Psychological and Neurobiological Foundations of Consciousness

John-Dylan Haynes

Contents

G. Roth et al. (eds.), *Psychoneuroscience*, https://doi.org/10.1007/978-3-662-65774-4_8

This chapter deals with a highly popular and at the same time fiercely controversial psycho-neurobiological topic, namely consciousness and its possible neural foundations. In the following, we will set aside the philosophical dimensions of this topic, even though we appreciate their importance, and concentrate on approaches that are based on empirical psychological and neurobiological studies. Research on the neural correlates of consciousness seeks to determine what happens differently in the brain when a stimulus is consciously recognized. What differences characterize the neural processing of consciously perceived stimuli compared to stimuli that do not reach consciousness? Is there evidence that stimuli do not reach consciousness but are still processed by the brain? Can the brain decide what to be consciously aware of and what not? This chapter will deal with these and similar questions.

Learning Objectives

After reading this chapter, you should have gained a deeper insight into the findings of empirical consciousness research. This includes knowledge of the applied psychological and neuroscientific methods, as well as the experimental distinction between unconscious and conscious perceptual and cognitive performance. Likewise, you will become familiar with core questions in the neuroscience of consciousness.

8.1 Methodology of Consciousness Research

The neuroscience of consciousness aims to identify differences in how conscious and unconscious stimuli are processed in the brain. A simple first approach is to compare processing of two types of stimuli: First, weak or masked stimuli that don't cross the threshold to consciousness (i.e., **subliminal or subthreshold stimuli**, from *limes*, Latin for the boundary); second, stimuli that are unmasked or sufficiently strong to cross the threshold to consciousness (i.e., **suprathreshold or supraliminal stimuli**). Comparing these two cases should allow us to determine which brain regions are more involved in conscious stimulus processing than in subthreshold stimulus processing (e.g., Dehaene et al. 2001; Haynes and Rees 2005a).

This approach, however, is rather crude in providing only two conditions on the continuum between invisibility and visibility. A more fine-grained approach would be to cross the **perceptual threshold in more fine-grained steps**. This seems to be simple at first glance. One starts with a simple, faint and brief stimulus and then makes it either longer or more intense in order to determine the presentation duration or the intensity at which a person just begins to perceive a stimulus. One important finding of such studies is that the threshold of perception is not an abrupt, discontinuous transition from intensities at which the stimulus is never seen to intensities at which the stimulus is always seen (see Gescheider 1997). Around the threshold, there are intensities where perception of the stimulus is not fully determined and it is seen only with some probability. This results in an S-shaped rather than a step-shaped threshold function.

The continuous and gradual nature of the perceptual threshold raises the question how it can be defined technically. Typically, a threshold is defined as the stimulus intensity (duration) at which a certain percentage of responses are correct (e.g., 75%). Note that with such a definition performance is already above chance at the threshold. So why not define the threshold as the intensity at which the visibility first starts deviating from baseline? This has statistical reasons, because it is difficult to define the threshold as the first intensity or duration where a stimulus begins to be seen barely above chance. Take an accuracy of 51%. A lot of trials would be needed to tell if a measured accuracy of 51% really is above chance, or if this just reflects random fluctuations in people's behavior. In order to investigate truly unconscious processing and ensure that the subject has not seen a stimulus even on a few trials, the recognition of the stimulus must be at the chance level (e.g. 50% for two equally likely stimuli). For this, Bayesian statistics can be useful because they provide a framework for quanitifying evidence for the absence of an effect.

8.1.1 Criteria for Conscious Perception I: Subjective Threshold

One problem with determining the perceptual threshold is that there are different criteria for whether a stimulus has been perceived. At first glance, one could simply ask the subject whether they saw a stimulus or not and base the threshold on their judgment. There are numerous examples of the use of such *subjective judgments* in research on **unconscious stimulus processing**. One example is a study by Berti and Rizzolatti (1992) on neglect patients. Patients with right parietal lesions often show an attentional deficit for stimuli in the left visual field, especially when these stimuli are in competition for attention with stimuli in the right visual field. This visual hemineglect is not due to perceptual deficits, as single isolated stimuli in the left visual field can be readily detected. Berti and Rizzolatti (1992) presented prime stimuli to the neglected visual field and target stimuli to the intact visual field and investigated whether the invisible primes had an effect on target perception. They inferred the invisibility of the primes from the subjective reports of the patients, who reported seeing only the stimulus in the right visual field. From this it was possible to conclude that unconscious primes have an influence on the processing of conscious targets (see Box: Confidence Judgments).

A study by Moutoussis and Zeki (2002) on the processing of unconscious object stimuli was also based on subjective judgments about whether subjects were consciously aware of stimuli. Face and house stimuli were presented in two different visibility conditions. Stimuli had either the same color or opposite colors in both eyes (green on a red background in one eye and red on a green background in the other eye). As a measure of visibility, subjects were asked to give one of three possible responses depending on whether they thought they saw "house," "face," or "neither." In the counter-colored condition, the perceptibility of objects was greatly reduced. Subjects reported seeing no picture in most cases, so according to their subjective judgment they were unaware of the stimuli. However, an objective discrimination test revealed that some subjects were disproportionately good at guessing which picture had been shown. This typical dissociation between subjective reports and objective discrimination measures is discussed in more detail below in ► Sect. 8.2.

Experiments on unconscious stimulus processing were already conducted in the early days of experimental psychology, even with far more detailed distinctions between "conscious" and "unconscious" (Peirce and Jastrow 1884). As early as 1884, Peirce and Jastrow presented weight stimuli when investigating *tactile* perception and asked subjects to discriminate between positive and negative changes in weights. To investigate the role of consciousness in this, *confidence* judgments were collected, ranging from 0 ("No preference for one judgment over the other, the question seemed meaningless") to 3 ("High confidence of having answered correctly"). Subjects were disproportionately correct even at the lowest confidence level. Later experiments also showed that even at minimal confidence levels, discrimination performance can be disproportionately good (for an early review, see Adams 1957). Confidence judgments are also occasionally collected in experiments on signal detection theory (Green and Swets 1966). As a rule, subjects are overconfidently good even at the lowest confidence ratings (i.e., above the identity in the hits-versus-false alarms diagram).

8.1.2 Criteria for Conscious Perception II: Objective Threshold

However, it was already criticized in the 1960s that subjective methods could be influenced by possible conservative response tendencies (Eriksen 1954, 1960; Kunimoto et al. 2001). This is understood to mean that a subject may not be entirely sure that he or she has seen the stimulus when the perception is faint and unclear, but may prefer to answer "no" once too often rather than "yes" too often, i.e., have a tendency toward false negative rather than false positive judgments. This problem is particularly glaring in the field of *perceptual defense* research (e.g., McGinnies 1949), in which the perceptual threshold for taboo words (*dirty words*; Eriksen 1954) is reported to be elevated. Rather than a higher perceptual threshold for taboo words, it is also possible that subjects just have a reluctance to report these words (Eriksen 1954).

The problem of **partial information** also plays a major role here. Even if a subject did not *fully* recognize a stimulus, any partial or fragmentary information might be sufficient for the required discrimination (Kunimoto et al. 2001; Kouider and Dupoux 2004). For example, in an experiment by Sidis, subjects were asked to recognize numbers and letters on cards (Sidis 1898). The cards were presented at such a distance that subjects reported perceiving only a blurred dot. This was taken as evidence that the stimuli were not consciously perceived. Nevertheless, subjects were able to discriminate between letters and numbers disproportionately well. However, since the subjects had weak, albeit undifferentiated, perception, it may be that even fragmentary conscious perception allowed them to distinguish between numbers and letters (e.g., numbers tend to have more curves than letters). The problem of partial information is compounded by the fact that a dichotomous judgment of perception as "conscious" or "unconscious" forces subjects to split a possible continuous graduation of awareness between two categories (Kunimoto et al. 2001). Frame and anchor effects may occur, so that the separation between "conscious" and "unconscious" responses may simply be oriented towards a median of visibility, with the consequence that even partially conscious stimuli are classified as "unconscious".

8.1.3 Experimental Implementation

Even once a decision has been made in favor of subjective or objective threshold measurements, the determination of a perceptual threshold is fraught with numerous further difficulties. For example, as already pointed out by Fechner (1860), the result depends on the temporal order used in which stimuli of different intensities are presented. In the *method of adjustment* the subject themself is allowed to set the intensity at which a stimulus is just perceived. In the *method of limits*, stimuli of increasing intensity are presented and the time at which the stimulus is perceived is noted. To avoid hysteresis effects, one then switches from increasing to decreasing intensity and notes at what point the stimulus is no longer perceived. This is repeated a few times, and the threshold is obtained as the mean value of the measured values of ascending and descending series of measurements. The most reliable method is the *method of constant stimuli*, in which different intensities are presented randomly in the threshold range, which allows sequence effects and expectations to be excluded. Likewise, it should be taken into account that often during a longer experiment stimuli that cannot be detected at first are seen later after some experience, an effect that can be attributed to perceptual learning (Kahnt et al. 2011; Watanabe et al. 2001).

Background Information

Perceptual framing effects refer, for example, to the fact that a stimulus is judged differently depending on the composition of the set of other stimuli it is presented with in an experiment (anchor effects, see e.g. Cannon 1984; Gescheider 1988). When determining thresholds, one usually cannot isolate individual variables but must consider the entire physical physical properties of the stimulus. For example, sometimes people report a "flicker fusion frequency", above which the flickering of a light source is no longer perceived and it appears to be continuously illuminated. However, this frequency depends (among other factors) on stimulus intensity, so it is not possible to state a single, universal flicker fusion frequency (Watson 1986). Not only perceptual, but also cognitive context effects must be taken into account. Consider, for example, a priming experiment in which the effect of a subthreshold priming stimulus on the judgment of a suprathreshold target stimulus is to be examined. For example, one could work with pattern masking and present the following stimulus sequence: Mask (100 ms)—Prime (16 ms)—Mask (100 ms)—Target stimulus (200 ms). In order to ensure that the prime was truly invisible one could randomly test on some proportion of trials whether the participant can see the prime. In the remaining trials one can then test for a priming effect on the perception of the target. However, the difference between these two cognitive tasks (judging the prime versus the target) will have an effect on the visibility of the prime, presumably leading one to over-estimate prime visibility. An alternative practice is to test the visibility before or after the series of measurements. But again, with this different task prime visibility could be overestimated, and perceptual learning throughout the experiment might also influence the visibility.

8.1.4 Criteria for Subliminal Processing

To investigate subthreshold processing, it is necessary, on the one hand, to measure the threshold of consciousness, as explained in ► Sect. 8.1.3. On the other hand, some behavioral or neuroscientific evidence is also required that indicates whether the subliminal stimulus is still processed by the brain (◘ Fig. 8.1). For example, one could test whether the subliminal—despite being unseen—has an effect on word stem completion or on some signature of brain activity. Measuring the conscious perceptibility of a stimulus is called a *direct measure of* processing (Reingold and Merikle 1988)—*direct* because the task directly and explicitly refers to seeing the stimulus. The measure of hypothetical subliminal processing of the stimulus is called *indirect* because it does not refer directly to the stimulus but allows the indirect effects of a stimulus on a task to be measured.

One possibility is to look for **qualitative differences** between conscious and unconscious processing (e.g., Merikle et al. 1995), in which case implicit effects cannot simply reflect a weak form of perception because it has very different properties. Another method is the **dissociation between confidence judgments and discrimination performance** (Kunimoto et al. 2001), in which a subject's discrimination performance is assumed to be unconscious when he is unable to say in which trials his performance is good or bad. The process dissociation method is also applied to subliminal perception (Debner and Jacoby 1994). The logic of this procedure is that it should be impossible to rule out the effect of a stimulus that is presented unconsciously, so that it nevertheless has a nonconscious and therefore uncontrollable influence on a word stem completion. This connection between consciousness and intentional controllability has been pointed out by several other researchers (Marcel 1983; Holender 1986). But even these newer approaches are not without controversy, so that presumably only the development of explicit mathematical models will bring about a clarification of the measurement problems (Schmidt and Vorberg 2006).

As mentioned above, it is also possible to take a completely different approach and use brain processes triggered by unconscious stimuli as evidence for implicit processing

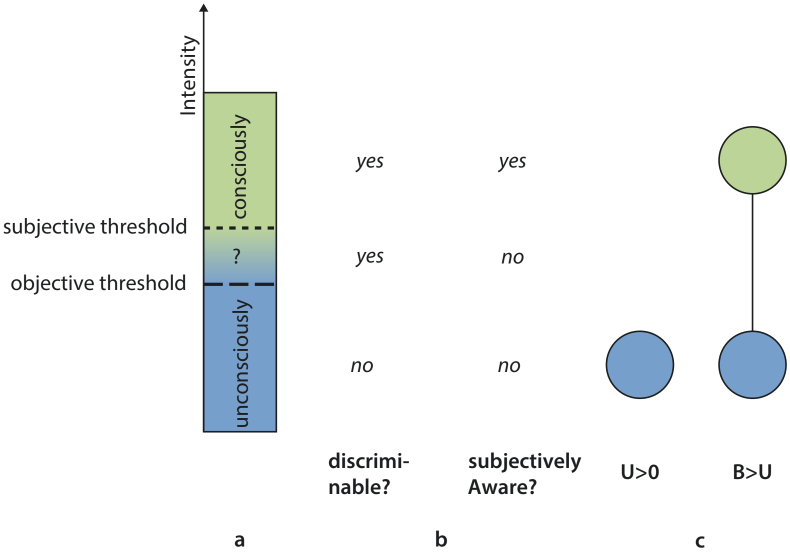


Fig. 8.1 Thresholds and experimental contrasts. **a**, **b** Differential awareness of stimuli presented with increasing intensity. Two ranges are uncontroversial: Below the objective threshold, a stimulus is no longer discriminable and possible residual processing is unconscious. Above the subjective threshold, the stimulus is consciously recognized and can be discriminated. There are different interpretations of the range between the subjective and objective thresholds in which a stimulus was not considered conscious by the subject, but he was still able to discriminate it correctly. According to one view, the stimulus is unconscious and the residual discriminability is due to unconscious processing. According to another interpretation, the stimulus in this domain is conscious, except that subjects are very conservative in their response behavior. **c** Two important statistical contrasts (comparisons) that examine different aspects of conscious and unconscious information processing and are commonly used in neuroimaging experiments. Contrast U > 0 tests whether significant activation (greater than zero) occurs with unconscious stimuli. This can be used to identify regions where unconscious stimuli are processed. The contrast B > U tests for regions where conscious stimuli elicit greater activity than unconscious stimuli. This allows us to determine what cortical activity is “added” during conscious processing

instead of implicit measures of behavior (e.g., Haynes and Rees 2005a). This will be illustrated in the following.

8.2 Neuronal Correlates of Conscious and Unconscious Stimulus Processing

Assuming that one has managed to avoid all the pitfalls and that one has experimentally generated a clean distinction between subliminal and suprathreshold stimuli. Then one can set about the task of comparing the neural processing of consciously perceived and subliminal stimuli in order to determine something about the neural correlates of consciousness. First, one can examine what additional neural processing occurs during conscious compared to unconscious stimulus processing (■ Fig. 8.1c). On the other hand, one can investigate the depth to which unconscious stimuli are processed (■ Fig. 8.1b).

8.2.1 Conscious Neural Processing

A number of studies have examined the effects of awareness of visual information on brain activity in humans and in nonhuman primates. Most notably, conscious stimuli have been found to undergo more extensive cortical processing compared to unconscious stimuli. In one study, Dehaene et al. (2001) used functional magnetic resonance imaging (fMRI) to examine the cortical processing of masked words that were either visible or invisible. The subthreshold invisible word stimuli activated the brain, but the activation remained restricted to visual brain regions. In contrast, the suprathreshold stimuli were characterized by widespread activation in parietal and prefrontal brain regions as well. This was interpreted by Dehaene et al. as an indication of a *distribution of* sensory information in the brain during conscious perception and was interpreted in terms of the global workspace theory (► Sect. 8.5).

Suprathreshold and subthreshold stimuli are physically different stimuli. As a rule, one has to increase the stimulus intensity or attenuate a masking in order to cross the threshold of consciousness. However, this complicates the interpretation of the measured differences: a brain region that responds more strongly to a consciously perceived stimulus than to a subliminal one could possibly be responding only to the fact that the latter has a higher stimulus intensity, that it entails a stronger exogenous allocation of attention, or that the subject prepares and executes a response to the stimulus (Dehaene et al. 2001).

Some of these problems can be avoided by a more fine-grained examination of the perception threshold. As explained in ► Sect. 8.1, the perception threshold does not jump abruptly from subthreshold to suprathreshold with increasing intensity, but there is a range in which perception is only partially determined and the subject consciously recognizes even a constant stimulus only in a portion of the runs. The S-shaped transition at threshold can be approximated mathematically as a cumulative Gaussian normal distribution (Gescheider 1997). This makes it possible to correlate the individual course of a threshold for a particular stimulus and a particular person with the neuronal fMRI responses in different brain regions in order to search for areas in which the activity level reflects the threshold course.

It was shown that in the perception of masked objects, the profile of the perceptual threshold correlated with the level of activity in the lateral occipital cortex (LOC) but not in the primary visual cortex (V1) (Grill-Spector et al. 2000). Similarly, masking of simple brightness stimuli showed that the shape of the masking function was reflected in the connectivity between early (V1) and later visual areas (in the fusiform gyrus) (Haynes et al. 2005a, b). This direct comparison between *psychometric* threshold functions and neural or *neurometric* response curves can also be applied to the response behavior of single cells and cell populations in sensory brain regions (Parker and Newsome 1998).

Fluctuations in perceptual judgments of the same physical stimulus have been shown to be accompanied by fluctuations in neural activity already in early visual cortex. For example, in trials in which the stimulus is seen, activity in primary visual cortex V1 is higher than in trials in which the stimulus is not seen (Ress and Heeger 2003).

Fluctuations in perceptual judgments not only occur between individual stimulus presentations, but there are also other, slower stochastic fluctuations in perception. Tononi et al. (1998) used MEG to investigate fluctuations in awareness of rival stimuli. Conscious stimuli were found to produce more extensive activation and coherence in regions beyond the visual system. In a study of *change blindness*, Beck et al. (2001) showed that conscious awareness of a change in a stimulus display leads to increased activity in frontoparietal networks. Similarly, Vuilleumier et al. (2001) found the same in a study of perception in a neglect patient. In runs in which the patient consciously recognized a stimulus in the neglected visual field, brain activity was also increased in parietal regions.

Moreover, there was a widespread increase in effective connectivity between visual, parietal, and prefrontal regions during conscious perception. Similar effects of conscious stimulus perception on functional connectivity measures were also found in other studies (e.g., Lumer and Rees 1999; Dehaene et al. 2001; Haynes et al. 2005a, b).

8.2.2 Unconscious Neural Processing

Research on the neural processing of unconscious stimuli shows that they are processed very deeply. On the one hand, stimulus properties that are not consciously recognized are already represented at the level of the primary visual cortex in the activity levels of individual neurons as well as in fMRI measurements of neuronal populations (Gur and Snodderly 1997; Haynes and Rees 2005a), which had already been indicated previously in behavioral studies (He et al. 1996). However, processing of unconscious stimuli is not restricted to early stages of the visual system, but also reaches higher, content-specific processing stages (Moutoussis and Zeki 2002).

Moutoussis and Zeki (2002) were able to show that invisible images of houses and faces are processed in the corresponding regions parahippocampal place area (PPA) and fusiform face area (FFA, Kanwisher et al. 1997), respectively. Activation for invisible stimuli was significant, but was significantly weaker than for visible stimuli. Studies in neurological patients have shown that even in neglect patients, stimuli that are detected in isolation but do not reach consciousness due to competition with stimuli in other regions of the visual field nevertheless selectively activate content-specific brain regions (Rees et al. 2002). Fang and He (2005) used interocular suppression stimuli to show that the *dorsal* visual pathway is also activated by invisible objects. Interestingly, activation was almost as high for invisibly presented images of tools as for visible images, which fits with the theory that the dorsal pathway is relevant to action control (Goodale and Milner 1992). Moreover, emotional aspects of invisible pictures can activate emotion networks in amygdala and orbitofrontal cortex (Vuilleumier et al. 2002). The unconscious neural processing of stimuli may even lead to conditioning processes in the basal ganglia (Pessiglione et al. 2008), which may have direct effects on choice behavior.

8.2.3 The Phenomenon of Blindsight

An interesting phenomenon is *blindsight* (Weiskrantz et al. 1974; Weiskrantz 2004). After V1 lesions, patients have perceptual deficits (scotomata) in subregions of the visual field corresponding to the deficient V1 region. They claim not to notice and not to be able to identify stimuli presented at these locations in the visual field. However, if they are allowed to guess the stimulus, their recognition performance may be above chance, even if they subjectively appear to be only guessing. One explanation for this perceptual performance is that visual information from the visual thalamus can reach the rest of the cortex not only via V1, but also via direct projection to extrastriate cortex areas (see, e.g., Bullier and Kennedy 1983). Consistent with this, stimuli in blind regions of the visual field can induce extensive activation of the extrastriate visual cortex (MT+, V4/V8, LOC) in blindsight patients (Goebel et al. 2001).

8.3 Contents of Consciousness

Conscious perception cannot only be characterized at the level of a dichotomous distinction into conscious-unconscious, but also involves the representation of **conscious content**. This calls for the study of the representation of conscious content in the brain, all the more so because various theories of neural correlates make specific predictions about the effect of consciousness on representation of content (Crick and Koch 1998; Tononi and Edelman 1998; Dehaene and Naccache 2001).

Three different “coding levels” must be distinguished in the representation of consciousness contents.

8.3.1 Coding Levels

At the top level of encoding are contents of different **sensory modalities**. In PET and fMRI, the cortical processing of the different modalities is clearly separable (Binder et al. 1994; Tootell et al. 1996). At the middle level, the **representation of the submodalities** occurs. In the visual system, these would be, for example, color, brightness, motion, or object perception. They can usually be traced back to separate regions within the individual processing pathways.

Background Information

An early example of the assignment of visual submodalities to specific brain regions can be found in a PET study by Zeki et al. (1991). Using positron emission tomography, brain activity was measured in healthy subjects while they viewed visual presentations with different content.

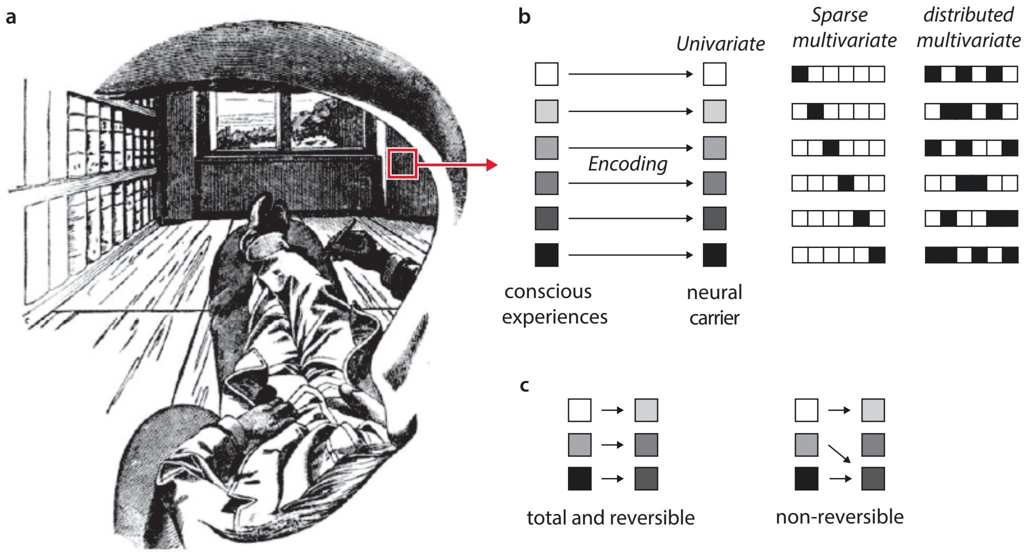
In addition, a closed-eye condition was measured, which served as a "baseline" to which the activities in the other conditions were compared. A brain region responsible for recognizing *motion* should show stronger activity for the moving stimuli than for the static stimuli. A region responsible for recognizing *color* should show stronger activity for viewing colored stimuli than for gray stimuli. This allowed the researchers to identify an area of movement they called V5 (now widely known as MT+). Similarly, at the ventral border of occipital and temporal cortex, they found a region that was more responsive to color stimuli, which they named V4. Although the designation of this region as "V4" is controversial (Hadjikhani et al. 1998), there is no doubt that a color-selective region is located in the ventral temporal cortex. This was the first evidence in humans of a specialization of the visual cortex for different submodalities.

Below the submodalities there is the **level of the specific contents**. Thus, in color perception, the specific color qualities (hue, saturation, brightness) that are present at a particular location in the visual field must be distinguished. In motion perception, the different directions and speeds of movement must be encoded, and in object perception, different shapes as well as exemplars and categories of objects must be encoded.

Another and more flexible way of exploring the contents of consciousness is offered by the class of "multistable stimuli", which allow different interpretations in perception. Most famous are the "Necker cube" (Necker 1832) and "My wife and mother-in-law" (Boring 1930). If one looks at the stimulus for a long time, the perception suddenly flips and one sees another possible interpretation. Since the two interpretations are based on the same physical stimulus, this allows changes in conscious perception to be studied without changes in physical stimulation. A variant of this multistable perception is binocular rivalry (Leopold and Logothetis 1996, 1999), in which conflicting images are presented to both eyes, with the effect that perception does not fuse these images but spontaneously switches back and forth between the two images. Single-derivation studies in awake monkeys (Leopold and Logothetis 1996) and fMRI studies in humans have shown that higher, content-specific cortical regions of the visual system change their activity in concert with conscious perception. Later, further evidence was added that even early stages of visual processing (V1, LGN) are involved in binocular rivalry (Tong and Engel 2001; Haynes et al. 2005a, b; Wunderlich et al. 2005), presumably due to altered representation of simple features of conscious perception such as edges and brightnesses. However, the role of attention in these early correlates has not been definitively established (Watanabe et al. 2011).

Binocular Rivalry

A type of multistable perception in which both eyes are presented with images that do not merge into a unified image, but rather the perception spontaneously switches back and forth between the two images.



▣ **Fig. 8.2** Coding principles for perceptual content. **a** A drawing by Ernst Mach (1886) in which he attempts to represent his visual experiences pictorially. **b** left: In a univariate code, ordinally scalable contents of consciousness (such as brightness) are represented in a continuous neural parameter (e.g., firing rate). Middle: In a *sparse* multivariate code, a separate, specialized neuron is active for each content (so-called grandmother or cardinal cell code). Right: in a distributed multivariate code, all neurons are involved in encoding all content. **c** The connection between consciousness contents and neuronal representations must meet certain mapping criteria. Totality means that a neural correlate can be identified for each content of consciousness. More important is reversibility, which means that no information is lost because each content of consciousness is assigned its own neural correlate and thus the content can be decoded again from the neural representation at any time. (Illustration from Haynes 2009)

8.3.2 Coding of Consciousness Contents

The more recent approach of multivariate decoding (► Sect. 8.3.3) makes it possible to investigate directly how certain contents of consciousness (colours, brightness, edges, movement, objects) are realised in the brain and how these representations change when people become conscious (Haynes 2009). For this purpose, the nervous system can be thought of as a kind of *carrier* into which the contents of consciousness are encoded. Encoding here does not mean that there is a homunculus that has to read out an encrypted code, but that there is a stable mapping relationship between states of consciousness (such as different sensations of brightness) and the states of the neuronal carrier. So you have to look for a neural carrier in the brain that has a stable association with some content, so that every time the content is in consciousness, the same state of the carrier goes with it (Haynes 2009).

► Example

Let us take as an example a drawing by Ernst Mach (1886). ▣ Figure 8.2a shows Mach's visual perceptual space as he sits on a sofa and looks out of his left eye into his study. The black and white drawing shows mainly edges and brightness. The brightness sensation in a region in the upper right quadrant of his visual field was presumably encoded in some parameter of neural activity of a particular brain area. One possibility would be that each brightness was encoded in the *activity rate of* individual cells in a region (say

V1 or V4) (see, e.g., Haynes et al. 2004), so that greater brightnesses would be associated with higher firing rates. This "univariate" code could be found by a simple correlation between brain activity level and perceptual intensity (Haynes et al. 2004). However, other coding formats are conceivable (◘ Fig. 8.2b). In a "multivariate" representation, multiple cells or cell populations are involved in the encoding, so that the content can no longer be explained by the activity of individual cells. In a "*sparse*" code, each perceptual content is associated with a dedicated cell, so that this cell (and only this cell) is always active when a person has a particular brightness experience. This representational format is also known as "grandmother cell" or "cardinal cell" code. In contrast, with a *distributed* multivariate representation, the assignment of experiences to individual cells is no longer possible at all; instead, a dedicated overall population activity state encodes each individual consciousness content. This code is called a *distributed* code. Importantly, sparse and distributed codes can no longer be explained by a simple correlation between activity in a region and perceptual content. ◄

Many studies have adopted a priori the **univariate encoding model** and searched for the neural "correlates" of specific perceptual content (e.g., Tong et al. 1998; Ress and Heeger 2003; Haynes et al. 2004). However, to identify any sparsely or distributedly encoded neural carriers, one must use dedicated analysis techniques, such as multivariate correlation, regression, and classification. The latter is also referred to as multivariate decoding and is presented in ► Sect. 8.3.3.

8.3.3 Multivariate Decoding

A general procedure for identifying neuronal populations that encode specific perceptual content is multivariate decoding (◘ Fig. 8.3). This involves determining how well a perceptual content can be reconstructed from a neuronal population response. If this works well, then the population has a lot of information about the content and there is a stable mapping relationship between brain states and a particular category of conscious content. Multivariate decoding can be driven by different types of population signals, i.e., multiple leads from single cells (Quiroga et al. 2005), fMRI voxel sets (Haynes and Rees 2006), or multiple EEG electrodes (Blankertz et al. 2003).

Multivariate Decoding

A procedure for determining how well a perceptual content can be reconstructed from a neural population response.

Voxel

Data point in a three-dimensional representation space, corresponds to a pixel in a two-dimensional representation space.

► Example

◘ Figure 8.3 shows the procedure using the example of an fMRI measurement. The aim is to determine in which cortical regions the content of consciousness is encoded when looking at one of the two vehicles shown on the top right (◘ Fig. 8.3a; Cichy et al. 2011a, b). First, one can start the analysis at a location in the visual system. One extracts there—averaged over a short measurement run—the activity level in the spatial neighborhood of this starting point while the subject is looking at the first car. Then one repeats this measurement a few times, while the subject sometimes sees the first car and sometimes the second car. Thus, one measures several "samples" of the local brain activity patterns for both images.

The individual measurements can be thought of as points in a high-dimensional coordinate system with as many dimensions

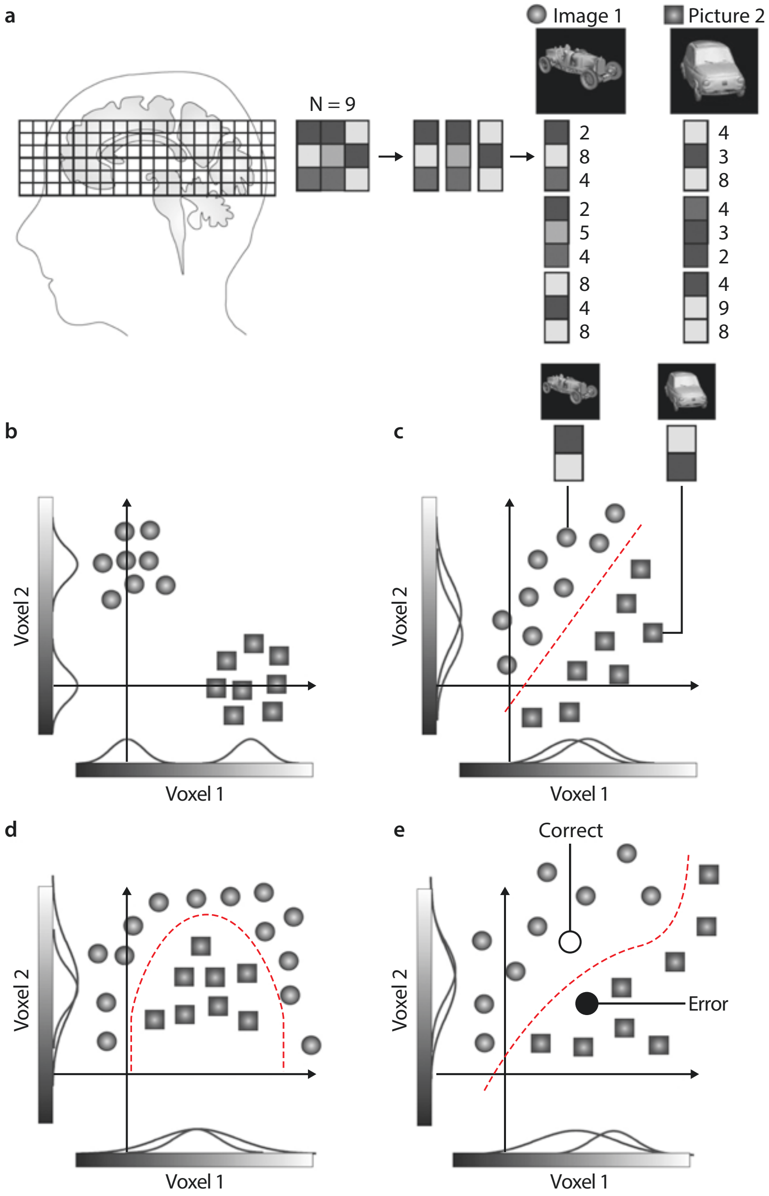


Fig. 8.3 Decoding of consciousness content using multivariate pattern recognition (see text). (Modified after Haynes and Rees 2006)

as voxels (so nine in this example). Since people find it difficult to imagine a 9-dimensional space, the analysis can be visualized for two dimensions. The measurements in voxel 1 can be plotted as *x-values* and the measurements in voxel 2 as *y-values* in a coordinate system. This would be a point (x, y) for each measurement resulting from the first two dimensions. Several possible coding examples are shown in the figure. ◘ Figure 8.3b shows a sparse code. High values in voxel 1 (x) occur only when image 2 is perceived, and high values in voxel 2 (y) occur only when image 1 is perceived. The perception of a content can always be attributed to a dedicated voxel, and the classification of whether the subject was looking at car 1 or car 2 during a given measurement would even be possible based on one voxel alone (if there were more than two objects, a sparse code in two dimensions would no longer be possible). Classification becomes more difficult in the case of ◘ Fig. 8.3c. Here, the assignment of consciousness contents to individual voxels is no longer possible. In this case the classification can also be done, but only if the values of both voxels are known at the same time. In this example, the decision can be made based on a linear separation between the two groups, but there may also be situations that have nonlinear decision boundaries and require the use of specialized nonlinear classification algorithms (◘ Fig. 8.3d).

To test whether the classification of consciousness content from a voxel population is possible, one uses a two-step procedure (◘ Fig. 8.3e). In the first step, one uses only a portion of the data, the **training dataset**, to train a classification algorithm to optimally separate the contents. The decision boundary can be learned via various algorithms, such as linear discriminant analysis or so-called support vector classification (Haynes and Rees 2006). In a second step, one then takes a statistically independent, new part of the data (*test data set*) and classifies it using the previously trained decoder. If the measurement points in the test data fall on the correct side of the decision boundary (◘ Fig. 8.3e, “correct”) and are thus correctly assigned, it is a hit. If they fall on the wrong side of the decision boundary (◘ Fig. 8.3e, “error”), it is a miss. The goodness with which the test data can be classified overrandomly well provides information about how much information about the perceptual content was contained in the voxel population. ◄

There are several ways to define the region used for classification. The **region** *of interest* (ROI) **method** uses an a priori defined region, such as the primary visual cortex, which can be defined on the basis of independent retinotopic mapping (Tootell et al. 1996). Alternatively, separate *localiser* runs can be considered, where a region is defined functionally (such as the fusiform face area, which is defined as the region that shows stronger responses to faces than to objects). The *searchlight* **method** starts with one point and notes the information content in the local environment around the starting point. The procedure is then repeated at many locations in the brain, creating a three-dimensional map showing the local information content at all locations in the brain (Haynes and Rees 2006). In the *whole brain* **method**, the voxels of the entire brain are used for classification. This last method is more commonly used in brain-computer interfaces that need to extract maximum information from brain signals for technical optimization. However, since the whole brain method is limited for clarifying regional hypotheses, it is less commonly used in cognitive neuroscience.

Multivariate decoding has been used in numerous studies to investigate the encoding of consciousness content (see Haynes 2009 for a review). In one study (Haynes and Rees 2005a), it was shown that subliminally presented masked orienting stimuli are nevertheless encoded in V1 in a feature-specific manner, confirming earlier results from psychophysics (He et al. 1996). Moreover, the encoding of specific consciousness cues can be clearly assigned to individual brain regions. Thus, different pictorial visual ideas

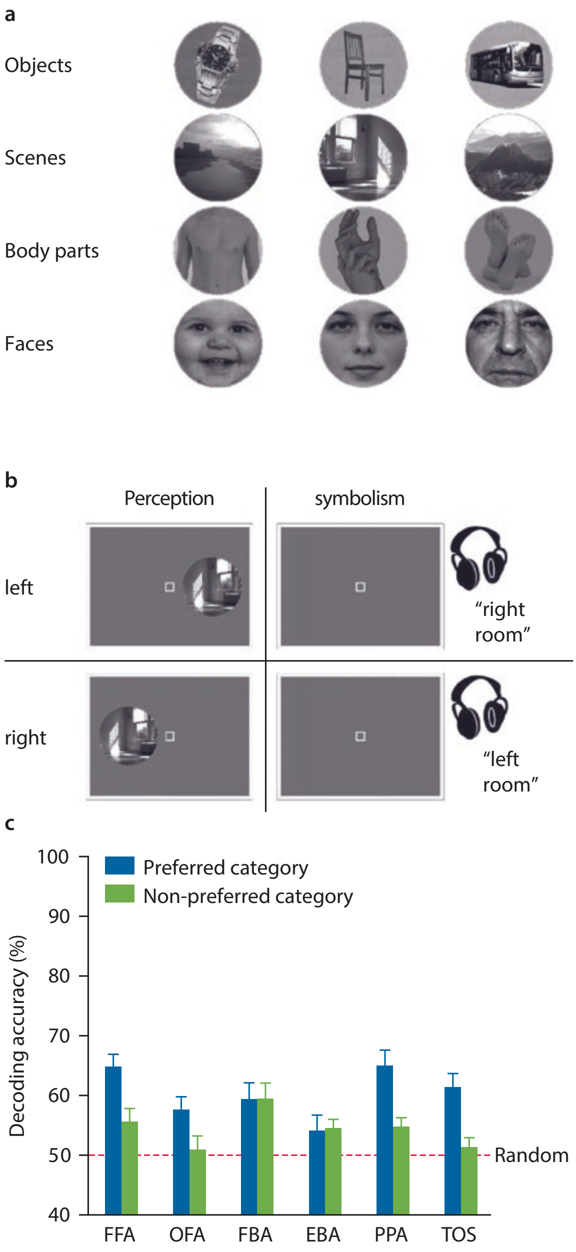


Fig. 8.4 Decoding of visual ideas from temporo-occipital brain regions (Cichy et al. 2011b). **a** Subjects were shown pictures of different categories. **b** During an fMRI measurement, they had to imagine them pictorially. **c** Although specialized regions had the most information about their optimal category (e.g., faces in the fusiform face area FFA), it was also possible to some extent to read out which of various non-preferred objects had been imagined (e.g., chair versus clock in the face area). Other abbreviations: *EBA* extrastriatal body area, *FBA* fusiform body area, *OFA* occipital face area, *PPA* parahippocampal place area, *TOS* transverse occipital sulcus

were found to be encoded in spatial brain activity patterns in specialized brain regions (Cichy et al. 2011b). However, it was also shown that other visual regions, apparently specialized in other ways, also represent the contents in question to a certain extent (Fig. 8.4).

An important property of higher-level conscious perception could also be investigated, namely invariance to presentation conditions (Cichy et al. 2011b). Subjects had to imagine objects in either the left or the right visual field. This made it possible to train pattern recognition on images in one hemifield and test it on another hemifield. The logic here is that object identity, as a higher-level feature, should be invariant to changes in the details of the presentation. Despite the change in position, it was possible to decode object category from patterns in the ventral path, arguing for a spatially invariant representation. In the ventral path, invariances to other stimulus properties, such as colors or textures, have also been demonstrated (Sáry et al. 1993).

8.4 Structure of Consciousness

Another important question is how similarities between consciousness contents are neuronally encoded. Let us take as an example two brightness gradations H_1 and H_2 of a grey area and their neuronal correlates N_1 and N_2. Put simply, the aforementioned strict correlation between consciousness content and brain activity means that whenever a person sees a certain brightness, the same neuronal activity is observed in the brain. In other words, H_1 would always measure N_1, H_2 would always measure N_2, and so on. A broader question is whether the *relations* and similarity ratios between consciousness contents also translate into the same relations and similarity ratios of activity patterns in the brain.

In the simplest case, this could mean, for example, the following: If area 2 is brighter than area 1 (i.e., $H_2 > H_1$), then the neuronal activity N_2 is also greater than N_1. The advantage of such an encoding would be that not only would the individual contents be explained by specific brain activities, but the perceived **similarity relations would** thus be automatically explained as well. Evidence for such an encoding exists in certain ases, however, according to imaging experiments, especially the brightness encoding in the visual system seems to be organized rather bipolar, namely in the form of deviations from an average gray value. Thus, if a surface is particularly dark or particularly bright, specialized neurons react more strongly to this than if the surface is gray (Haynes et al. 2004).

However, in other cases, particularly in object perception, there is strong evidence that the experienced similarities between perceptual states translate into similarities of neural representations. For example, in an early study Edelman et al. (1998) showed their subjects a series of images of different categories. They then asked the subjects to rate how similar they found the images. Based on the similarity judgments, the images were then arranged on a two-dimensional surface using multidimensional scaling in such a way that the relative positions of the objects in this space reflected their *subjectively perceived similarity to* each other (▣ Fig. 8.5a). In a second step, fMRI was used to measure brain activity patterns in the object perception cortex LOC for each object. Then, these activity patterns of different objects were compared with each other, and similarity was determined. Now, the objects were arranged with a multidimensional scaling such that the spatial proximity indicated the *similarity of their neural representation* (▣ Fig. 8.5b). It was found that the space of subjectively perceived similarities and the space of neural similarities reflect each other well. This means that the subjectively perceived similarity between objects can be explained by neural representations in this region, the LOC.

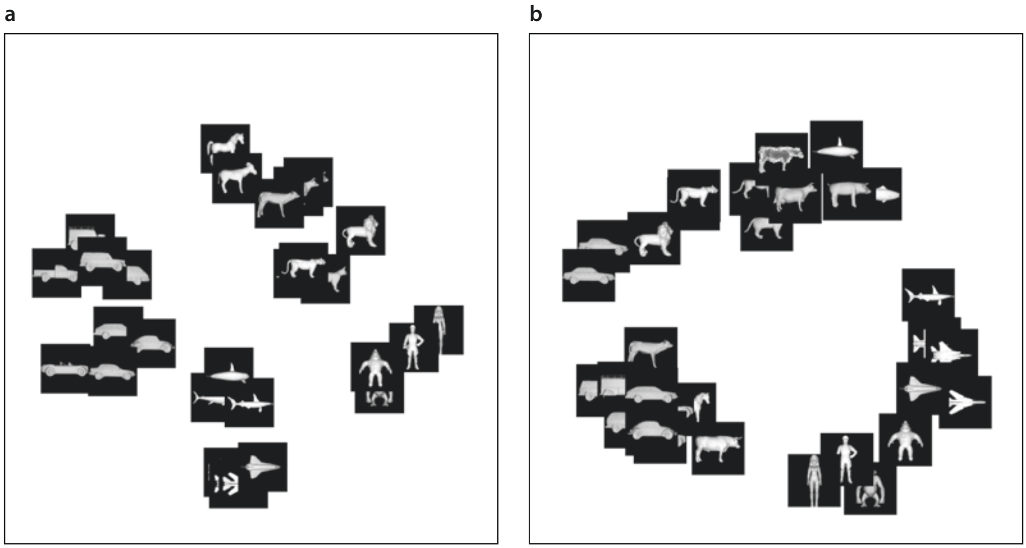


Fig. 8.5 Psychophysical and neuronal perceptual spaces. A comparison of the perceived similarities of objects **a** with their neuronal similarity in the lateral occipital complex (**b** see text). (Adapted from Edelman et al. 1998)

8.5 Models of Consciousness

A central question in research on the neural correlates of consciousness is what specific difference in the neural processing of a stimulus determines whether it enters consciousness or not. There are a number of candidates, and some evidence can be found for all of them. Some representative theories are briefly outlined below:

Activity Level

The core idea of this theory is that the awareness of a stimulus is tied to how strong the activity is in the brain that the stimulus triggers. Accordingly, a weak activation in a brain region may remain unconscious, but above a certain level of arousal, the neuronal representations it contains become conscious.

There is ample evidence to support this theory. The perceptual threshold for masked stimuli is directly correlated with the level of activity in object processing cortex, in monkeys (Kovács et al. 1995) and in humans (Grill-Spector et al. 2000). Already in primary visual cortex, there is evidence that activity level determines perception, even for small fluctuations in perception around threshold (Ress and Heeger 2003). However, work by Leopold and Logothetis (1996) shows that the link between activity and awareness may be more complex, at least in binocular rivalry. They found cells whose activity increased when the corresponding representation became conscious, but there were also cells whose activity dropped.

Activity theory can explain well an important property of conscious perception, namely its availability for access. We can respond to conscious stimuli intentionally, we can remember them, we can describe them verbally. This means that conscious stimuli undergo further processing in the brain. If the associated neural representations are “stronger” according to activity theory, it could also explain that they influence later regions more strongly.

Communication and Synchronization

Another and very popular model states that neuronal stimulus representations reach consciousness when the distributed subrepresentations synchronize their firing rate (Engel and Singer 2001). The cortical activity associated with this is thought to be primarily in the high-frequency range (gamma band, 30–80 Hz). Accordingly, consciousness would not be reducible to individual brain regions, but would be a network phenomenon. This model can explain particularly well why, despite the distributed, quasi-modular processing of individual feature dimensions (color, shape, motion), an object is nevertheless perceived as integrated. The explanation would be that the uniformity of perception can be attributed to the synchronization of neuronal partial representations. This synchronization theory could also well explain the "access property" of consciousness, because neurons that become synchronously active have a higher probability of suprathreshold excitation of downstream regions (König et al. 1996). However, the neurophysiological processes by which long-range, high-frequency synchronization is thought to occur remain unclear. Although neurons that fire in the gamma band are found in narrowly defined populations of neurons, such as in the visual cortex (V1, V4), these are inhibitory interneurons that act on pyramidal cells as the respective output neurons of the population. However, apart from specialized pyramidal cells, the latter fire at low frequency in the 1–10 Hz range and thus cannot transmit gamma acivity over a long range (cf. Buszáki and Schomburg 2015; Ray and Maunsell 2015). However, they can be "forced" by high-frequency input from interneurons to fire with low-frequency pulses during specific time windows. Coupling of closely neighboring populations via interneurons could lead to synchronous firing of pyramidal cells in neighboring populations, thereby enhancing certain activities, such as attentional, within circumscribed cortex areas (see Taylor et al. 2005; Fries 2015; Ni et al. 2016). Longer-range synchronization phenomena in the gamma band region, however, cannot yet be explained by this.

■ Recurrent Processing and Feedback

Another theory also relates to the **dynamics** of **long-range neural processes** (Lamme et al. 2000; Pascual-Leone and Walsh 2001) and had previously been used to explain masking phenomena. Super et al. observed that conscious recognition of texture-defined shapes is expressed by increased activity at late stages of processing in primary visual cortex (Supèr et al. 2001). They interpreted this late activity as a "recurrent" phase of processing on which forward, backward, and lateral processing overlap. Further evidence for the importance of feedback signals after V1 was provided by a paper by Pascual-Leone and Walsh (2001). They showed that disrupting backward projection after V1 to a late stage of stimulus processing can prevent conscious awareness of motion stimuli. A related perceptual model is the reverse hierarchy theory (Hochstein and Ahissar 2002). If recurrence should also be a necessary condition for conscious perception, feedback after V1 might not be required for it, since in certain cases conscious experiences of motion can occur even without V1 (Zeki and Ffytche 1998).

■ Specific Regions and Micro-awareness(es)

Semir Zeki (2001) formulated a theory according to which different modules in the visual cortex can produce independent "*micro-consciousness*". This is attributed, among other things, to the fact that lesions can be so selective that only a certain category of consciousness content is precipitated (e.g., color or face recognition). Further evidence is that different sensory feature dimensions (e.g., motion and color) reach consciousness at different times (Moutoussis and Zeki 1997). Most notable are the rare cases of patients with Riddoch syndrome (Zeki and Ffytche 1998). These patients have a lesion in the primary visual cortex and are blind because of it. However, their MT+/V5 motor area is intact, and they report being able to see dimly fast motion and can discriminate motion stimuli correctly. They describe movement as being like a "black shadow against a dark background" (Zeki and Ffytche 1998). This presumably means that even when visual perception is largely absent, an isolated dimension can remain intact and be accompanied by activity in the corresponding cortex area. Moreover, a content of consciousness without V1 means that feedback to V1 cannot be a necessary condition for visual consciousness, which contradicts the aforementioned V1 feedback theories of consciousness (Pascual-Leone and Walsh 2001; Lamme et al. 2000).

Dynamic Core

Tononi and Edelman (1998) formulated a theory that consciousness is equivalent to a "*dynamic* core" of brain activity. This was based on two observations: First, we experience our consciousness as integrated or "bound." Thus, we experience a visual scene as a single entity rather than as an unconnected collection of colors, brightnesses, and edges, even though these features are processed in different places in the brain. Similarly, we find it difficult to perform more than a single task at a time unless the processing is highly automated. The second observation is that consciousness is differentiated despite integration. We can have a variety of experiences, hence our consciousness contains a lot of "information".

They bring these two aspects together in their theory of the dynamic core. According to this theory, there is an information-theoretically irreducible core in brain activity that exhibits a high degree of integration and recurrence between posterior and anterior thalamocortical loops, thus linking sensory categorization with action planning, evaluation, and memory. Evidence for this comes from an experiment on binocular rivalry in which two images flickering at different frequencies are presented separately to the left and right eyes (Tononi et al. 1998). It showed a response strength and an increase in coherence between parietal and frontal MEG sensors at the frequency corresponding to the image when one image was conscious. A body of evidence, according to Tononi and Edelman (1998), suggests that the core process is dynamic, that is, it can alternately integrate different brain regions into a functional unit. This includes, in particular, the encapsulation of automated, unconscious processing, as expressed, for example, in the activation of extensive sensory cortex regions by unconscious stimuli.

The theory has been developed several times into a **theory of information integration** that gives explicit mathematical definitions of integration (Tononi 2004). In this, an attempt is made to also explain the diversity of qualitative dimensions of experience (qualia) within the same theoretical framework by stating that each dynamic unit spans its own coding space. The dynamic core theory explains the dynamism, integratedness and metastability of consciousness very well. One problem, however, is that the mathematical measures of integration are challenging to measure.

Global Workspace

An important property of conscious perception is that we can cognitively access conscious representations differently from unconscious representations. When we consciously recognize a visual stimulus, we know we have seen it and can remember it later, we can describe it verbally, or we can otherwise use it to control behavior, e.g., by pressing a response button. To explain this property, Baars formulated the theory of the global workspace (Baars 2002), which was then further developed by Dehaene and Naccache (2001). The idea is that visual stimulation that does not reach consciousness is informationally encapsulated in sensory cortex regions. Information about stimuli that do reach consciousness, on the other hand, is broadly "distributed" in the cortex, as if on a global worksurface where sensory information is available to multiple brain regions (such as for memory and action control) for readout. The *global workspace theory* (GWS; Baars 2002) is supported by numerous evidences. For example, there is direct evidence from Dehaene et al. (2001) that visually masked, unseen words activate only the sensory cortex. In contrast, consciously recognized words excite broad regions of the brain, including the prefrontal cortex (Dehaene et al. 2001). Baars (2002) lists a number of examples where it has been shown that conscious stimuli undergo more extensive processing. These include the extensive cortical distribution of information about binocular rivalry stimuli when they enter consciousness (Tononi et al. 1998), or the greater frontoparietal activation when small changes in dynamic visual displays are detected compared to the situation in which they are not detected (Beck et al. 2001).

■ Specific Regions

A number of authors have discussed the importance of specific regions for visual consciousness. Crick and Koch (1995) argued for V1 not being directly involved in conscious processes. Among the main arguments for this is that V1 has no direct projections to the prefrontal cortex, which is incompatible with the availability of conscious information for complex behavioral control. Moreover, V1 is also activated by unconscious stimuli (He et al. 1996; Haynes and Rees 2005a, b). It would be difficult to explain why V1 should be involved in consciousness and yet these stimulus representations do not reach consciousness. Goodale and Milner (1992) argued against involvement of the parietal-dorsal visual processing pathway in visual consciousness. This was based on a double dissociation in the behavioral profile of two patients who had a lesion in the dorsal and ventral pathways, respectively. One patient with a dorsal visual lesion was able to describe a visual stimulus but was unable to use it to control behavior. In contrast, a patient with a lesion in the ventral visual pathway was able to incorporate the stimulus into behavioral programs but was unable to describe it. The authors interpreted this as evidence that only the ventral visual pathway is conscious.

Overview

An assessment of the various theories of consciousness is complicated by the fact that they are good at explaining different aspects of consciousness. Some theories focus on the availability of consciousness content for multiple complex behavioral performances. Activation theory and synchronization theory predict that subsequent regions are more amenable to suprathreshold arousal and explain this in terms of greater activity or synchronization of conscious neural representations. Both aspects are in principle compatible with global workspace theory, although other physiological realizations have been proposed by the authors. Moreover, in empirical data, activation and synchronization (and presumably recurrence) are strongly correlated in most cases (e.g., Siegel and König 2003), making it difficult to distinguish between the theories.

The theories mentioned in this section make very different predictions about what happens to content-specific neural representations during conscious versus unconscious processing (◘ Fig. 8.6). For example, according to the microconsciousness theory, it is sufficient for awareness to occur if a neural representation arises in a dedicated perceptual area. For example, content-specific activation in the fusiform face area would already automatically lead to face perception

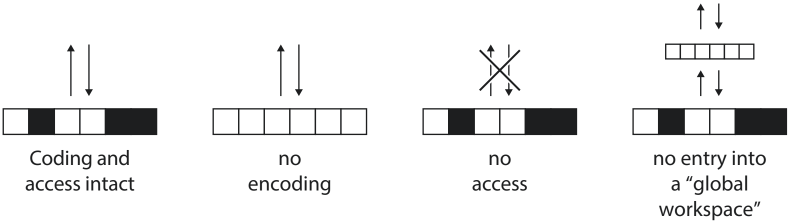


Fig. 8.6 Contents of consciousness in different models of consciousness. (Adapted from Haynes 2009). From left: In the state of normal awareness, both the stimulus representation and the access to the stimulus representation are intact. If a stimulus is not consciously perceived, there could be several reasons for this: A lack of encoding of the content in content-specific brain regions, a lack of attional or executive access to the information despite an intact representation, no long-range distribution of the information in the sense of a “global workspace”

without requiring additional conditions such as attention. In contrast, for theories that focus on *access* to visual representations, a selection or distribution mechanism must be operative in addition to a representation. Accordingly, it could happen that a content is encoded in sensory regions but does not reach consciousness.

To distinguish between these two theories, it would be important to investigate whether encoding in sensory regions actually necessarily leads to awareness of the stimulus. Some experiments have shown that even unconscious stimuli can lead to content-specific activation in the visual system (Rees et al. 2002; Moutoussis and Zeki 2002; Haynes and Rees 2005a). Moreover, awareness leads to more widespread activations outside sensory brain regions as well. This suggests that additional processes are required to bring a representation into consciousness. However, whether these additional processes operate in the sense of a global workspace is currently unclear. This theory suggests that content must be distributed widely in the brain in order to become conscious.

Accordingly, not only would one have to find widespread activity in the brain upon awareness of a stimulus, but these processes would have to be shown to be content-specific and, in fact, to encode *information* about the sensory process. More recently, advocates of the different positions on consciousness have started to engage in adversarial collaborations in order to define crucial experiments that will help decide between the different views (Melloni et al. 2023).

8.6 Awareness, Selection and Attention

According to a widely held view, selective attention regulates access to consciousness and is, in effect, the selective gatekeeper of consciousness. This would mean that we can only consciously recognize stimuli when we attend to them.

There is indeed some evidence for a close connection between consciousness and attention. The clearest evidence is provided by so-called cueing experiments (Sperling 1960; Posner et al. 1980; Hawkins et al. 1990). Experiments by Hawkins et al. (1990), for example, showed that the accuracy with which target stimuli presented at different positions could be recognized increased when subjects were given a cue indicating the most likely target position beforehand. Earlier, in a similar experiment, George Sperling (1960) had shown subjects random sequences of letters and asked them to recall as many letters as possible afterwards. It turned out that their hit rate was considerably better when the subjects were given a cue as to which rows they should report, even when this cue was given only *after the* row of letters had been removed. Accordingly, attention can even operate on representations in iconic memory and select its contents for conscious perception.

Another example of the close relationship between awareness and attention is provided by experiments on inattentional blindness (*Inattentional Blindness*; Mack and Rock 1998) and change blindness (*Change* Blindness; McConkie and Currie 1996; Rensink et al. 1997; Niedeggen et al. 2001; Beck et al. 2001). In inattentional blindness, subjects complete a few runs of a difficult fixation task, and in a final, crucial single run, a stimulus is presented in the periphery of the visual field during the fixation task. Few subjects report seeing the unexpected peripheral stimulus at all (Mack and Rock 1998). The lack of conscious recognition of the peripheral stimulus is attributed to lack of attention. In the very similar phenomenon of change blindness, even striking changes in visual images are not detected until they are processed with attention.

However, there is also evidence in these experiments that attention does not automatically lead to consciousness. Thus, a similar "blindness" can also occur with fully attended stimuli.

► Example

In one experiment, subjects were engaged in fake conversations with a stranger (Simons and Levin 1998). The conversation was briefly interrupted by two construction workers carrying a door between them so that their view was briefly obscured. Unbeknownst to the subject, the interlocutor was exchanged, which was not noticed by numerous subjects. Since the subject had been looking at the interlocutor (with only one minor interruption) with full attention, this argues that the faulty recognition was not so much due to a lack of attention but to a lack of *expectation* that an interlocutor's identity might change. Other experiments show that attention can also affect the processing of unconscious stimuli (e.g., Martens et al. 2011). Together with the other experiments elaborated above, this would imply that attention is a necessary but not a sufficient condition for consciousness. ◄

However, in recent years some dissociations have been elaborated suggesting that attention is not necessary for consciousness either, because consciousness can occur without attention (Lamme 2003; Koch and Tsuchiya 2007; Srinivasan 2008). A number of experimental findings lead to this conclusion:

1. *Some stimulus properties can be recognized even without attention*. Even without attentional guidance (*Full Report*), the recognition rate in Sperling's experiments was well above chance (Sperling 1960), which could argue for residual processing, presumably by the subject's spontaneous attention. On the other hand, it could also mean that minor processing is possible even in the absence of attention. This is confirmed by studies with dual tasks (Braun and Julesz 1998). Here, subjects' attention is bound by a difficult primary discrimination task at

the fixation point. The extent to which additional peripheral stimuli can be detected without diverting attention from the central task is then investigated. It is first shown that simple stimulus features (such as color) can be detected without attentional cost, whereas complex stimulus features (such as shape) cannot. This suggests that conscious recognition of simple features is not dependent on attention. Later, it was shown that even the core content of extended and complex visual scenes can be reproduced without attention (Li et al. 2002).

2. *Attention can be a hindrance to awareness*. Awareness can in some cases be improved by paying less attention to the target stimulus (Yeshurun and Carrasco 1998; Olivers and Nieuwenhuis 2005). For example, Yeshurun and Carrasco (1998) have shown that foveal texture detection can be degraded by attention, presumably because processing of coarse-resolution shape information suffers from the increased spatial resolution induced by attention.
3. *Attention and consciousness can have opposite effects on visual processing*. Some experiments suggest a double dissociation between the effects that attention and conscious stimulus detection have on visual information processing. For example, a number of studies on afterimages showed that attention shortens their duration, whereas consciousness can increase their duration (see Koch and Tsuchiya 2007 for a review), although there are also opposite findings (Kaunitz et al. 2011).
4. *Attention can also be controlled by unconscious stimuli*. For example, McCormick (1997) directly compared attentional control by suprathreshold and subliminal stimuli in a study. It was shown that even invisible stimuli can trigger exogenous orienting of attention (see Mulckhuyse and Theeuwes 2010 for a detailed review). Attention can even be guided by complex properties of invisible stimuli, which suggests that subliminal stimuli and can be processed deep enough to control attention (Jiang et al. 2006). Interestingly, the effect may depend on the unconscious stimulus that attracts attention being relevant to the subject's task (Ansorge et al. 2010). Taken together, then, the above points show a clear dissociability between attention and consciousness, even though the two processes are usually closely intertwined.

Summary

In this chapter, we first considered the psychological conditions of the occurrence of conscious awareness and found that it is not an all-or-nothing phenomenon, but that there is a gradual transition between completely unconscious and clearly conscious stimuli: Thus, some features of a stimulus may be recognized but not others, or seemingly completely unconscious "masked" stimuli may have a marked influence on choice decisions between alternative consciously presented stimuli (priming). There is evidence that certain stimuli are not consciously perceived at all, but nevertheless influence our behavior, for example in unconscious learning and conditioning tasks.

The crucial cognitive neuroscience question is whether there are stable differences in the brain activities evoked by unconscious or conscious stimuli. It has been established that unconscious stimuli can activate not only low-level but also high-level processing areas of the brain, e.g. the temporal face recognition area (fusiform gyrus), but that activation by conscious stimuli is always more extensive and involves a larger number of cortical areas. This mainly concerns the so-called fronto-parietal network, which extends between the superior frontal cortex (PFC) and the posterior parietal cortex (PPC). However, even in primary sensory areas, e.g. the primary visual cortex,

responses to stimuli that become conscious are stronger than those that remain unconscious.

Neuroscience has also addressed the question whether there are discrete single-unit representations, called “cardinal cells”, not only for very simple stimuli, but also for certain complex contents such as faces, persons and situations, or whether such contents are encoded in a distributed code involving activity patterns of wide cortical populations of neurons. Recent research has confirmed the latter view for many domains as the rule using the method of multivariate decoding. It shows that even in the case of distributed coding, the same perceptual contents give rise to the same neuronal representations.

Another central question of neuroscientific consciousness research is to what extent differences and similarities between contents of consciousness are also expressed in differences and similarities of neuronal coding, i.e. whether the structure of our experience is reflected in the relational structure of our brain activity patterns. Subjectively perceived similarities are indeed expressed in similar neural responses. Thus, the link between subjective experience and brain activity can be established also at the level of detailed contents.

Consciousness and attention are usually closely linked: Attention acts to amplify the neural responses to sensory stimuli. Attention and awareness are not identical, however, because we can miss information in stimuli even when they are fully focused. This can happen, for example, when the information was not expected.

Neurotheorists and neurophilosophers have developed numerous models for the emergence of consciousness and its relationship to brain processes. Some authors assume that everything that affects the brain strongly enough and for long enough also penetrates into consciousness. Other authors assume that this requires “recurrent” activity between primary sensory areas and associative areas in the cortex. According to one older model such areas “bind together” in a specific way using high-frequency synchronization in the so-called gamma range (30–80 Hz) to produce meaningful conscious content. This can also be integrated with the idea that there is a “core mechanism” underlying all consciousness content that is active between the cortex and thalamus and constantly aligns with new current content. The various models can explain many aspects of consciousness well, but by no means all of them. According to a different view, consciousness can be understood as a “global workspace” involving special mechanisms of information processing, enabling current detailed perception, its memorability and linguistic reportability, and greater access to behavioral responses, including complex action planning.

Current neuroscience provides strong evidence against so-called dualistic positions, according to which there is an independence of brain and mind. Rather, the contents of consciousness are precisely related to neuronal processes in the brain.

References

Adams JK (1957) Laboratory studies of behavior without awareness. Psychol Bull 54:383–405

Ansorge U, Horstmann G, Worschech F (2010) Attentional capture by masked colour singletons. Vis Res 50(19):2015–2027. (Epub 2010 Jul 24)

Baars BJ (2002) The conscious access hypothesis: origin and recent evidence. Trends Cogn Sci 8:47–52

Beck DM, Rees G, Frith CD, Lavie N (2001) Neural correlates of change detection and change blindness. Nat Neurosci 4(6):645–650

Berti A, Rizzolatti G (1992) Visual processing without awareness: evidence from unilateral neglect. J Cogn Neurosci 4:345–351

Binder JR, Rao SM, Hammeke TA, Yetkin FZ, Jesmanowicz A, Bandettini PA, Wong EC, Estkowski LD, Goldstein MD, Haughton VM et al (1994) Functional magnetic resonance imaging of human auditory cortex. Ann Neurol 35(6):662–672

Blankertz B, Dornhege G, Schäfer C, Krepki R, Kohlmorgen J, Müller KR, Kunzmann V, Losch F, Curio G (2003) Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. IEEE Trans Neural Syst Rehabil Eng 11(2):127–131

Boring EG (1930) A new ambiguous figure. Am J Psychol 42:444

Braun J, Julesz B (1998) Withdrawing attention at little or no cost: detection and discrimination tasks. Percept Psychophys 60(1):1–23

Bullier J, Kennedy H (1983) Projection of the lateral geniculate nucleus onto cortical area V2 in the macaque monkey. Exp Brain Res 53(1):168–172

Buszáki G, Schomburg EW (2015) What does gamma coherence tell us about interregional neural communication? Nat Neurosci 18:484–489

Cannon MW (1984) A study of stimulus range effects in free modulus magnitude estimation of contrast. Vis Res 24(9):1049–1055

Cichy RM, Chen Y, Haynes JD (2011a) Encoding the identity and location of objects in human LOC. NeuroImage 54(3):2297–2307

Cichy RM, Heinzle J, Haynes JD (2011b) Imagery and perception share cortical representations of content and location. Cereb Cortex 22:372–380

Crick F, Koch C (1995) Are we aware of neural activity in primary visual cortex? Nature 375(6527):121–123

Crick F, Koch C (1998) Consciousness and neuroscience. Cereb Cortex 8(2):97–107

Debner JA, Jacoby LL (1994) Unconscious perception: attention, awareness, and control. J Exp Psychol Learn Mem Cogn 20(2):304–317

Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. Cognition 79(1–2):1–37

Dehaene S, Naccache L, Cohen L, Bihan DL, Mangin JF, Poline JB, Rivière D (2001) Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci 4(7):752–758

Edelman S, Grill-Spector K, Kushnir T, Malach R (1998) Towards direct visualization of the internal shape space by fMRI. Psychobiology 26:309–321

Engel AK, Singer W (2001) Temporal binding and the neural correlates of sensory awareness. Trends Cogn Sci 5(1):16–25

Eriksen CW (1954) The case for perceptual defense. Psychol Rev 61:175–182

Eriksen CW (1960) Discrimination and learning without awareness: a methodological survey and evaluation. Psych Rev 67:279–300

Fang F, He S (2005) Cortical responses to invisible objects in the human dorsal and ventral pathways. Nat Neurosci 8(10):1380–1385. (Epub 2005 Sep 4)

Fechner GT (1860) Elemente der Psychophysik. Breitkopf & Härtel, Leipzig

Fries P (2015) Rhythmus for cognition: communication for coherence. Neuron 88:220–235

Gescheider GA (1988) Psychophysics: the fundamentals. Annu Rev Psychol 39:169–200

Gescheider G (1997) Psychophysics: the fundamentals. Erlbaum, New Jersey

Goebel R, Muckli L, Zanella FE, Singer W, Stoerig P (2001) Sustained extrastriate cortical activation without visual awareness revealed by fMRI studies of hemianopic patients. Vis Res 41(10–11):1459–1474

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15(1):20–25

Green DM, Swets JA (1966) Signal detection theory and psychophysics. Wiley, New York

Grill-Spector K, Kushnir T, Hendler T, Malach R (2000) The dynamics of object-selective activation correlate with recognition performance in humans. Nat Neurosci 3(8):837–843

Gur M, Snodderly DM (1997) A dissociation between brain activity and perception: chromatically opponent cortical neurons signal chromatic flicker that is not perceived. Vis Res 37(4):37782

Hadjikhani N, Liu AK, Dale AM, Cavanagh P, Tootell RB (1998) Retinotopy and color sensitivity in human visual cortical area V8. Nat Neurosci 1(3):235–241

Hawkins HL, Hillyard SA, Luck SJ, Mouloua M, Downing CJ, Woodward DP (1990) Visual attention modulates signal detectability. J Exp Psychol Hum Percept Perform 16:802–811

Haynes JD (2009) Decoding visual consciousness from human brain signals. Trends Cogn Sci 13(5):194–202

Haynes JD, Rees G (2005a) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8(5):686–691

Haynes JD, Rees G (2005b) Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15(14):1301–1307

Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7(7):523–534

Haynes JD, Lotto RB, Rees G (2004) Responses of human visual cortex to uniform surfaces. Proc Natl Acad Sci U S A 101(12):4286–4291

Haynes JD, Deichmann R, Rees G (2005a) Eye-specific effects of binocular rivalry in the human lateral geniculate nucleus. Nature 438(7067):496–499

Haynes JD, Driver J, Rees G (2005b) Visibility reflects dynamic changes of effective connectivity between V1 and fusiform cortex. Neuron 46(5):811–821

He S, Cavanagh P, Intriligator J (1996) Attentional resolution and the locus of visual awareness. Nature 383(6598):334–337

Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36(5):791–804

Holender D (1986) Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: a survey and appraisal. Behav Brain Sci 9:1–23

Jiang Y, Costello P, Fang F, Huang M, He S (2006) A gender- and sexual orientation-dependent spatial attentional effect of invisible images. Proc Natl Acad Sci U S A 103(45):17048–17052. (Epub 2006 Oct 30)

Kahnt T, Grueschow M, Speck O, Haynes JD (2011) Perceptual learning and decision-making in human medial frontal cortex. Neuron 70(3):549–559

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17(11):4302–4311

Kaunitz L, Fracasso A, Melcher D (2011) Unseen complex motion is modulated by attention and generates a visible aftereffect. J Vis 11(13):10. https://doi.org/10.1167/11.13.10

Koch C, Tsuchiya N (2007) Attention and consciousness: two distinct brain processes. Trends Cogn Sci 11(1):16–22

König P, Engel AK, Singer W (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. Trends Neurosci 19(4):130–137

Kouider S, Dupoux E (2004) Partial awareness creates the “illusion” of subliminal semantic priming. Psychol Sci 15:75–81

Kovács G, Vogels R, Orban GA (1995) Cortical correlate of pattern backward masking. Proc Natl Acad Sci U S A 92(12):5587–5591

Kunimoto C, Miller J, Pashler H (2001) Confidence and accuracy of near-threshold discrimination responses. Conscious Cogn 10:294–340

Lamme VA (2003) Why visual attention and awareness are different. Trends Cogn Sci 7(1):12–18

Lamme VA, Supèr H, Landman R, Roelfsema PR, Spekreijse H (2000) The role of primary visual cortex (V1) in visual awareness. Vis Res 40(10–12):1507–1521

Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys’ percepts during binocular rivalry. Nature 379(6565):549–553

Leopold DA, Logothetis NK (1999) Multistable phenomena: changing views in perception. Trends Cogn Sci 3(7):254–264

Li FF, VanRullen R, Koch C, Perona P (2002) Rapid natural scene categorization in the near absence of attention. Proc Natl Acad Sci U S A 99(14):9596–9601

Lumer ED, Rees G (1999) Covariation of activity in visual and prefrontal cortex associated with subjective visual perception. Proc Natl Acad Sci U S A 96(4):1669–1673

Mach E (1886) Beiträge zur Analyse der Empfindungen. Fischer, Berlin

Mack A, Rock I (1998) Inattentional blindness. MIT Press, Cambridge

Marcel AJ (1983) Conscious and unconscious perception: experiments on visual masking and word recognition. Cogn Psychol 15(2):197–237

Martens U, Ansorge U, Kiefer M (2011) Controlling the unconscious: attentional task sets modulate subliminal semantic and visuomotor processes differentially. Psychol Sci 22(2):28291

Melloni L, Mudrik L, Pitts M, Bendtz K, Ferrante O, Gorska U, Hirschhorn R, Khalaf A, Kozma C, Lepauvre A, Liu L, Mazumder D, Richter D, Zhou H, Blumenfeld H, Boly M, Chalmers DJ, Devore S, Fallon F, de Lange FP, Jensen O, Kreiman G, Luo H, Panagiotaropoulos TI, Dehaene S, Koch C, Tononi G (2023) An adversarial collaboration protocol for testing contrasting predictions of global neuronal workspace and integrated information theory. PLoS One. 18(2):e0268577. https://doi.org/10.1371/journal.pone.0268577. PMID: 36763595; PMCID: PMC9916582

McConkie GW, Currie CB (1996) Visual stability across saccades while viewing complex pictures. J Exp Psychol Hum Percept Perform 22(3):563–581

McCormick PA (1997) Orienting attention without awareness. J Exp Psychol Hum Percept Perform 23(1):168–180

McGinnies E (1949) Emotionality and perceptual defense. Psychol Rev 56:244–251

Merikle PM, Joordens S, Stolz JA (1995) Measuring the relative magnitude of unconscious influences. Conscious Cogn 4(4):422–439

Moutoussis K, Zeki S (1997) A direct demonstration of perceptual asynchrony in vision. Proc Biol Sci 264(1380):393–399

Moutoussis K, Zeki S (2002) The relationship between cortical activation and perception investigated with invisible stimuli. Proc Natl Acad Sci U S A 99(14):9527–9532

Mulckhuyse M, Theeuwes J (2010) Unconscious attentional orienting to exogenous cues: a review of the literature. Acta Psychol 134(3):299–309

Necker LA (1832) Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. Lond Edinb Philos Mag J Sci 1(5):329–337

Ni J, Wunderle T, Lewis CM, Desimone R, Diester I, Fries R (2016) Gamma-rhythmic gain modulation. Neuron 92:240–251

Niedeggen M, Wichmann P, Stoerig P (2001) Change blindness and time to consciousness. Eur J Neurosci 14(10):1719–1726

Olivers CN, Nieuwenhuis S (2005) The beneficial effect of concurrent task-irrelevant mental activity on temporal attention. Psychol Sci 16(4):265–269

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. Ann Rev Neurosci 21:227–277

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

Peirce CS, Jastrow J (1884) On small differences in sensation. Mem Natl Acad Sci 3:73–83

Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD (2008) Subliminal instrumental conditioning demonstrated in the human brain. Neuron 59(4):561–567

Posner MI, Snyder CR, Davidson BJ (1980) Attention and the detection of signals. J Exp Psychol Gen 109:160–174

Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. Nature 435(7045):1102–1107

Ray RD, Maunsell JH (2015) Do gamma oscillations play a role in cerebral cortex? Trends Cogn Sci 19:78–85

Rees G, Wojciulik E, Clarke K, Husain M, Frith C, Driver J (2002) Neural correlates of conscious and unconscious vision in parietal extinction. Neurocase 8(5):387–393

Reingold EM, Merikle PM (1988) Using direct and indirect measures to study perception without awareness. Percept Psychophys 44:563–575

Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: the need for attention to perceive changes in scenes. Psychol Sci 8:368–373

Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. Nat Neurosci 6(4):414–420

Sáry G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. Science 260(5110):995–997

Schmidt T, Vorberg D (2006) Criteria for unconscious cognition: three types of dissociation. Percept Psychophys 68(3):489–504

Sidis B (1898) The psychology of suggestion. Appleton, New York

Siegel M, König P (2003) A functional gamma-band defined by stimulus-dependent synchronization in area 18 of awake behaving cats. J Neurosci 23(10):4251–4260

Simons DJ, Levin DT (1998) Failure to detect changes to people in a real-world interaction. Psychon Bull Rev 5:644–649

Sperling G (1960) The information available in brief visual presentations. Psychol Monogr 74:1–29

Srinivasan N (2008) Interdependence of attention and consciousness. Prog Brain Res 168:65–75

Supèr H, Spekreijse H, Lamme VA (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat Neurosci 4(3):304–310

Taylor K, Mandon S, Freiwald WA, Kreiter AK (2005) Coherent oscillatory activity in monkey area v4 predicts successful allocation of attention. Cereb Cortex 15:1424–1437

Tong F, Engel SA (2001) Interocular rivalry revealed in the human cortical blind-spot representation. Nature 411(6834):195–199

Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21(4):753–759

Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 2(5):42

Tononi G, Edelman GM (1998) Consciousness and complexity. Science 282(5395):1846–1851

Tononi G, Srinivasan R, Russell DP, Edelman GM (1998) Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. Proc Natl Acad Sci U S A 95(6):3198–3203

Tootell RB, Dale AM, Sereno MI, Malach R (1996) New images from human visual cortex. Trends Neurosci 19(11):481–489

Vuilleumier P, Sagiv N, Hazeltine E, Poldrack RA, Swick D, Rafal RD, Gabrieli JD (2001) Neural fate of seen and unseen faces in visuospatial neglect: a combined event-related functional MRI and eventrelated potential study. Proc Natl Acad Sci U S A 98(6):3495–3500

Vuilleumier P, Armony JL, Clarke K, Husain M, Driver J, Dolan RJ (2002) Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. Neuropsychologia 40(12):2156–2166

Watanabe T, Náñez JE, Sasaki Y (2001) Perceptual learning without perception. Nature 413(6858):844–848

Watanabe M, Cheng K, Murayama Y, Ueno K, Asamizuya T, Tanaka K, Logothetis N (2011) Attention but not awareness modulates the BOLD signal in the human V1 during binocular suppression. Science 334(6057):829–831

Watson AB (1986) Temporal sensitivity. In: Boff K, Kaufman L, Thomas J (eds) Handbook of perception and human performance. Wiley, New York

Weiskrantz L (2004) Roots of blindsight. Prog Brain Res 144:229–241

Weiskrantz L, Warrington EK, Sanders MD, Marshall J (1974) Visual capacity in the hemianopic field following a restricted occipital ablation. Brain 97(4):709–728

Wunderlich K, Schneider KA, Kastner S (2005) Neural correlates of binocular rivalry in the human lateral geniculate nucleus. Nat Neurosci 8(11):1595–1602

Yeshurun Y, Carrasco M (1998) Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396(6706):72–75

Zeki S (2001) Localization and globalization in conscious vision. Annu Rev Neurosci 24:57–86

Zeki S, Ffytche DH (1998) The Riddoch syndrome: insights into the neurobiology of conscious vision. Brain 121(Pt 1):25–45

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11(3):641–649