different mental state. Second, the mapping cannot assign the same brain state to two different mental states. Otherwise the brain states would not be able to explain the different mental states. Technically this means the mapping has to be invertible, or injective. Every brain state should be assigned to no more than one mental state. However, it is possible—in the sense of multiple realisation—that multiple brain states are assigned to the same mental state, as long as neither of these brain states co-occurs with other mental states. The brain states referred to here only mean brain states that are relevant for explaining a set of mental states. If we want to explain thoughts about six animals, say, it might not be necessary that brain states in the motor cortex are different for the different animals. However, if one wants to propose one set of neurons (say, those in the lateral occipital complex, Malach et al. 1995) as a candidate for explaining animal experiences, then this can only hold if the abovementioned mapping requirements are fulfilled.

In practice it will be very difficult to establish this mapping directly. One major problem is that we can't measure brain states in sufficient detail with current neuroscience techniques. Non-invasive measurement techniques such as electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) have very limited spatial resolution. fMRI for example resolves the brain with a measurement grid of around 1–3mm, so that each measurement unit (or voxel) contains up to a million cells. And the temporal resolution of fMRI is restricted because fMRI measures the delayed and temporally-extended hemodynamic response to neural stimulation. While it is possible—to some degree—to reconstruct visual experiences from fMRI signals (e.g., Miyawaki et al. 2008), fMRI cannot resolve temporal details of neural processes, such as the synchronized activity of multiple cells. But it is not only EEG and fMRI

that have limited resolution: Invasive recording techniques are typically restricted to individual well-circumscribed locations, where surgery is performed. And even with multielectrodes it is not possible to identify the state of each individual neuron in a piece of living tissue.

Another important limitation lies in our ability to precisely characterize and cognitively penetrate *phenomenal states* (e.g., Raffman 1995). There is currently no psychophysical technique that would allow us to characterize the full details of a person's visual experiences at each location in the visual field. Verbal reports or button presses can convey only a very reduced picture of the true complexity of visual experiences. So ultimately, the mapping requires precision from both psychology and neuroscience, and any imprecision in either approach will blur the mapping and distort the interpretation.

The next best option short of establishing full mapping is to use decoding techniques that follow a similar logic. Brain-based decoding is also referred to as “brain reading” or “multivoxel pattern analysis” (see Haynes & Rees 2005 for a review). The basic idea is to see to which degree it is possible to infer a mental state from a measurement of a brain state. Say you want to test whether the lateral occipital complex is a suitable candidate for encoding visual thoughts about animals. You test if it is possible to infer which animal a person is currently seeing by training a classifier to learn the association between animal and brain activation pattern, and then one needs to test whether the classifier can correctly assign the animal that belongs to a new measurement of brain activity. In the following, this approach will be explained in detail.

Take for example a hypothetical fMRI-measurement of the human brain within a three by three grid of locations, amounting to nine voxels (Figure 4a). These nine voxels can be systematically resorted into a column of numbers (or vectors), where each entry denotes the activation at one location (high values correspond to strong fMRI responses). Say one was interested in testing whether these nine voxels contain information about two different visual

images, perhaps a dog and a cat. The question that needs to be addressed is whether the response patterns (i.e., the vectors) are sufficiently different to allow for distinguishing between the animals, based on these brain activity measurements alone. The vector is not a useful way to see whether this classification is possible. It can help to visualize the same information in a two-dimensional coordinate system. Take the responses to the dog. One can think of the first and second entries in the vector as x- and y-values that define points in a coordinate system. The response in the first voxel (x) to the dog is a low value (2), while the second value (y) is a high value (8). When plotted in a two-dimensional coordinate system (Figure 4b), this yields a point in the top left of the coordinate system, shown here in red. Repeated measurements of the brain response to the dog yield a small cloud of red points. Repeatedly measured brain responses to the cat have high values in voxel 1 (x) and low values in voxel 2 (y). In the two-dimensional coordinate system this yields a cloud of blue points in the bottom right of the coordinate system. Clearly the responses are separable in this two-dimensional coordinate system, so the two animals enjoy reliably separate neural representations in this set of nine voxels. In this hypothetical example, each of the two voxels alone would be sufficiently informative about the category of animal. By collapsing the points for voxel one onto the x-axis it becomes clear that the two distributions of points (red and blue) are sufficiently different to allow for telling the two apart. The same holds for voxel two by collapsing to the y-axis. This is akin to a labelled line code, with one line for “dog” and one line for “cat”.

However, there are cases where the two distributions will not be so easily separable. Figure 4c shows an example where the individual voxels do not have information about the animals. The collapsed or “marginal” distributions largely overlap. There is no way to tell a cat response from a dog response by looking at either voxel one or two alone. However, by taking into account the *joint activity* in both voxels, the two animals become clearly separ-

able. Responses to the dog all cluster to the top left of the diagonal and responses to the cat cluster to its bottom right. This joint consideration of the information contained in multiple voxels is the underlying principle of *multivariate decoding*. The line separating the two distributions is known as the decision boundary. Decision boundaries are not necessarily straight lines. Many other types of distributions of responses are possible. Figure 4d, for example, shows a non-linear decision boundary. Finding the optimal decision boundary is the key objective in the field of machine learning (Müller et al. 2001), where many different types of classifiers have been developed (most well known are support vector classifiers). In order to identify the decision boundary the available data are split into training and test data. The test data are put aside and only the training data are then used to find a decision boundary, as, for example, is shown in Figure 4e. The crucial test is then performed with the remaining test data. The classifier is applied to these data to see to which degree it is able to correctly assign the labels. Depending on which side of the decision boundary a test data point falls upon, it will yield either a correct or an incorrect classification.

Please note the similarity between the mapping of mental states and brain states. The red cloud of points in Figure 4b shows a two-dimensional response pattern that corresponds to the neural code for percepts of dogs. The spread of the point cloud (i.e. the fact that repeated measurements don't yield identical values) could mean two things. Either the spread reflects *noise and uncertainty* that is typically inherent in measurements of neural data. This could, for example, reflect the fact that single fMRI voxels can sample many thousand cells, only few of which might be involved in processing. Additionally, physiological background rhythms can influence the signals and contribute to noise (Fox et al. 2006). Alternatively, however, the spread of the points could also be an inherent property of the representation. This would suggest that every time a person sees or visual imagines a dog, a slightly different activation pattern is observed in the brain. This would then

be evidence of *multiple realization*. Current measurement techniques do not have sufficient precision to distinguish between these two accounts. One difference between the multivariate mapping shown in Figure 3a (right) and the classification shown in Figure 4 is that the classification shows response distributions where each individual variable (voxel, channel) can adopt a graded value, whereas the values in Figure 3a (right) are only binary.

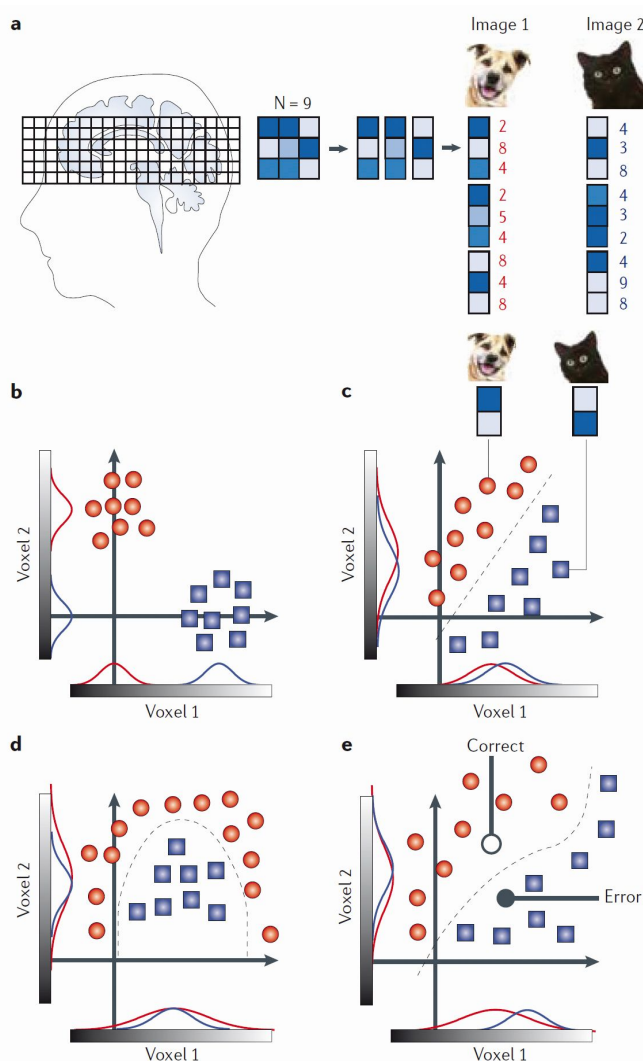


Figure 4: Mental state decoding using classification techniques (image adapted from Haynes & Rees 2005).

5 What does multivariate decoding reveal about NCCCs?

The importance of information theory for understanding the neural correlates of consciousness has been stressed repeatedly, most notably

by Giulio Tononi (2005). His information integration theory focuses on the information-based properties of neural processes and uses these special properties to provide a general explanation of consciousness. In contrast, the multivariate decoding account presented here attempts to solve the much more basic question of which neural populations provide the best account for which visual experiences. As mentioned above, this can be thought of as a search for the core neural correlates of the contents of consciousness (NCCCs), which have been postulated in similar forms by previous authors (Chalmers 2000; Block 2007; Koch 2004). While these proposals for core NCCCs have been influential in theoretical discussions on consciousness, they have only rarely been directly linked to neuroscience research, which requires spelling out how the NCCC can be established in empirical data. In the following, various studies of multivariate decoding from our lab will be presented that have implications for identifying NCCCs.

5.1 Example 1: Encapsulated information in V1

There has long been a debate as to whether the primary visual cortex (V1) is a neural correlate of visual consciousness. Crick & Koch (1995) postulated that V1 does not encode visual experiences for several reasons. First, V1 does not have the anatomical projections to the prefrontal cortex that would allow for a direct read-out of information in V1. This would be required to explain a key distinguishing feature of conscious experiences: that we can voluntarily act upon them. A second reason is that V1 encodes information of which we are not aware. Psychophysical experiments, for example, show that V1 can encode orientation information of which we are not aware (He et al. 1996). We thus directly assessed the link between information encoding in V1 and visual awareness (Haynes & Rees 2005). Specifically, we investigated the effects crossing the threshold to awareness has on the neural coding of simple visual features. Participants viewed oriented “grating” images (Figure 5) and had to tell whether they were tilted to the left or to the right. In one

condition the images were clearly visible, in the other condition they were rendered invisible by rapidly alternating the orientation stimulus with a mask. In this condition participants were not able to tell the difference between the two orientation stimuli (Figure 5).

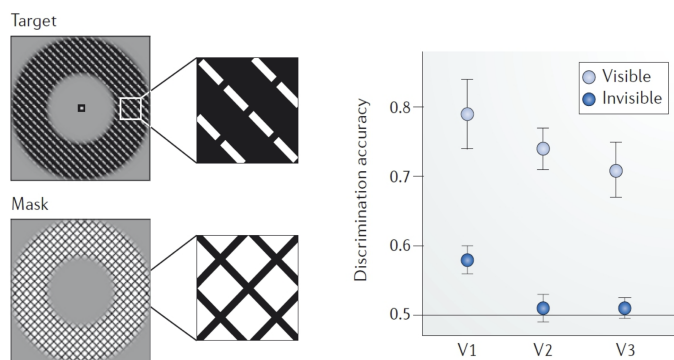


Figure 5: Decoding the orientation of invisible grating stimuli from patterns of activity in early visual areas. Target stimuli were line patterns that were either tilted top left to bottom right, or top right to bottom left. They were rapidly alternated with mask stimuli so that participants were unable to identify the target orientation. The classification accuracy for these “invisible” gratings was above chance in area V1, but not in V2 or V3. For visible orientation stimuli the classification was above chance in all three early visual areas (figure taken from Haynes & Rees 2005).

We then applied a classifier to fMRI-signals from early visual regions V1, V2, and V3 to see if it would be possible to decode the orientation of stimuli. We found that orientation for the visible stimuli could be decoded from all early visual regions, V1, V2, and V3 (Figure 5, right). This is in line with previous research on encoding of orientation information in early visual areas (Bartfeld & Grinvald 1992). Interestingly, we were able to decode the orientation from V1 even for invisible stimuli. This means that V1 presumably continues to carry low-level feature information even when a participant can’t access this information. V2 and V3, however, only had information for visible stimuli, not for invisible stimuli. Please note that an alternative interpretation could be that subjects perceive the subtle differences between masked stimuli, but they cannot report or reason about them. However, in psychophysics an absence of

discriminability is typically considered a strong criterion for absence of awareness. This finding is interesting for several reasons. First, it demonstrates that information can be encapsulated in a person’s early visual cortex, without them being able to access this information. This suggests that V1 is not an NCCC for conscious orientation perception. Second, it shows that one explanation why stimuli are rendered invisible by visual masking is that the information that is available at early stages of processing (V1) is not passed on to the next stages of processing in V2 and V3. Similar encapsulation of information has also been observed for parietal extinction patients in higher-level visual areas with more conventional neuroimaging approaches (Rees et al. 2000).

5.2 Example 2: Imagery and perception

There has also been a long debate on the neural mechanisms underlying visual imagery. One important question is whether the NCCCs underlying imagery are the same—or at least overlapping—with those for veridical perception. One study (Kosslyn et al. 1995) found that imagery activated even very early stages of the visual cortex. This fits with a mechanism that encodes visual images as a replay of representations of veridical percepts. However, this does not reveal whether the activation of the early visual cortex really participates in encoding the imagined contents. Instead, these regions might be involved in ensuring the correct spatial distribution of attention across the visual field (Tootell et al. 1998). The question of whether the neural representations of veridical percepts are the same as those for visual imaginations needs to be established in addition.

We conducted a study to directly address the overlap of NCCCs for veridical perception and imagery (Cichy et al. 2012). Participants were positioned inside an MRI scanner and had to perform one of two tasks: Either they were asked to *observe* visual images presented to the left or right of fixation (Figure 6), or they were asked to imagine visual images in the same locations. Twelve different images from four categories were used: three objects, three visual

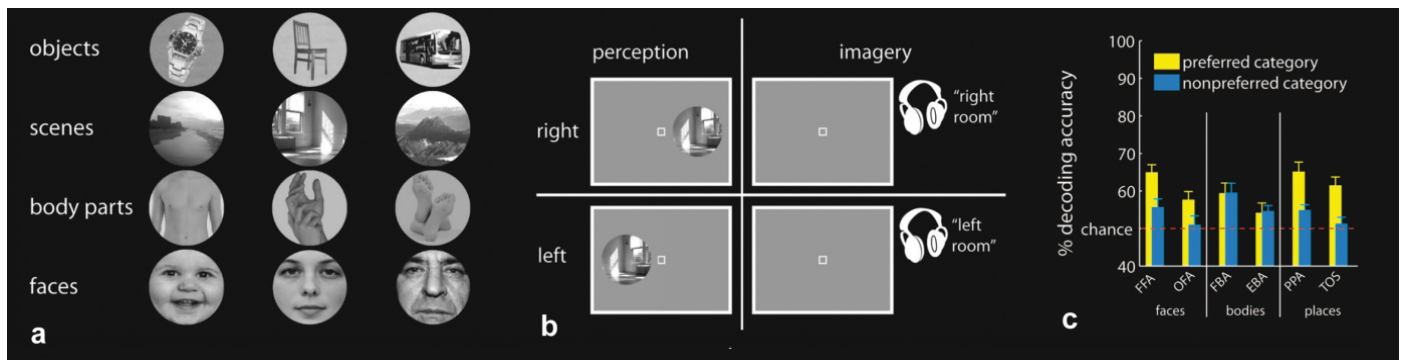


Figure 6: Visual imagery. (a) Visual stimuli used in the experiment consisted of three selections from four categories. (b) In different trials participants either saw the images to the left or right of fixation or they received an auditory instruction to imagine a specific image. (c) A classifier trained on the brain responses of different imagined images could be used to correctly cross-classify which image a person was currently seeing on the screen in the perception condition. Information was higher for the images “preferred” by a visual area, but there was still information, esp. in FBA, about the non-preferred categories (FFA=fusiform face area; OFA=occipital face area; FBA=fusiform body area; EBA=extrastriate body area; PPA=parahippocampal place area; TOS=transverse occipital sulcus)(figure from Cichy et al. 2012).

scenes, three body parts, and three faces. We found that multiple higher-level visual regions had information about the images. Furthermore, it was possible to decode seen visual images using a classifier that had only been trained on imagined visual images. This suggests that imagery and veridical perception share *similar* neural representations for perceptual contents, at least in high-level visual regions. Please note, however, that the cross-classification between veridical perception and imagery is not perfect. It is currently unclear whether this reflects imperfections in the measurement of brain signals with fMRI, or whether it reflects residual differences in the contents of consciousness between imagery and veridical perception, for example the higher vividness of perception based on external visual stimuli (Perkey 1910).

5.3 Example 3: Perceptual learning

Another interesting riddle of sensory awareness is perceptual learning (Sagi 2011; see also Lamme this collection). When we are first exposed to a novel class of sensory stimuli our ability to differentiate between nuances is highly limited. When one tastes the first glass of wine, all wines taste the same. But with increased exposure and experience we learn to distinguish even subtle differences between different wines.

The interesting question here is whether the sensory information was there all along, and we just failed to notice it, or whether the sensory representation of the wines actually improves (see Dennett 1991).

We addressed this question, but with visual grating stimuli instead of different wines (Kahnt et al. 2011). Participants performed two fMRI sessions, where they had to distinguish small differences in the orientation of lines presented on the screen. They had to tell whether they were rotated clockwise or counter-clockwise with respect to a template. During the first fMRI session their ability to distinguish between the line patterns was quite limited. Afterwards we trained them in two sessions outside the MRI scanner on the same line patterns, and their performance continually improved. In a final second fMRI session they had then substantially improved their ability to tell even subtle differences between the orientations apart. But what explains this improvement: Better sensory coding, or better interpretation of the information that was there all along?

To address this question we first looked into the responses in the early visual cortex to the different line stimuli. As expected from our above-mentioned study on orientation coding (Haynes & Rees 2005), it was possible to decode the orientation of the line elements from signals in early

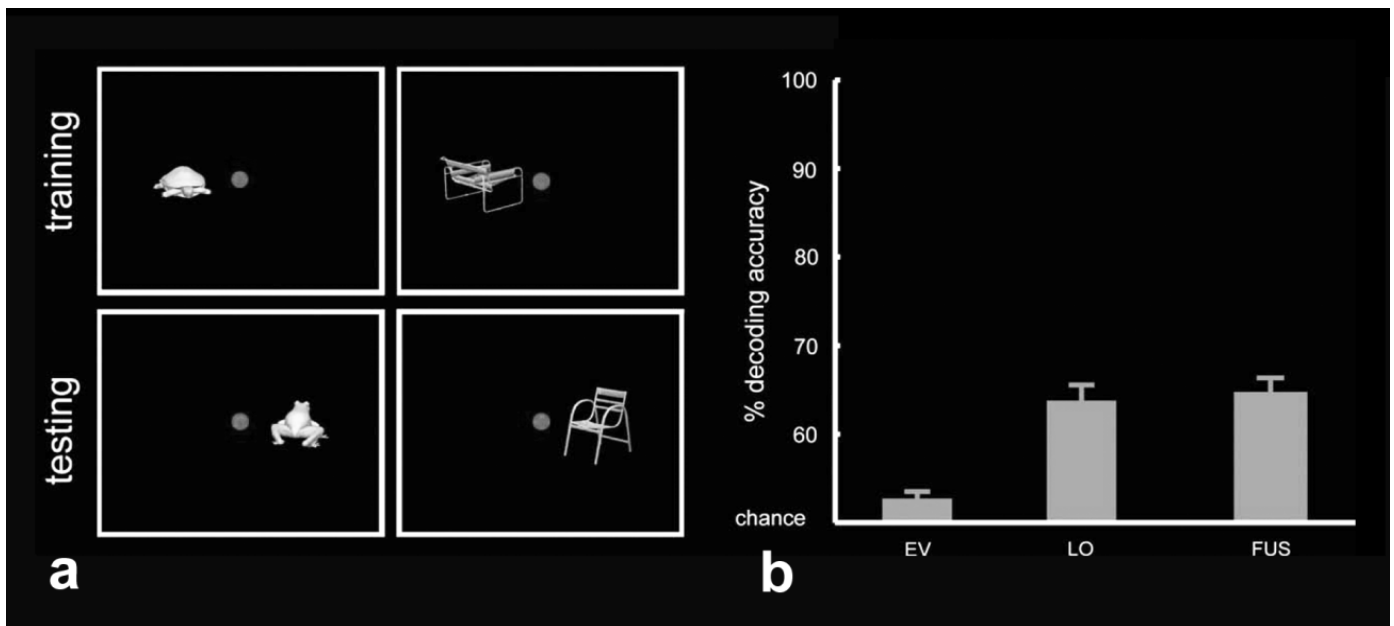


Figure 7: FMRI evidence for invariance of object-representations in the high-level visual regions lateral occipital (LO) and fusiform gyrus (FUS) as compared to early visual cortex (EV; figure from Cichy et al. 2011).

visual areas. It is well established that these areas have information about such simple visual features (Bartfeld & Grinvald 1992). However, we found no improvement in our ability to decode the orientation of the stimuli with learning. There is some divergence in the literature with some studies finding effects of learning in early sensory areas (see Sasaki et al. 2010). Other recent findings in monkeys are in line with our findings and do not find improved information coding in sensory areas (e.g., Law & Gold 2009). In our case, it seems as if the sensory representation of orientation remains unchanged and that some other mechanism has to be responsible for the improvement in perceptual discrimination. We found a region in the medial prefrontal cortex where signals followed the learning curve, thus suggesting that the improvement was not so much a question of stimulus coding but of the *read-out* of information from the sensory system. This study suggests that representation of a feature in an NCCC might not automatically guarantee it enters visual awareness.

5.4 Example 4: Invariance in human higher-level visual cortex

As mentioned above, one important challenge to the idea that the contents of visual awareness

are encoded exclusively late in the visual system is the invariance of responses to low-level visual features (Sáry et al. 1993). We directly investigated the invariance of fMRI responses in the regions lateral occipital (LO) and fusiform gyrus (FUS) of the higher-level object-selective visual cortex (Malach et al. 1995; Grill-Spector et al. 2001). In this study (Cichy et al. 2011) participants viewed objects presented either to the left or the right of the fixation spot (Figure 7). These objects consisted of three different exemplars from four different categories (animals, planes, cars, and chairs). For example, the category “animal” contained images of a frog, a tortoise, and a cow. With these data we were able to explore two different aspects of invariance. First, we wanted to know whether object representations are invariant to changes in spatial location. This is important because a low-level visual representation that focuses exclusively on the distribution of light in the visual field would not be able to generalize from one position to another. So we assessed whether a classifier trained to recognize an object at one position in the visual field would be able to generalize to another position in the visual field. We found that a classifier was able to generalize to a different position, however with reduced accuracy. This indicates that the representations

were at least partially invariant with respect to low-level visual features. Next, we investigated whether the representations would generalize from one exemplar to another. This goes even further in testing for the level of abstraction of the representation. A classifier that can generalize not only to a different location but even to a different exemplar (say from a frog to a cow) needs to operate at a higher level of abstraction that is largely independent from low-level visual features. Again we found that the classifier was able to generalize between exemplars of the same category, further supporting the abstraction of representations in the higher visual regions LO and FUS (Figure 7). This makes it again less plausible that the contents of visual awareness are encoded exclusively in the higher-level visual cortex. Encoding in these regions is invariant (or at least tolerant) to low-level feature changes, and thus this level of perceptual experience has to be encoded at a different, presumably lower, level of visual processing.

5.5 Example 5: No sensory information in PFC

A further case where multivariate decoding might inform theories of visual awareness becomes apparent when we confront the question of whether sensory information is distributed throughout the brain when a stimulus crosses the threshold of awareness. The global neuronal workspace theory (e.g., Dehaene & Naccache 2001; see also Baars 2002) posits that sensory signals are made globally *available* across large-scale brain networks, especially in the prefrontal and parietal cortices, when they reach awareness. An interesting and open question is whether this global availability of sensory information means that the sensory information about a stimulus can be actually decoded from these prefrontal and parietal brain regions to which the information is made available. In theory, one might be able to distinguish between a “streaming model” of global availability, where information is broadcast throughout the brain (e.g., Baars 1988), and which should thus be decodable from multiple brain regions; an alternative would be an “on demand” model of global

availability, where sensory signals are only propagated into prefrontal and parietal cortex when selected by attention (e.g., Dehaene & Naccache 2001).

We performed three fMRI studies to test this question (Bode et al. 2012; Bode et al. 2013; Hebart et al. 2012). In the first study (Bode et al. 2012), participants were briefly shown images of pianos and chairs that were temporally embedded in scrambled mask stimuli. There were two conditions. In one condition, the timing of visual stimuli was chosen such that the target stimuli were clearly visible. In the other condition, the timing of scrambled masks and targets was such that the targets were effectively rendered invisible. We attempted to decode the sensory information about the presented objects. Under high visibility we were able to decode which image was being shown from fMRI signals in the so-called lateral occipital regions of the human brain, where complex object recognition takes place. Under low visibility, there was no information in these brain regions. This suggests a possible mechanism for explaining why the stimuli failed to reach awareness. Presumably their sensory representations were already cancelled out at the visual processing stages. The “streaming model” mentioned above would mean that sensory information about the object category is distributed into parietal and prefrontal brain regions when the stimulus crosses the threshold of awareness. However, we found no information in the prefrontal cortex—under either high or low visibility (Bode et al. 2012). This finding was repeated in two different studies, one also using objects as stimuli (Bode et al. 2013) and one using drifting motion stimuli (Hebart et al. 2012). In contrast, in animal studies sensory information has been found in the prefrontal cortex (Pasternak & Greenlee 2005). It is currently unclear whether this reflects a species-difference or whether it is due to limitations in the resolution of human neuroimaging techniques.

5.6 Example 6: Unconscious processing of preferences

It is well known that unattended and even invisible visual stimuli can undergo substantial processing. We investigated whether informa-

tion about high-level, more interpretative and subjective properties of visual stimuli would also be traceable using decoding techniques. For this we aimed to decode the degree to which *preferences* for certain visually presented images of cars can be decoded, even when these stimuli were unattended and were not task-relevant (Tusche et al. 2010).

For this experiment we carefully pre-selected our participants, who were self-reporting car-enthusiasts. Then we ensured that we chose stimuli where different participants had maximally-divergent opinions as to which car they preferred. This was necessary in order to de-correlate the classification of the preference from the classification of the specific vehicles. Subjects were divided into two groups. Participants from the first group were presented with the car images in the scanner and had to actively evaluate whether they liked them. The second group was also presented with the car images, but they were distracted from them. They were required to solve a very difficult task that required them to focus their attention elsewhere in the visual field, on fixation. The car stimuli were thus task-irrelevant and presented outside of the attentional focus. This group of subjects could not recall which cars had been shown during the experiment, suggesting that they were indeed not actively deliberating about the cars. After the experiment, participants from both groups were asked to rate how much they would like to buy each car. This served as a gold standard for their preference.

We then tried to decode whether individual subjects liked the cars or not. For this, we looked into patterns of brain activity throughout the brain, to see where there might be information regarding preferences. This was done in order to reduce the bias when only looking into pre-specified brain regions. We found that it was possible to decode the preferred cars with 75% accuracy from brain regions far outside the visual system, in the medial prefrontal and in the insular cortex. This was true for the subjects who had been actively deliberating about their preferences for the cars, but also for the participants who been distracted from thinking about them. Presumably, this

means that the brain automatically processes the car-images all the way up to the stage of encoding preferences, even in the absence of visual attention. Please note that this finding of *preference* information in the prefrontal cortex is quite different to that in the previous experiment, where there was no *sensory* information in PFC. Here, in contrast, there is information in PFC, but (a) not about a sensory property and (b) even for unattended stimuli. Thus, it appears that the informational dividing line between sensory and prefrontal brain regions is not one of awareness, but rather one of the type of information coded.

6 Complicating factors

When searching for the neural correlates of contents of consciousness (NCCCs), there are several complicating factors. One might a priori have an assumption of modularity, meaning that one feature is encoded in one dedicated NCCC area. The idea that such single, maximally-informative regions exist for different features, however, is no more than an assumption. It might turn out that perceptual coding—even for single features—inherently involves processes in multiple brain regions.

Empirically, it is known that information about objects is distributed across multiple regions. One question is whether one brain region can have information about more than one content (e.g., Haxby et al. 2001; Cichy et al. 2013). In a study inspired by Haxby et al. (2001) we investigated whether object-selective brain regions have information only about objects from their preferred category (Cichy et al. 2013). Participants viewed images from four different categories: objects, visual scenes, body parts, and faces. These categories were chosen because faces, body parts, and places are believed to be processed by highly selective brain regions. We found that a classifier not only contained information about a region's preferred category: take the example of the face-selective region FFA. It was not only possible to classify faces from this region, it was also possible to classify the difference between other, non-face-related objects, say between a chair and a window. The

flipside of this finding that individual regions encode multiple contents is that individual perceptual contents were found in multiple regions. For example, information about faces was also found in supposedly “non-face-selective” brain regions (e.g., in the PPA). This presents a challenge to the idea that each content is represented in one region only.

However, the problem might not be as severe as it first appears. It is actually expected that multiple regions will contain information about each type of content. Different brain regions do not exist in isolation, but are densely causally interconnected (Felleman & van Essen 1991). Furthermore, in the visual pathway, stimulus-related information will reach higher-level brain regions by way of low-level regions. Even if the FFA is the visual region that responds most (albeit not fully) selectively to faces, the presence of a face could also be inferred from the discharge pattern of ganglion cells in the retina. Thus, vertical and horizontal propagation of information is expected. One crucial criterion, which has not received much attention, is whether the information in different regions is *redundant* or whether it is *independent* with respect to a person’s perceptual experience. If one hypothetical brain region, say the uniform unicorn area (UUA), is directly responsible for visual experiences of unicorns, it should have more information about a person’s unicorn experiences than any other region.

The relationship between information in the UUA and other areas will reveal a lot about the nature of representation. If other regions also have information about unicorn experiences, and they receive their information about unicorn experiences via the UUA, then the unicorn-related information in the other regions should be partially redundant to that in the UUA. A classifier should not be able to extract more information about a person’s unicorn experiences by additionally taking other regions into account, over and above the information available from the UUA. If, in contrast, other regions have information that goes beyond that in the UUA that allow the system to improve the classification of unicorn experiences, it is likely that the representation itself is distrib-

uted across multiple brain regions. Another way to put it is to distinguish between representational and causal entanglement. A change in neural activity in one region will typically be propagated to any neighbouring regions with which it is connected. This *causal* entanglement, however, does not directly implicate *representational* entanglement. Only if it were not possible to find an individual region where neural activity patterns is not fully informative of a specific feature, and if taking into account the joint activity of this region and another region did provide full information, would this provide evidence for representational entanglement.

7 Putting it together

As outlined above, when attempting to identify the neural correlate of a particular content of conscious experience, it is important to ensure that brain representations in any candidate region fulfil certain mapping requirements. Because we have no direct way of establishing this mapping, multivariate decoding provides a rough approximation that allows the linking of perceptual contents to population brain responses in different regions, and allows us to explore their properties. The data from our lab provide several constraints for a theory of NCCs. Consistent with previous suggestions (Crick & Koch 1995), the very early stages of processing in V1 are presumably not directly involved in encoding visual experiences. Representations in these regions have more detail than enters consciousness (Haynes & Rees 2005) and might not change their information content during perceptual learning when contents are successively represented with more detail in consciousness (Kahnt et al. 2011). Please note that early regions beyond V1 have to be NCCs, because higher-level visual areas are invariant to low-level visual features. This has not only been shown in animals (Sáry et al. 1993), but also in humans using classification techniques (e.g., Cichy et al. 2011). This invariance means that high-level regions cannot simultaneously encode the high-level, more abstract phenomenal properties (such as whether a cloud of points resembles a dog or a cat) and the low-level phe-

nomenal properties (colour or brightness sensations). Multiple regions are needed to account for the full multilevel nature of our perceptual experience. While V1 is presumably excluded from visual awareness, early extrastriate regions (such as V2) are likely to be involved, because they still encode low-level visual information. They also appear to filter out sensory information that does not enter awareness, thus again closely matching perceptual experience. For example, V2 and V3 do not encode the orientation of invisible lines, whereas V1 does (Haynes & Rees 2005). Similarly, neural object representations in the lateral occipital complex were wiped out by visual masking that rendered an object stimulus invisible (Bode et al. 2012). The role of extrastriate and higher-level visual areas in visual awareness is further highlighted by the fact that they exhibit a certain convergence of different aspects of awareness. Most notably, they employ a shared code for visual perception and visual imagery (Cichy et al. 2012).

While extrastriate and higher-level visual regions jointly encode different feature levels of visual awareness, there is evidence that a representation in these regions is not sufficient for visual awareness. For example, our experiments on perceptual learning (Kahnt et al. 2011)—where subjects are unable to access certain details of visual stimuli—show that improved sensory perception is not necessarily associated with improved representation of information in these early areas. The mechanism through which perception of details might be improved lies beyond the sensory encoding stage, in the prefrontal cortex. The mechanism of this improvement is not an improved sensory representation in the prefrontal cortex. Contrary to several experiments on animals (Pasternak & Greenlee 2005), our experiments consistently fail to show any sensory information in the frontal cortex. For example, when a stimulus survives visual masking and is consciously perceived, there is no evidence for the additional distribution of information into the prefrontal cortex (Bode et al. 2012; Bode et al. 2013; Hebart et al. 2012) as would be expected if information is indeed made globally available in the sense of a “streaming model” of a global

workspace (Dehaene & Naccache 2001). Even in a more conventional experimental task, based on visual working memory, we were not able to identify sensory information in the prefrontal cortex. Thus, the direct encoding of the visual contents of consciousness, the NCCCs appear to lie in sensory brain regions, at least as far as can be told with the resolution of non-invasive human neuroimaging techniques. On the other hand, our results suggest that the prefrontal cortex is involved in the decision-making—as has been suggested before (Heekeren et al. 2004)—and in learning about sensory contents (Kahnt et al. 2011). Thus, it appears to do so without re-representing or encoding sensory information itself.

References

- Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T. & Haynes, J. D. (2011). Flow of affective information between communicating brains. *NeuroImage*, *54* (1), 439-446. [10.1016/j.neuroimage.2010.07.004](https://doi.org/10.1016/j.neuroimage.2010.07.004)
- Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., Filho, W. S., Lent, R. & Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, *513* (5), 532-541. [10.1002/cne.21974](https://doi.org/10.1002/cne.21974)
- Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge, UK: Cambridge University Press.
- (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, *6* (1), 47-52. [10.1016/s1364-6613\(00\)01819-2](https://doi.org/10.1016/s1364-6613(00)01819-2)
- Bartfeld, E. & Grinvald, A. (1992). Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *89* (24), 11905-11909. [10.1073/pnas.89.24.11905](https://doi.org/10.1073/pnas.89.24.11905)
- Block, N. (2007). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30* (5-6), 481-499. [10.1017/S0140525X07002786](https://doi.org/10.1017/S0140525X07002786)
- Bode, S., Bogler, C., Soon, C. S. & Haynes, J. D. (2012). The neural encoding of guesses in the human brain. *NeuroImage*, *59* (2), 1924-1931. [10.1016/j.neuroimage.2011.08.106](https://doi.org/10.1016/j.neuroimage.2011.08.106)
- Bode, S., Bogler, C. & Haynes, J. D. (2013). Similar neural mechanisms for perceptual guesses and free decisions. *NeuroImage*, *65*, 456-465. [10.1016/j.neuroimage.2012.09.064](https://doi.org/10.1016/j.neuroimage.2012.09.064)
- Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.) *Neural Correlates of Consciousness: Conceptual and Empirical Questions* (pp. 17-40). Cambridge, MA: MIT Press.
- Cichy, R. M., Chen, Y. & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *NeuroImage*, *54* (3), 2297-2307. [10.1016/j.neuroimage.2010.09.044](https://doi.org/10.1016/j.neuroimage.2010.09.044)
- Cichy, R. M., Heinzle, J. & Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex*, *22* (2), 372-380. [10.1093/cercor/bhr106](https://doi.org/10.1093/cercor/bhr106)
- Cichy, R. M., Sterzer, P., Heinzle, J., Elliot, L. T., Ramirez, F. & Haynes, J. D. (2013). Probing principles of large-scale object representation: category preference and location encoding. *Human Brain Mapping*, *34* (7), 1636-1651. [10.1002/hbm.22020](https://doi.org/10.1002/hbm.22020)
- Crick, F. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375* (6527), 121-123. [10.1038/375121a0](https://doi.org/10.1038/375121a0)
- Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79* (1-2), 1-37. [10.1016/S0010-0277\(00\)00123-2](https://doi.org/10.1016/S0010-0277(00)00123-2)
- Dennett, D. (1991). *Consciousness explained*. Boston, MA: Penguin.
- Engel, A. K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5* (1), 16-25. [10.1016/S1364-6613\(00\)01568-0](https://doi.org/10.1016/S1364-6613(00)01568-0)
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1* (1), 1-47. [10.1093/cercor/1.1.1](https://doi.org/10.1093/cercor/1.1.1)
- Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K. & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40* (4), 859-869. [10.1016/S0896-6273\(03\)00669-X](https://doi.org/10.1016/S0896-6273(03)00669-X)
- Fox, M. D., Snyder, A. Z., Zacks, J. M. & Raichle, M. E. (2006). Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. *Nature Neuroscience*, *9* (1), 23-25. [10.1038/nn1616](https://doi.org/10.1038/nn1616)
- Fujita, I., Tanaka, K., Ito, M. & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, *360* (6402), 343-346. [10.1038/360343a0](https://doi.org/10.1038/360343a0)
- Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41* (10-11), 1409-1422. [10.1016/S0042-6989\(01\)00073-6](https://doi.org/10.1016/S0042-6989(01)00073-6)
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L. & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293* (5539), 2425-2430. [10.1126/science.1063736](https://doi.org/10.1126/science.1063736)
- Haynes, J. D. (2009). Decoding visual consciousness from human brain signals. *Trends in Cognitive Sciences*, *13* (5), 194-202. [10.1016/j.tics.2009.02.004](https://doi.org/10.1016/j.tics.2009.02.004)
- Haynes, J. D. & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary

- visual cortex. *Nature Neuroscience*, 8 (5), 686-691. [10.1038/mm1445](https://doi.org/10.1038/mm1445)
- (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, 7 (7), 523-534. [10.1038/nrn193](https://doi.org/10.1038/nrn193)
- Hebart, M. N., Donner, T. H. & Haynes, J. D. (2012). Human visual and parietal cortex encode visual choices independent of motor plans. *NeuroImage*, 63 (3), 1393-1403. [10.1016/j.neuroimage.2012.08.027](https://doi.org/10.1016/j.neuroimage.2012.08.027)
- Heekeren, H. R., Marrett, S., Bandettini, P. A. & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, 431 (7010), 859-862. [10.1038/nature02966](https://doi.org/10.1038/nature02966)
- He, S., Cavanagh, P. & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383 (6598), 334-337. [10.1038/383334a0](https://doi.org/10.1038/383334a0)
- Kahnt, T., Grueschow, M., Speck, O. & Haynes, J. D. (2011). Perceptual learning and decision-making in human medial frontal cortex. *Neuron*, 70 (3), 549-559. [10.1016/j.neuron.2011.02.054](https://doi.org/10.1016/j.neuron.2011.02.054)
- Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CL: Roberts & Company.
- Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, 378 (6556), 496-498. [10.1038/378496a0](https://doi.org/10.1038/378496a0)
- Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10 (11), 494-501. [10.1016/j.tics.2006.09.001](https://doi.org/10.1016/j.tics.2006.09.001)
- (2015). The crack of dawn: Perceptual functions and neural mechanisms that mark the transition from unconscious processing to conscious vision. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Law, C. T. & Gold, J. I. (2009). Reinforcement learning can account for associative and perceptual learning on a visual-decision task. *Nature Neuroscience*, 12 (5), 655-663. [10.1038/nn.2304](https://doi.org/10.1038/nn.2304)
- Leopold, D. A. & Logothetis, N. K. (1996). Logothetis N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, 379 (6565), 549-553.
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R. & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 92 (18), 8135-8139.
- Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A., Morito, Y., Tanabe, H. C., Sadato, N. & Kamitani, Y. (2008). Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. *Neuron*, 60 (5), 915-925. [10.1016/j.neuron.2008.11.004](https://doi.org/10.1016/j.neuron.2008.11.004)
- Müller, K. R., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, 12 (2), 181-201. [10.1109/72.914517](https://doi.org/10.1109/72.914517)
- Pascual-Leone, A. & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292 (5516), 510-512. [10.1126/science.1057099](https://doi.org/10.1126/science.1057099)
- Pasternak, T. & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6 (3), 97-107. [10.1038/nrn1637](https://doi.org/10.1038/nrn1637)
- Penfield, W. & Rasmussen, T. (1950). *The cerebral cortex of man*. New York, NY: Macmillan.
- Perkey, C. (1910). An experimental study of imagination. *American Journal of Psychology*, 21 (3), 422-452. [10.1037/h0041622](https://doi.org/10.1037/h0041622)
- Quiroga, R. Q., Kreiman, G., Koch, C. & Fried, I. (2008). Sparse but not 'grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences*, 23 (3), 87-91. [10.1016/j.tics.2007.12.003](https://doi.org/10.1016/j.tics.2007.12.003)
- Raffman, D. (1995). On the persistence of phenomenology. In T. Metzinger (Ed.) *Conscious experience* (pp. 293-308). Paderborn, GER: Schöningh Verlag.
- Rees, G., Wojciulik, E., Clarke, K., Husain, M., Frith, C. & Driver, J. (2000). Unconscious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain*, 23 (8), 1624-1633. [10.1093/brain/123.8.1624](https://doi.org/10.1093/brain/123.8.1624)
- Sagi, D. (2011). Perceptual learning in vision research. *Vision Research*, 51 (13), 1552-1566. [10.1016/j.visres.2010.10.019](https://doi.org/10.1016/j.visres.2010.10.019)
- Sasaki, Y., Nanez, J. E. & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews Neuroscience*, 11 (1), 53-60. [10.1038/nrn2737](https://doi.org/10.1038/nrn2737)
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., Rosen, B. R. & Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268 (5212), 889-893. [10.1126/science.7754376](https://doi.org/10.1126/science.7754376)
- Sheinberg, D. L. & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization.

- Proceedings of the National Academy of Sciences of the United States of America*, 94 (7), 3408-3413.
- Singer, W. (2015). The ongoing search for the neuronal correlate of consciousness. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Sáry, G., Vogels, R. & Orban, G. A. (1993). Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, 260 (5110), 995-997.
- Tong, F., Nakayama, K., Vaughan, J. T. & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21 (4), 753-759.
[10.1016/S0896-6273\(00\)80592-9](https://doi.org/10.1016/S0896-6273(00)80592-9)
- Tononi, G. (2005). Consciousness, information integration, and the brain. *Progress in Brain Research*, 150, 109-126.
- Tononi, G., Srinivasan, R., Russell, D. P. & Edelman, G. M. (1998). Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (6), 3198-3203.
- Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T. & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron*, 21 (6), 1409-1422.
[10.1002/\(SICI\)1097-0193\(1997\)5:4<280::AID-HBM13>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0193(1997)5:4<280::AID-HBM13>3.0.CO;2-I)
- Tusche, A., Bode, S. & Haynes, J. D. (2010). Neural responses to unattended products predict later consumer choices. *Journal of Neuroscience*, 30 (23), 8024-8031.
[10.1523/JNEUROSCI.0064-10.2010](https://doi.org/10.1523/JNEUROSCI.0064-10.2010)
- Yacoub, E., Harel, N. & Ugurbil, K. (2008). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (30), 10607-10612.
[10.1073/pnas.0804110105](https://doi.org/10.1073/pnas.0804110105)

What's up with Prefrontal Cortex?

A Commentary on John-Dylan Haynes

Caspar M. Schwiedrzik

The prefrontal cortex is perhaps one of the most intriguing areas of the brain, and considered by many to be involved in a whole battery of higher cognitive functions. However, evidence for a direct involvement in conscious perception, although often postulated, remains inconclusive. In his paper, John-Dylan Haynes presents results from experiments using multivariate decoding techniques on human functional magnetic resonance imaging data that speak against the assertion that prefrontal cortex broadcasts the contents of consciousness throughout the brain. I consider potential reasons for these null results, as well as where else we may look for the neural correlates of consciousness. Specifically, I propose that conscious perception arises when distributed neurons are bound into coherent assemblies—a process that does not require relay through specific brain areas.

Keywords

Multivariate pattern analysis | Neuronal correlates of consciousness | Neuronal synchrony | Prefrontal cortex

Commentator

Caspar Schwiedrzik

cschwiedrz@rockefeller.edu

The Rockefeller University
New York, NY, U.S.A.

Target Author

John-Dylan Haynes

haynes@bccn-berlin.de

Charité – Universitätsmedizin Berlin
Berlin, Germany

Editors

Thomas Metzinger

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

Jennifer M. Windt

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

1 Introduction

There is a striking parallel between the hierarchical organization of behavior and the hierarchical organization of the cerebral cortex (Botvinick 2008). It is thus tempting to assign consciousness, at least historically often considered to be one of our highest functions (Jackendoff 1987; Markowitsch 1995), to the prefrontal cortex (PFC), which is positioned at the top of the cortical hierarchy. While the idea that consciousness can be localized to a single brain area has now been discredited, many current theories of consciousness still consider the PFC a key player in the emergence of conscious perception (Dehaene & Changeux 2011; Lau & Rosenthal 2011). And indeed, a multitude of

neuroimaging studies has shown differential activation for perceived vs. unperceived stimuli in various parts of the PFC (Dehaene et al. 2001; Lau & Passingham 2006; Sahraie et al. 1997; Schwiedrzik et al. 2014). A very prominent theoretical proposal on the neural correlates of consciousness, the Global Neuronal Workspace (GNW) model by Stanislas Dehaene and colleagues, posits that the PFC (in conjunction with parietal cortex) serves to distribute information that is processed in unconscious modules to the entire brain, and that it is this broadcasting of information that gives rise to conscious experience (Dehaene & Changeux 2011). The PFC may be particularly well equipped to

do so, for example because it hosts an abundance of neurons with long-distance connections, so called “von Economo” neurons, which seem ideally suited to both receive and deliver information from all areas of the brain to all areas of the brain (Dehaene & Changeux 2011). A prediction that can be directly derived from this account and that has been eloquently put forward by John-Dylan Haynes is that the PFC should at least temporarily represent the information that we consciously perceive, i.e., it should directly encode the contents of consciousness (Haynes 2009; this collection). To test this idea, Haynes and his coworkers have used a neuroimaging technique that allows for exquisite access to perceptual content, namely multivariate pattern analysis (MVPA) of functional magnetic resonance imaging (fMRI) signals. In this technique, powerful machine learning algorithms are used to analyze spatially-distributed patterns of brain activity, and a brain region is said to represent the content of interest if its activity patterns allow the reliable classification—in the case of consciousness—of which stimulus the subject perceived on a given trial. This contrasts with previous fMRI studies not using MVPA: because these studies do not directly address content, activity in the PFC (and other regions) that differentiates perceived from not perceived trials could in principle reflect other aspects of conscious experience, for example the allocation of attentional resources or working memory. The stunning result of Haynes’ investigations is that while MVPA shows that perceptual content can be decoded from higher sensory areas, PFC activity does not yield decoding accuracies higher than chance level. So what’s up with PFC?

2 Neuronal representations in the PFC

Indeed, the inability to decode perceptual content from the PFC runs counter intuitions about PFC functions we have from animal models such as the macaque monkey, where representations of (perceptual) content can be even more directly assessed than with MVPA, by using electrophysiological recordings from single neurons. These studies show that PFC neurons

are tuned for and thus represent perceptual features such as visual motion direction (Zaksas & Pasternak 2006) or somatosensory flutter frequency (Romo et al. 1999). Even more direct evidence for the representation of perceptual content in the PFC comes from a recent study by Theofanis Panagiotaropoulos et al. (2012), which shows that single PFC neurons exhibit stimulus-specific activity modulations as a function of subjective perception under flash suppression, a technique that can render visual stimuli temporarily invisible.

In the absence of direct electrophysiological recordings from human PFC, one possible explanation for this discrepancy is that the macaque brain is organized in a totally different way to the human brain. But while theoretically possible, this seems highly unlikely (Passingham 2009; Roelfsema & Treue 2014). Alternatively, one may consider whether certain properties of the neural representations in the PFC may pose limitations to the ability of the fMRI MVPA technique to decode their content. This is in light of the fact that decoding of content from human PFC has been unsuccessful not only in the context of conscious perception, but also in the context of working memory, which has led to a radical reinterpretation of the role of the PFC in this domain (Sreenivasan et al. 2014).

It has been hypothesized that successful decoding of stimulus features such as orientation or motion direction from sensory areas relies upon the presence of orderly spatial arrangements of these features in cortical columns or maps (Freeman et al. 2011; Kamitani & Tong 2005). It is thus worth asking the question whether PFC exhibits a similar map-like structure, or whether the spatial arrangement of features in the PFC already renders the likelihood of decoding any kind of information from its fMRI activity unlikely. For example, while maps representing space have been identified in the human PFC, they are much smaller than retinotopic maps in early visual areas, and intersubject variability in their location is much higher (Hagler & Sereno 2006). Furthermore, it is known from experiments in monkeys that only a subset of the neurons within the PFC subregions in which these maps have been found ac-

tually displays any spatial preference (Funahashi et al. 1989; Rainer et al. 1998). Importantly, the PFC also has a more complicated cytoarchitecture than sensory areas, with longer and more complex dendrites that allow for sampling of information from a wider range of inputs (Jacobs et al. 2001), which may affect the spatial scale at which information is represented and can be read out. Nevertheless, the overall picture that arises from studies employing optical imaging and microstimulation in monkeys is that at least several subregions of the PFC are topographically organized in a similar fashion as sensory areas (Roe 2010). However, while the topography of the PFC may be favorable to MVPA, neural representations in the PFC seem to exhibit more complex features and dynamics on a single neuron and population level than the representations in sensory areas where MVPA has been particularly successful. For example, recent studies in monkeys show that PFC representations are very high dimensional (Rigotti et al. 2013), that selectivity is not fixed but can be acquired (Bichot et al. 1996), that selectivity can change over time even within a trial (Stokes et al. 2013), and that populations of PFC neurons represent multiple stimulus dimensions at the same time even if one dimension is unattended (Mante et al. 2013). Thus, the dimensionality and temporal instability of neural representations in the PFC may pose a serious challenge to fMRI MVPA experiments, given that they rely on an inherently slow, hemodynamic signal that integrates neural activity over time.

Putting these and other (Anderson & Oates 2010; Vilarroya 2013) potential limitations of the MVPA approach aside, what other evidence do we have that the PFC is actually involved in conscious perception? In particular, is there *causal* evidence for a role of the PFC in conscious perception?

3 Beyond decoding: Causal evidence for a role of the PFC in conscious perception?

Early studies in macaque monkeys have found that lesions to the PFC can increase the lumin-

ance threshold (Latto & Cowey 1971) and degrade detection performance (Kamback 1973). More recently, studies in humans using transcranial magnetic stimulation have similarly found that stimulation of the PFC can impair the visibility of stimuli (Rounis et al. 2010), but also improve detection rates during visual masking (Grosbras & Paus 2003). Finally, Antoine Del Cul et al. have shown that perceptual thresholds are increased in patients with relatively small prefrontal lesions even when attentional effects are tightly controlled for (2009). However, none of these studies has shown dramatic impairments, but rather modulations of performance or perception. Total blindness has only been reported after removing the entire frontal cortex including (parts of) the underlying cingulate cortex in monkeys, and only lasted for a few days in several cases (Nakamura & Mishkin 1986). Importantly, other lesion studies in humans have not reported perceptual deficits at all (Heath et al. 1949; Markowitsch & Kessler 2000). Taken together with the fact that the PFC is also active during unconscious processing (Diaz & McCarthy 2007; Lau & Passingham 2007; van Gaal et al. 2008), not deactivated under Thiopental anesthesia (Veselis et al. 2004), but deactivated during rapid eye movement sleep when vivid (non-lucid) dreams can be experienced (Braun et al. 1998; Desseilles et al. 2011; Muzur 2002), this indicates that the evidence for a direct, specific involvement of the PFC in conscious perception is currently inconclusive at best.

4 An alternative to localization

Luckily, we can do without PFC, at least for the purposes of explaining conscious perception, while still maintaining many of the other, more compelling aspects of the GNW model. One central component that the GNW model shares with several other proposals about the neural correlates of consciousness (for example Melloni & Singer 2010; Tononi 2004; von der Malsburg 1997) is the concept of global integration of information. In light of the modular organization of the brain, a mechanism is required that brings information together such that an integrated, coherent percept

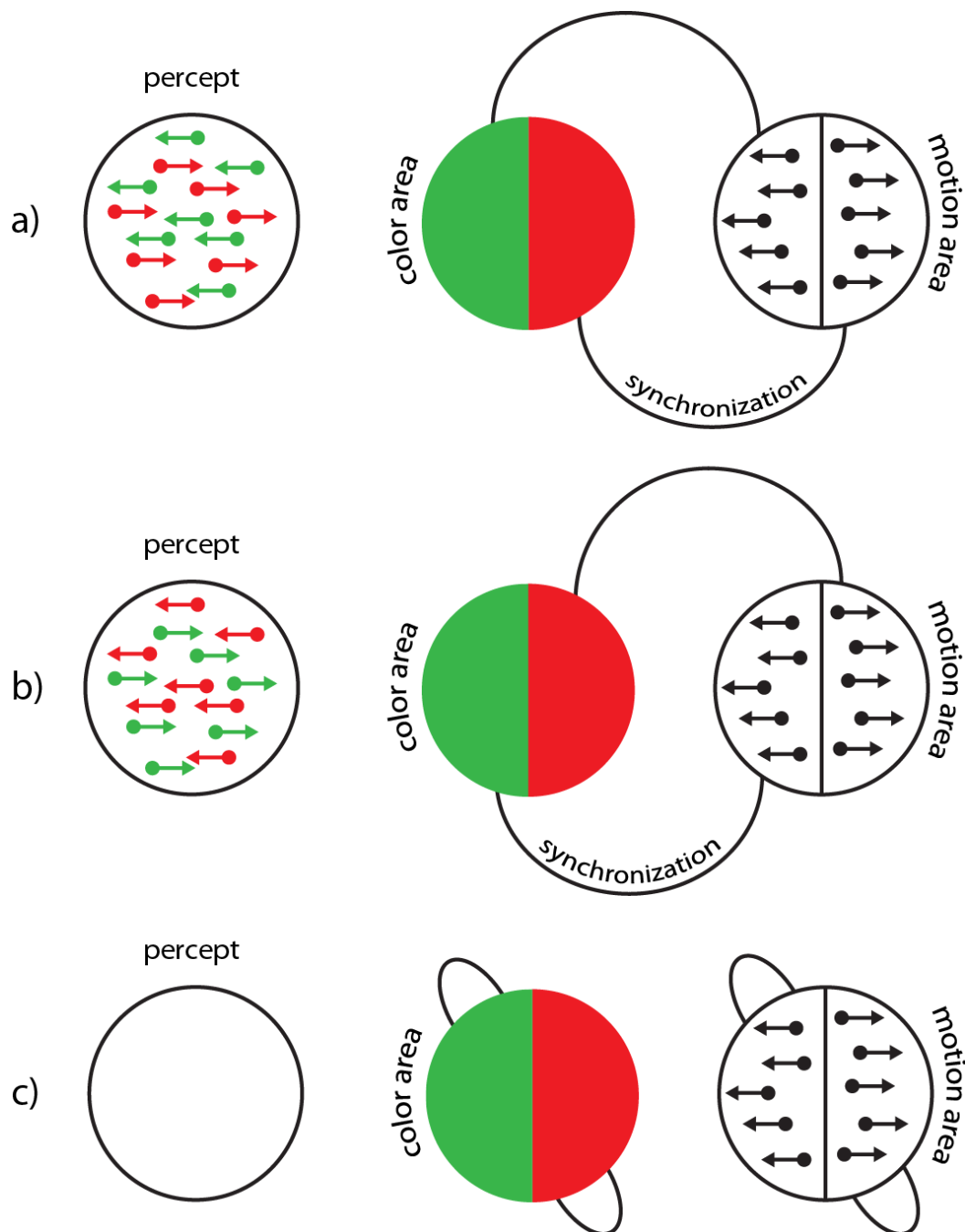


Figure 1: Neuronal synchrony binds distributed neurons into coherent assemblies, giving rise to conscious experience. Consider an experiment in which the subject is confronted with two superimposed, transparent surfaces of moving dots, as shown in the first column. **(a)** The dots on one surface are green and move to the left, and the dots on the other surface are red and move to the right. The two colors of dots are represented in a brain area coding for color, while the two motion directions are represented in an area coding for motion. If the neurons coding for green in the color area synchronize with the neurons coding for motion to the left in the motion area, and the neurons coding for red synchronize with the neurons coding for motion to the right, the two surfaces are consciously perceived. **(b)** A change in experience does not require a change in activity levels within areas, but a change of which neurons are synchronized. The opposite percept of (a) arises if neurons coding for green are synchronized with neurons coding for motion to the right, and if neurons coding for red are synchronized with neurons coding for motion to the left. Such content-specific synchronization between neurons has for example been observed in working-memory tasks in monkeys (Salazar et al. 2012). **(c)** Even when activity is synchronized within the color or motion area, respectively, a coherent conscious percept does not arise unless the areas are synchronized with each other.

can be formed. One attractive neural mechanism that can account for this requirement is neuronal synchrony (Bosman et al. 2012; Bressler et al. 1993; Salazar 2012). As has been discussed in greater detail elsewhere (Melloni & Singer 2010; Melloni this collection; Singer this collection), areas can be brought into *direct* contact with each other by synchronizing their neuronal activity, for example by phase alignment of neuronal oscillations, thus binding them into a functionally coherent assembly that forms a distributed representation of perceptual content. This self-organizing process can flexibly create and dissolve assemblies on top of a fixed anatomical architecture and does so without the need for anatomical convergence or broadcasting bottlenecks.

For example, imagine that a subject is confronted with two superimposed, transparent surfaces of moving dots (Figure 1). The dots on one surface are green and move to the left, and the dots on the other surface are red and move to the right. The two colors of dots are represented in a brain area coding for color, while the two motion directions are represented in an area coding for motion. For the subject to become conscious of the two surfaces, the neurons coding for green in the color area would need to synchronize their activity with the neurons coding for motion to the left in the motion area, and the neurons coding for red would need to synchronize with the neurons coding for motion to the right (Figure 1a). If the dots change direction, a new state of synchronization needs to be established, this time linking neurons coding for green with neurons coding for motion to the right, and neurons coding for red with neurons coding for motion to the left (Figure 1b). Hence, while the contents of the subject's experience are determined by the specific neuronal assemblies being active, conscious perception would be an emergent property of the state of synchronization. Recent tracing and modelling work in the macaque brain suggests that the kind of direct connectivity required to flexibly instantiate numerous, high-dimensional combinations of features is indeed afforded by high-density, reciprocal connections between brain areas (Markov et al. 2013).

Theoretical considerations and empirical evidence further suggest that the critical feature

differentiating conscious from unconscious processing is the spatial scale at which information is exchanged: while the integration of information in local modules, even in higher sensory areas (Sterzer et al. 2008), does not give rise to conscious experience by itself, large-scale integration over long distances does (Del Cul et al. 2007; Melloni et al. 2007). In the example of the transparent surfaces, this implies that even when activity is synchronized within the color or motion area, respectively, a coherent conscious percept cannot arise unless the areas are synchronized with each other (Figure 1c). Taken together, functional connectivity between distributed brain areas (i.e., connectivity that does not imply that one drives or controls the other) is an attractive alternative to localization in PFC as a candidate for the neural correlate of consciousness.

Coming back to the MVPA technique, this proposal makes a clear prediction that could be tested using decoding algorithms: specifically, one would predict that the large-scale connectivity patterns *between* brain regions for different percepts should differ, even if only slightly, for different conscious contents, and hence that conscious content should be decodable from them. This may well be the case in light of the fact that, at a much coarser scale, the neural correlates of auditory and visual awareness involve different brain networks (Eriksson et al. 2007), and that higher decoding accuracy for a subject's percept can be achieved if the joint activity patterns of several areas are considered instead of only singular patterns (Pessoa & Padmala 2007). The MVPA technique could in principle also be applied to other neuroimaging techniques that afford higher time resolution, such as electro- or magnetoencephalography or electrocorticography, thus potentially resolving the problem that arises because the representational carriers of perceptual content are highly dynamic and thus require a time-resolved analysis.

5 Conclusions

In summary, a complete theory for the neural correlates of consciousness should be able to account for the neural implementation of the contents of consciousness. John-Dylan Haynes has proposed a

clever experimental approach to localizing the contents of consciousness in the human brain, and has found that the PFC does not seem to be involved in this representation. Although surprising at first sight, this null result lines up well with the overall inconclusive evidence for a direct involvement of the PFC in conscious perception. However, it remains to be seen whether localization is the most fruitful approach to identifying the neural correlates of consciousness, or whether a more dynamic view that embraces the importance of communication between brain areas will bring us closer to solving the enigma of consciousness in the brain.

Acknowledgements

CMS is supported by a Human Frontier Science Program Long-term Fellowship (LT001118/2012-L). I would like to thank Lucia Melloni for her continuous support, insightful discussions, and comments on this manuscript.

References

- Anderson, M. L. & Oates, T. (2010). *A critique of multi-voxel pattern analysis*. Portland: Paper presented at the 32nd Annual Meeting of the Cognitive Science Society.
- Bichot, N. P., Schall, J. D. & Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, *381* (6584), 697-699. [10.1038/381697a0](https://doi.org/10.1038/381697a0)
- Bosman, C. A., Schoffelen, J. M., Brunet, N., Oostenveld, R., Bastos, A. M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P. & Fries, P. (2012). Attentional stimulus selection through selective synchronization between monkey visual areas. *Neuron*, *75* (5), 875-888. [10.1016/j.neuron.2012.06.037](https://doi.org/10.1016/j.neuron.2012.06.037)
- Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12* (5), 201-208. [10.1016/j.tics.2008.02.009](https://doi.org/10.1016/j.tics.2008.02.009)
- Braun, A. R., Balkin, T. J., Wesensten, N. J., Gwady, F., Carson, R. E., Varga, M., Baldwin, P., Belenky, G. & Herscovitch, P. (1998). Dissociated pattern of activity in visual cortices and their projections during human rapid eye movement sleep. *Science*, *279* (5347), 91-95. [10.1126/science.279.5347.91](https://doi.org/10.1126/science.279.5347.91)
- Bressler, S. L., Coppola, R. & Nakamura, R. (1993). Episodic multiregional cortical coherence at multiple frequencies during visual task performance. *Nature*, *366* (6451), 153-156. [10.1038/366153a0](https://doi.org/10.1038/366153a0)
- Dehaene, S. & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70* (2), 200-227. [10.1016/j.neuron.2011.03.018](https://doi.org/10.1016/j.neuron.2011.03.018)
- Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B. & Riviere, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4* (7), 752-758. [10.1038/89551](https://doi.org/10.1038/89551)
- Del Cul, A., Baillet, S. & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology*, *5* (10), e260. [10.1371/journal.pbio.0050260](https://doi.org/10.1371/journal.pbio.0050260)
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E. & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain*, *132* (9), 2531-2540. [10.1093/brain/awp111](https://doi.org/10.1093/brain/awp111)
- Desseilles, M., Dang-Vu, T. T., Sterpenich, V. & Schwartz, S. (2011). Cognitive and emotional processes during dreaming: A neuroimaging view. *Consciousness and Cognition*, *20* (4), 998-1008. [10.1016/j.concog.2010.10.005](https://doi.org/10.1016/j.concog.2010.10.005)
- Diaz, M. T. & McCarthy, G. (2007). Unconscious word processing engages a distributed network of brain regions. *Journal of Cognitive Neuroscience*, *19* (11), 1768-1775. [10.1162/jocn.2007.19.11.1768](https://doi.org/10.1162/jocn.2007.19.11.1768)
- Eriksson, J., Larsson, A., Ahlstrom, K. R. & Nyberg, L. (2007). Similar frontal and distinct posterior cortical regions mediate visual and auditory perceptual awareness. *Cerebral Cortex*, *17* (4), 760-765. [10.1093/cercor/bhk029](https://doi.org/10.1093/cercor/bhk029)
- Freeman, J., Brouwer, G. J., Heeger, D. J. & Merriam, E. P. (2011). Orientation decoding depends on maps, not columns. *Journal of Neuroscience*, *31* (11), 4792-4804. [10.1523/JNEUROSCI.5160-10.2011](https://doi.org/10.1523/JNEUROSCI.5160-10.2011)
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, *61* (2), 331-349.
- Grosbras, M. H. & Paus, T. (2003). Transcranial magnetic stimulation of the human frontal eye field facilitates visual awareness. *European Journal of Neuroscience*, *18* (11), 3121-3126. [10.1111/j.1460-9568.2003.03055.x](https://doi.org/10.1111/j.1460-9568.2003.03055.x)
- Hagler, D. J. jr. & Sereno, M. I. (2006). Spatial maps in frontal and prefrontal cortex. *NeuroImage*, *29* (2), 567-577. [10.1016/j.neuroimage.2005.08.058](https://doi.org/10.1016/j.neuroimage.2005.08.058)

- Haynes, J. D. (2009). Decoding visual consciousness from human brain signals. *Trends in Cognitive Sciences*, 13 (5), 194-202. [10.1016/j.tics.2009.02.004](https://doi.org/10.1016/j.tics.2009.02.004)
- (2015). An information-based approach to consciousness: Mental state decoding. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Heath, R. G., Carpenter, M. B., Mettler, F. A. & Kline, N. S. (1949). Visual apparatus: Visual fields and acuity, color vision, autokinesis. *Selective partial ablation of the frontal cortex, a correlative study of its effects on human psychotic subjects* (pp. 489-491). New York, NY: Paul B. Hoeber.
- Jackendoff, R. S. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.
- Jacobs, B., Schall, M., Prather, M., Kapler, E., Driscoll, L., Baca, S., Jacobs, J., Ford, K., Wainwright, M. & Trembl, M. (2001). Regional dendritic and spine variation in human cerebral cortex: A quantitative golgi study. *Cerebral Cortex*, 11 (6), 558-571. [10.1093/cercor/11.6.558](https://doi.org/10.1093/cercor/11.6.558)
- Kamback, M. C. (1973). Detection of brief light flashes by monkeys (*Macaca nemestrina*) with dorsolateral frontal ablations. *Neuropsychologia*, 11 (3), 325-329. [10.1016/0028-3932\(73\)90044-4](https://doi.org/10.1016/0028-3932(73)90044-4)
- Kamitani, Y. & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, 8 (5), 679-685. [10.1038/nn1444](https://doi.org/10.1038/nn1444)
- Latto, R. & Cowey, A. (1971). Visual field defects after frontal eye-field lesions in monkeys. *Brain Research*, 30 (1), 1-24. [10.1016/0006-8993\(71\)90002-3](https://doi.org/10.1016/0006-8993(71)90002-3)
- Lau, H. C. & Passingham, R. E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, 103 (49), 18763-18768. [10.1073/pnas.0607716103](https://doi.org/10.1073/pnas.0607716103)
- (2007). Unconscious activation of the cognitive control system in the human prefrontal cortex. *Journal of Neuroscience*, 27 (21), 5805-5811. [10.1523/JNEUROSCI.4335-06.2007](https://doi.org/10.1523/JNEUROSCI.4335-06.2007)
- Lau, H. C. & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, 15 (8), 365-373. [10.1016/j.tics.2011.05.009](https://doi.org/10.1016/j.tics.2011.05.009)
- Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503 (7474), 78-84. [10.1038/nature12742](https://doi.org/10.1038/nature12742)
- Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z. & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science*, 342 (6158), 1238406-1238406. [10.1126/science.1238406](https://doi.org/10.1126/science.1238406)
- Markowitsch, H. J. (1995). Cerebral bases of consciousness: A historical view. *Neuropsychologia*, 33 (9), 1181-1192. [10.1016/0028-3932\(95\)00057-A](https://doi.org/10.1016/0028-3932(95)00057-A)
- Markowitsch, H. J. & Kessler, J. (2000). Massive impairment in executive functions with partial preservation of other cognitive functions: The case of a young patient with severe degeneration of the prefrontal cortex. *Experimental Brain Research*, 133 (1), 94-102. [10.1007/s002210000404](https://doi.org/10.1007/s002210000404)
- Melloni, L. (2015). Consciousness as inference in time—A commentary on Victor Lamme. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W. & Rodriguez, E. (2007). Synchronization of neural activity across cortical areas correlates with conscious perception. *Journal of Neuroscience*, 27 (11), 2858-2865. [10.1523/JNEUROSCI.4623-06.2007](https://doi.org/10.1523/JNEUROSCI.4623-06.2007)
- Melloni, L. & Singer, W. (2010). Distinct characteristics of conscious experience are met by large scale neuronal synchronization. In E. Perry, D. Collerton, F. E. N. LeBeau & H. Ashton (Eds.) *New horizons in the neuroscience of consciousness* (pp. 17-28). Amsterdam, NL: John Benjamins.
- Muzur, A., Pace-Schott, E. F. & Hobson, J. A. (2002). The prefrontal cortex in sleep. *Trends in Cognitive Science*, 6 (11), 475-481. [10.1016/s1364-6613\(02\)01992-7](https://doi.org/10.1016/s1364-6613(02)01992-7)
- Nakamura, R. K. & Mishkin, M. (1986). Chronic 'blindsight' following lesions of nonvisual cortex in the monkey. *Experimental Brain Research*, 63 (1), 173-184. [10.1007/BF00235661](https://doi.org/10.1007/BF00235661)
- Panagiotaropoulos, T. I., Deco, G., Kapoor, V. & Logothetis, N. K. (2012). Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron*, 74 (5), 924-935. [10.1016/j.neuron.2012.04.013](https://doi.org/10.1016/j.neuron.2012.04.013)
- Passingham, R. (2009). How good is the macaque monkey model of the human brain? *Current Opinion in Neurobiology*, 19 (1), 6-11. [10.1016/j.conb.2009.01.002](https://doi.org/10.1016/j.conb.2009.01.002)
- Pessoa, L. & Padmala, S. (2007). Decoding near-threshold perception of fear from distributed single-trial brain activation. *Cerebral Cortex*, 17 (3), 691-701. [10.1093/cercor/bhk020](https://doi.org/10.1093/cercor/bhk020)
- Rainer, G., Asaad, W. F. & Miller, E. K. (1998). Memory fields of neurons in the primate prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (25), 15008-15013. [10.1073/pnas.95.25.15008](https://doi.org/10.1073/pnas.95.25.15008)

- Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K. & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497* (7451), 585-590. [10.1038/nature12160](https://doi.org/10.1038/nature12160)
- Roe, A. W. (2010). Optical imaging of short-term working memory in prefrontal cortex of the macaque monkey. In A. W. Roe (Ed.) *Imaging the brain with optical methods* (pp. 119-133). New York, NY: Springer.
- Roelfsema, P. R. & Treue, S. (2014). Basic neuroscience research with nonhuman primates: a small but indispensable component of biomedical research. *Neuron*, *82* (6), 1200-1204. [10.1016/j.neuron.2014.06.003](https://doi.org/10.1016/j.neuron.2014.06.003)
- Romo, R., Brody, C. D., Hernandez, A. & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, *399* (6735), 470-473. [10.1038/20939](https://doi.org/10.1038/20939)
- Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E. & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs meta-cognitive visual awareness. *Cognitive Neuroscience*, *1* (3), 165-175. [10.1080/17588921003632529](https://doi.org/10.1080/17588921003632529)
- Sahraie, A., Weiskrantz, L., Barbur, J. L., Simmons, A., Williams, S. C. & Brammer, M. J. (1997). Pattern of neuronal activity associated with conscious and unconscious processing of visual signals. *Proceedings of the National Academy of Sciences of the United States of America*, *94* (17), 9406-9411. [10.1073/pnas.94.17.9406](https://doi.org/10.1073/pnas.94.17.9406)
- Salazar, R. F., Dotson, N. M., Bressler, S. L. & Gray, C. M. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science*, *338* (6110), 1097-1100. [10.1126/science.1224000](https://doi.org/10.1126/science.1224000)
- Schwiedrzik, C. M., Ruff, C. C., Lazar, A., Leitner, F. C., Singer, W. & Melloni, L. (2014). Untangling perceptual memory: hysteresis and adaptation map into separate cortical networks. *Cerebral Cortex*, *24* (5), 1152-1164. [10.1093/cercor/bhs396](https://doi.org/10.1093/cercor/bhs396)
- Singer, W. (2015). The ongoing search for the neuronal correlate of consciousness. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Sreenivasan, K. K., Curtis, C. E. & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Science*, *18* (2), 82-89. [10.1016/j.tics.2013.12.001](https://doi.org/10.1016/j.tics.2013.12.001)
- Sterzer, P., Haynes, J. D. & Rees, G. (2008). Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. *Journal of Vision*, *8* (15), 1-12. [10.1167/8.15.10](https://doi.org/10.1167/8.15.10)
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D. & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78* (2), 364-375. [10.1016/j.neuron.2013.01.039](https://doi.org/10.1016/j.neuron.2013.01.039)
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5* (42). [10.1186/1471-2202-5-42](https://doi.org/10.1186/1471-2202-5-42)
- van Gaal, S., Ridderinkhof, K. R., Fahrenfort, J. J., Scholte, H. S. & Lamme, V. A. (2008). Frontal cortex mediates unconsciously triggered inhibitory control. *The Journal of Neuroscience*, *28* (32), 8053-8062. [10.1523/JNEUROSCI.1278-08.2008](https://doi.org/10.1523/JNEUROSCI.1278-08.2008)
- Veselis, R. A., Feshchenko, V. A., Reinsel, R. A., Dnistrian, A. M., Beattie, B. & Akhurst, T. J. (2004). Thiopental and propofol affect different regions of the brain at similar pharmacologic effects. *Anesthesia & Analgesia*, *99* (2), 399-408. [10.1213/01.ANE.0000131971.92180.DF](https://doi.org/10.1213/01.ANE.0000131971.92180.DF)
- Vilarroya, O. (2013). The challenges of neural mind-reading paradigms. *Frontiers in Human Neuroscience*, *7* (306). [10.3389/fnhum.2013.00306](https://doi.org/10.3389/fnhum.2013.00306)
- von der Malsburg, C. (1997). The coherence definition of consciousness. In M. Ito, Y. Miyashita & E. T. Rolls (Eds.) *Cognition, computation and consciousness* (pp. 193-204). Oxford, UK: Oxford University Press.
- Zaksas, D. & Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area MT during a working memory for visual motion task. *The Journal of Neuroscience*, *26* (45), 11726-11742. [10.1523/JNEUROSCI.3420-06.2006](https://doi.org/10.1523/JNEUROSCI.3420-06.2006)

Can Synchronization Explain Representational Content?

A Reply to Caspar M. Schwiedrzik

John-Dylan Haynes

Multivariate decoding provides an important tool for studying the representation and transformation of mental contents in the human brain. Specifically, decoding can be used to identify the neural correlates of contents of consciousness (NCCs). Decoding of functional magnetic resonance imaging (fMRI) signals has so far mostly revealed content-selectivity in sensory brain regions, but not in prefrontal cortex. The limitations of fMRI-decoding only permit cautious conclusions because fMRI signals are only indirectly related to neural coding. However, the role of prefrontal cortex in visual awareness is also questioned by other findings, reviewed in [Schwiedrzik \(this collection\)](#). Neural synchronization might offer an alternative to solving the binding problem by providing a computational means of integrating information encoded in distributed brain regions. However, it is unclear whether synchronization in itself serves as a coding dimension for visual features. Furthermore, other alternatives to synchronization, especially the role of spatial codes, need to be considered as potential solutions to the feature binding problem.

Keywords

Bias | Binding problem | Contents of consciousness | Dynamically changing coding space | Fmri-decoding | Functional magnetic resonance imaging | Global workspace theory | Multivariate decoding | Prefrontal cortex | Spatial code | Synchronization | Tolerance | Visual awareness

Author

[John-Dylan Haynes](#)

haynes@bccn-berlin.de

Charité – Universitätsmedizin Berlin
Berlin, Germany

Commentator

[Caspar M. Schwiedrzik](#)

cschwiedrz@rockefeller.edu

The Rockefeller University
New York, NY, U.S.A.

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

1 Introduction

Information-based approaches to brain function have been very successful in recent years ([Pouget et al. 2000](#); [Haynes & Rees 2006](#); [Kriegeskorte et al. 2006](#)). Most importantly, they allow to study how mental contents are represented and transformed during information processing in the brain. My target article in this volume ([Haynes this collection](#)) emphasized the importance of an information-based approach for the study of human consciousness, especially for understanding the neural mechanisms of

visual awareness. Whereas many previous studies mainly aimed to establish which additional processing needs to occur for a stimulus to reach awareness, a second question is equally important: how and where the brain encodes the specific contents of consciousness. Research on these neural correlates of the contents of consciousness (NCCs; [Chalmers 2000](#); [Block 2007](#); [Koch 2004](#)) has been sparse. For identifying NCCs, simply establishing that a brain area responds stronger under conscious than un-

der unconscious processing is not sufficient, because this could merely reflect unspecific processes such as attention or memory (Corbetta & Shulman 2002; Goldman-Rakic 1995). Instead, for identifying the neural code of contents several specific questions need to be addressed: Which brain regions encode sensory information in a representational space that exactly matches perception? And under which circumstances does a crossing of the threshold to awareness involve changes of representations in these specialized coding spaces?

2 The role of prefrontal cortex

One example of the importance of considering content-based processing is the global workspace theory (Dehaene & Naccache 2001; Baars 2002). In specific readings of the theory, consciousness involves a distribution of sensory information from sensory cortices to parietal and prefrontal cortex (Haynes this collection). Increased activity in frontoparietal regions under conscious perception is seen as evidence for such a “broadcasting” of sensory information (Baars 2002). However, without additional support by information-based or representational analyses, increased activity in frontoparietal regions with increased awareness might simply reflect unspecific processes, say as in detecting or reporting a change in perception, rather than coding the sensory information itself. In several studies with functional magnetic resonance imaging (fMRI) we found no evidence for changes in prefrontal representation of sensory information under increased levels of awareness (reviewed in Haynes this collection). Thus, we found no evidence that sensory information is re-represented in prefrontal or parietal cortex.

At this important point, the comment by Schwiedrzik (this collection) adds further important details on the potential role of prefrontal cortex (PFC) in visual awareness (Crick & Koch 1995; Dehaene & Naccache 2001). In a first line of arguments Schwiedrzik provides more detail on a point briefly sketched in the original article (Haynes this collection), whether absence of decodable information in PFC might reflect limitations of fMRI-based pattern classi-

fication. fMRI decoding will only be able to access neural information that is encoded in specific formats and topologies (Chaimow et al. 2011). For example, if neurons with different tuning properties are randomly distributed within a voxel, then the voxel will not be able to pick up any information about these properties. Thus, a macroscopic clustering of cells with similar tuning properties is required for fMRI-decoding to pick up information. So information could be present in prefrontal cortex, but in a format that is not accessible to fMRI, not even with the increased sensitivity of multivariate analyses. This might explain the discrepancy between the absence of information in many fMRI studies and differential responsivity to stimulus features of cells in PFC in non-human primates (Pasternak & Greenlee 2005). Please note, however, that most of the evidence for sensory tuning in PFC is obtained under working memory paradigms, which also includes temporal two-alternative forced choice tasks (Romo et al. 1999). Thus, it is unclear whether this generalizes to “realtime” perceptual experience.

There are further challenges in accessing neural information. Schwiedrzik (this collection) brings forward an important point already raised previously (Duncan & Owen 2000): coding in prefrontal cortex might be dynamic and thus the code might change across time. Such dynamically changing coding spaces might again not be detectable in classification analyses that assume a constant population code across the period analysed (Stokes et al. 2013). These points raised by Schwiedrzik are fully valid: It is highly important to consider these limitations when interpreting the results of fMRI decoding studies. To some degree these challenges might be alleviated with future technical developments. For example, columnar-level information can be accessed following recent advances in high-resolution fMRI (Yacoub et al. 2008). However, many limitations of fMRI will remain due to its vascular origin that only samples neural information indirectly. Please note that the limitations go far beyond the points raised by Schwiedrzik (this collection). For example, fMRI might not only miss information, but it might

also tap into information that is not available at the level of single neurons. Say, if an fMRI voxel samples a homogenous group of cells with highly similar tuning properties, the voxel might reflect a degree of averaging that is not available at the level of single neurons.

Please note, that the target article was not restricted to decoding approaches in fMRI alone. Instead, the aim was to outline a more general approach to studying the neural correlates of the contents of consciousness. If suitable recording techniques were available the information could be assessed based on a whole family of potential representational signals, including especially axonal and dendritic population activity. Please further note, that any recording technique has its blind spots. For example single-cell electrophysiology is biased towards large cells (Bartels et al. 2008), or optical imaging with voltage sensitive dyes is restricted to the surface of the brain (Grinvald & Hildesheim 2004). Thus, only a combination of techniques will be able to provide a full picture of the changes in neural coding with varying levels of visual awareness. Importantly, in a second line of argumentation Schwiedrzik (this collection) provides additional support and plausibility to the finding that the absence of information in PFC is real. For example the effects of lesions in PFC on visual recognition can be strikingly weak, which would not be expected if representation or routing of information in PFC were a necessary condition for awareness.

A third line of argumentation brings neural integration between spatially separated brain regions into play as a different potential mechanism of visual awareness (Engel & Singer 2001). The basic idea is that representation might involve a dynamic binding involving a synchronization of neural activity. According to this model, as Schwiedrzik points out, a distribution of sensory information into prefrontal brain regions might not be necessary. Many studies have related changes in neural synchronization to changes in visual awareness (see e.g., Engel & Singer 2001; Uhlhaas et al. 2009). However, it is important to look more closely at the explanation that can be obtained by changes in synchronization, specifically if the aim is to

provide an explanation of the neural correlates of contents of consciousness (NCCCs). Typically, synchronization is not viewed as a coding dimension for contents, but as a code for binding and integration of features that are distributed across multiple content-specific regions (von der Malsburg 1999). The example provided in figure 1 of the comment by Schwiedrzik illustrates this nicely. A person views two superimposed clouds of moving dots, one green cloud moving left and a red cloud moving right. The features are encoded in content-specific fashion with two different activation patterns in the color area coding the two colors and two different activation patterns in the motion area coding the two motion directions. Synchronization between the neural representations of “green” and “left” on the one hand and “red” and “right” enables a separate binding of these two distributed features and also allows them to jointly be more effective in activating any downstream brain regions (König et al. 1996). Here, the contents are represented as differential activation states in the content-specific regions and their binding is achieved by synchronization. In this example representation and binding are separable problems based on separate computational mechanisms. However, Schwiedrzik (this collection) also goes one step further by suggesting that the large-scale connectivity patterns between brain regions might themselves code for different conscious contents. It has already been shown with fMRI that connectivity patterns between remote brain regions reflect changes in visual awareness (e.g., Haynes et al. 2005; Imamoglu et al. 2012). It has also been shown that connectivity matrices obtained with fMRI can be used to classify cognitive states (Richiardi et al. 2011; Heinzle et al. 2012). However, to date I am not aware of any evidence that fine-grained perceptual contents are encoded in differential patterns of brain connectivity. Furthermore, synchronization sometimes fails to explain perception (Thiele & Stoner 2003) and there are other solutions to the binding problem besides synchronization. For example, high-level content-selective brain regions that achieve a certain degree of tolerance to variations in spatial location, still have con-

siderable information about the spatial location of features (e.g., [Cichy et al. 2011](#)). Thus, the spatial maps and their associated differential anatomical (as opposed to functional) connectivity patterns provide a plausible alternative hypothesis to synchronization ([Treisman & Gelade 1980](#)).

3 Conclusions

Prefrontal cortex is often considered vital for visual awareness ([Crick & Koch 1995](#); [Dehaene & Naccache 2001](#)), however multivariate decoding studies have revealed a marked absence of sensory information in prefrontal cortex ([Haynes this collection](#)). Neural synchronization ([Schwiedrzik this collection](#)) might provide an alternative account for feature binding and selective routing of information. However, it is currently unclear whether any form of functional connectivity can itself code specific sensory contents.

References

- Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 6 (1), 47-52.
- Bartels, A., Logothetis, N. K. & Moutoussis, K. (2008). fMRI and its interpretations: An illustration on directional selectivity in area V5/MT. *Trends in Neurosciences*, 31 (9), 444-453. [10.1016/j.tins.2008.06.004](https://doi.org/10.1016/j.tins.2008.06.004)
- Block, N. (2007). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, 30 (5-6), 481-499. [10.1017/S0140525X07002786](https://doi.org/10.1017/S0140525X07002786)
- Chaimow, D., Yacoub, E., Ugurbil, K. & Shmuel, A. (2011). Modeling and analysis of mechanisms underlying fMRI-based decoding of information conveyed in cortical columns. *NeuroImage*, 56 (2), 627-642. [10.1016/j.neuroimage.2010.09.037](https://doi.org/10.1016/j.neuroimage.2010.09.037)
- Chalmers, D. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.) *Neural correlates of consciousness: Conceptual and empirical questions* (pp. 17-40). Boston, MA: MIT.
- Cichy, R. M., Chen, Y. & Haynes, J. D. (2011). Encoding the identity and location of objects in human LOC. *NeuroImage*, 54 (3), 2297-2307. [10.1016/j.neuroimage.2010.09.044](https://doi.org/10.1016/j.neuroimage.2010.09.044)
- Corbetta, M. & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3 (3), 201-2015.
- Crick, F. & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375 (6527), 121-123.
- Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79 (1-2), 1-37.
- Duncan, J. & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, 23 (10), 475-483.
- Engel, A. K. & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, 5 (1), 16-25.
- Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, 14 (3), 477-485.
- Grinvald, A. & Hildesheim, R. (2004). VSDI: A new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience*, 5 (11), 874-885.
- Haynes, J. D. (2015). An information-based approach to consciousness: Mental state decoding. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a.M., GER: MIND Group.

- Haynes, J. D., Driver, J. & Rees, G. (2005). Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. *Neuron*, *46* (5), 811-821.
- Haynes, J. D. & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7* (7), 523-534.
- Heinzle, J., Wenzel, M. A. & Haynes, J. D. (2012). Visuo-motor functional network topology predicts upcoming tasks. *Journal of Neuroscience*, *9960* (9968), 475-483. [10.1523/JNEUROSCI.1604-12.2012](https://doi.org/10.1523/JNEUROSCI.1604-12.2012)
- Imamoglu, F., Kahnt, T., Koch, C. & Haynes, J. D. (2012). Changes in functional connectivity support conscious object recognition. *NeuroImage*, *63* (4), 1909-1917. [10.1016/j.neuroimage.2012.07.056](https://doi.org/10.1016/j.neuroimage.2012.07.056)
- Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood: Roberts.
- Kriegeskorte, N., Goebel, R. & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the USA*, *103* (10), 3863-3868.
- König, P., Engel, A. K. & Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends in Neurosciences*, *19* (4), 130-137.
- Pasternak, T. & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6* (2), 97-107.
- Pouget, A., Dayan, P. & Zemel, R. (2000). Information processing with population codes. *Nature Reviews Neuroscience*, *1* (2), 125-132.
- Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P. & Van De Ville, D. (2011). Decoding brain states from fMRI connectivity graphs. *NeuroImage*, *56* (2), 616-626. [10.1016/j.neuroimage.2010.05.081](https://doi.org/10.1016/j.neuroimage.2010.05.081)
- Romo, R., Brody, C. D., Hernández, A. & Lemus, L. (1999). Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature*, *399* (6735), 470-473.
- Schwiedrzik, C. (2015). What's up with prefrontal cortex? – A commentary on John-Dylan Haynes. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a.M., GER: MIND Group.
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D. & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78* (2), 364-375.
- Thiele, A. & Stoner, G. (2003). Neuronal synchrony does not correlate with motion coherence in cortical area MT. *Nature*, *421* (6921), 366-370.
- Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12* (1), 97-136.
- Uhlhaas, P. J., Pipa, G., Lima, B., Melloni, L., Neunenschwander, S., Nikolic, D. & Singer, W. (2009). Neural synchrony in cortical networks: History, concept and current status. *Frontiers in Integrative Neuroscience*, *3* (17). [10.3389/neuro.07.017.2009](https://doi.org/10.3389/neuro.07.017.2009)
- von der Malsburg, C. (1999). The what and why of binding: The modeler's perspective. *Neuron*, *24* (1), 95-104.
- Yacoub, E., Harel, N. & Ugurbil, K. (2008). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences of the USA*, *105* (30), 10607-10612. [10.1073/pnas.0804110105](https://doi.org/10.1073/pnas.0804110105)