



Neural events and perceptual awareness

Nancy Kanwisher*

Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, USA

Received 18 December 1999; accepted 27 September 2000

Abstract

Neural correlates of perceptual awareness, until very recently an elusive quarry, are now almost commonplace findings. This article first describes a variety of neural correlates of perceptual awareness based on fMRI, ERPs, and single-unit recordings. It is then argued that our quest should ultimately focus not on mere correlates of awareness, but rather on the neural events that are both necessary and sufficient for perceptual awareness. Indeed, preliminary evidence suggests that although many of the neural correlates already reported may be necessary for the corresponding state of awareness, it is unlikely that they are sufficient for it. The final section considers three hypotheses concerning the possible sufficiency conditions for perceptual awareness. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Neural events; Perceptual awareness; Correlates of awareness

1. Introduction

The quest for the neural correlates of consciousness (Crick & Koch, 1995), or at least the neural correlates of perceptual awareness, has suddenly become wildly successful. A variety of striking correlations have been reported in just the last few years between specific neural signals and perceptual experiences. But the success of this enterprise leads to a much more difficult question: now that we have found a set of neural correlates of perceptual awareness, what are we to do with them? What if anything do they tell us about awareness?

It is helpful to consider what exactly it is that we want to understand about perceptual awareness in the first place. If the scientific investigation of awareness

* Fax: +1-617-253-9767.

E-mail address: ngk@psyche.mit.edu (N. Kanwisher).

is different from the scientific investigation of perception, then the two phenomena must not be identical. (In keeping with the possibility that they are distinct, the word ‘perception’ will be used throughout this article to refer to the extraction and/or representation of perceptual information from a stimulus, without any assumption that such information is necessarily experienced consciously.) So the most basic question is whether all perception is accompanied by awareness, or whether the two phenomena can be uncoupled. Extensive evidence from behavioral studies of both normal subjects (see Merikle, Smilek, & Eastwood in this volume) and neurological patients (Farah, 1994; Milner & Rugg, 1992) shows that perceptual information can indeed be represented in the mind/brain without the subject being aware of that information. This fact opens up for exploration a broad landscape of additional questions. What subset of the information that is perceived reaches awareness? More pointedly, what factors determine which information reaches awareness and which information does not? Is awareness of a perceptual representation a simple monotonic increasing function of the strength or quality (Baars, 1988; Farah, 1994) of the underlying representation (the ‘activation strength hypothesis’)? How is information within awareness represented and processed differently from information that is not within awareness?

In this article a number of recent studies will be reviewed that use neurophysiological techniques (fMRI, ERPs, and single-unit recording) to investigate these questions. Section 2 describes studies demonstrating neural signals that are strongly correlated with the content of the subject’s awareness under conditions in which the stimulus itself does not change. These findings then lead to a consideration of whether the neural correlates of awareness are localized in a particular location (or set of locations) in the brain that play some special role in awareness. I hypothesize to the contrary that the neural correlates of awareness of a particular visual attribute are found in the very neural structure that perceptually analyzes that attribute. Section 3 describes several recent studies using fMRI and ERPs that show that many of the same regions that show strong correlations with awareness under some conditions can also be activated in the absence of the subjects’ awareness of the stimulus. Results of this kind argue that activations in these regions may not be sufficient for awareness. This raises the question of what is needed beyond the mere existence of a neural representation for that representation to be experienced consciously. In Section 4 several possible answers to this question are considered. I argue – contrary to the activation strength hypothesis – that even a strong neural representation may not be sufficient for awareness unless other parts of the mind/brain have access to the information so represented (see also Baars, 1988). Behavioral evidence is presented that perceptual awareness involves not only activation of the relevant perceptual properties, but the further construction of an organized representation in which these visual properties are attributed to their sources in external objects and events (see also Kahneman & Treisman, 1984; Marcel, 1983).

I hope in this article to show that scientific evidence can bear importantly on a number of questions about the nature of perceptual awareness. However, it probably can not answer all such questions. In particular, I will not tackle the question of why perceptual awareness *feels* like anything at all (Chalmers, 1995; Nagel, 1974),

because it is not clear that even a rich understanding of the cognitive and neural events that constitute perceptual awareness will provide any clues about how to answer it.

2. Neural correlates of perceptual awareness

When we look at an ambiguous stimulus, such as a Necker cube or Rubin's famous face/vase our perceptual experience alternates between two different states. Yet the stimulus itself does not change. What is the difference in the neural response to the same stimulus when it is seen first as one object (e.g. a face) and then a moment later as a completely different object (e.g. a vase)?

2.1. Evidence for neural correlates of awareness

2.1.1. Binocular rivalry

A particularly striking example of perceptual bistability arises in the long-known phenomenon of binocular rivalry (DuTour, 1763; von Helmholtz, 1962), in which a different image is projected to each eye. When human observers view such displays, instead of seeing a blend of the two images, their perceptual experience seems to reflect a dynamic competition between the two inputs. If vertical stripes are presented to the left eye and horizontal stripes to the right eye, the viewer is likely to see not a superimposition of the two patterns (i.e. a crosshatching plaid pattern), but an alternating sequence in which only vertical stripes will be seen for one moment, and only horizontal stripes the next. Although the precise mechanisms underlying binocular rivalry are a matter of some debate (Blake, Yu, Lokey, & Norman, 1998; Leopold & Logothetis, 1999; Wolfe, 1986), it is clear that experience alternates in a bistable fashion between being dominated by the input to one eye and being dominated by the input to the other eye. Because the retinal input remains constant throughout, binocular rivalry provides an excellent domain in which to search for the neural correlates of perceptual awareness unconfounded by variations in the stimulus hitting the retina.

In a series of classic experiments, Logothetis and colleagues recorded from single neurons in visual areas of the monkey brain as the monkey viewed rivalrous displays (Logothetis, 1998). The monkeys were trained to report by pulling on a lever which of two stimuli they saw each moment. Logothetis and colleagues used a variety of stimuli (moving gratings, faces, etc.) that were selected because they either drove a particular neuron very strongly (a 'preferred' stimulus for that neuron), or because they drove that neuron only very weakly (a 'non-preferred' stimulus). Logothetis and colleagues then asked how the neural response to each stimulus varied as a function of the monkey's reported awareness of the stimulus when it was presented in a rivalrous display. They found that while some cells in the visual pathway responded to stimuli in a fashion independent of the monkey's state of awareness, other neurons showed activity correlated with the monkey's reported percept. For example, if a moving stimulus was delivered to one eye and a stationary stimulus to the other, a motion-sensitive neuron might respond more strongly when the monkey

reported seeing motion than when he did not. Further, the percentage of neurons showing correlations with awareness varied across different stages in the visual pathway, from about 20% in V1 and V2 to about 90% in inferotemporal cortex. These results suggest that neurons in later stages of the visual pathway are more closely correlated with the monkey's state of awareness than are neurons earlier in the visual pathway.

It seems reasonable to assume that when a monkey reports the presence of a particular stimulus, he is aware of the stimulus in something like the way that a human would be. Nonetheless, it would be reassuring to find similar results in the human brain. Opportunities for direct electrical recording from human brains are very limited (Allison, Puce, Spencer, & McCarthy, 1999; Fried, MacDonald, & Wilson, 1997). However, Tong, Nakayama, Vaughn, and Kanwisher (1998) used fMRI to run an experiment on humans that was modeled after the monkey experiments just described. Instead of recording the response of single neurons to preferred and non-preferred stimuli, we measured the responses from two regions of human visual cortex that have highly selective responses to specific stimulus classes. One region of extrastriate cortex called the fusiform face area (FFA) responds at least twice as strongly to faces as to other classes of non-face stimuli such as hands, objects, and houses (Allison et al., 1999; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). Another region on the ventral surface of the brain, the parahippocampal place area (PPA), responds strongly to images of places including houses, but only weakly to non-place stimuli, and not at all to faces (Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998). Thus, these two cortical regions, which can be found in almost all subjects, have opposite stimulus preferences: faces are preferred and houses are non-preferred for the FFA, and the opposite pattern holds for the PPA. By displaying a face stimulus to one eye and a house stimulus to the other eye, we could therefore simultaneously monitor with fMRI the neural response to each stimulus during binocular rivalry.

In our experiment subjects viewed a single rivalrous face–house stimulus for an entire scan, while reporting with a button press each switch in the content of their awareness. As in numerous previous studies of binocular rivalry, subjects reported that every few seconds their percept flipped, in this case from the face to the house, then back to the face. We then averaged the MR signal from each subject's FFA and PPA across all the face-to-house flips, and (separately) all the house-to-face flips, time-locked to the button press. For each subject we saw a clear rise in neural activity in each of the two cortical regions when the preferred stimulus for that region (i.e. the face for the FFA, and the house for the PPA) popped into awareness. A fall in the activity in each area was found when the preferred stimulus for that area dropped out of awareness. Thus, the activity in these two cortical areas was clearly correlated with the content of the subject's awareness, even though the retinal stimulus remained unchanged throughout the experiment.

We then asked how these neural correlates of awareness in binocular rivalry compared to the neural correlates of a change in the stimulus itself. In scans carried out on the same subjects in the same session, we recreated the same sequence of

perceptual states the subject had reported via button presses in a previous rivalry scan, but in this case we changed the stimulus itself (from just a face to just a house and so on). To our surprise, the data obtained from these stimulus alternation scans not only qualitatively resembled the data from the rivalry scans, but were also quantitatively indistinguishable. That is, the magnitude of the neural responses in the FFA and PPA to a rivalrous change in awareness with the stimulus held constant was as great as the corresponding non-rivalrous change when the stimulus itself changed from face to house or vice versa. Our data thus demonstrated not only a neural correlate of awareness, but a neural response that was just as strongly correlated with the subjects' state of awareness as it was with the stimulus. These results parallel the earlier work by Logothetis and colleagues (Logothetis, 1998), extending them to humans and further demonstrating even stronger correlations between neural activity and awareness.

But what exactly do these data tell us about the neural basis of perceptual awareness? The FFA and PPA were selected for this study not because of any presumed link to awareness, but instead because the strong stimulus selectivity of these regions provided the markers we needed to do the experiment at all. It would therefore be a monumental coincidence if these two areas just happened to play a special role in awareness. Further, it is unlikely that the FFA and PPA play a major role in awareness of stimuli that are neither faces nor places because most other stimuli that have been tested produce a similar and relatively low response in these areas (Kanwisher, Downing, Epstein, & Kourtzi, in press). Thus, the more reasonable conjecture would be that if these two areas play any particular role in perceptual awareness, that role is likely to be largely restricted to awareness of faces (for the FFA) and of places (for the PPA).

Is neural activity in other extrastriate areas also correlated with perceptual awareness of the stimulus attributes that are processed in that area? Indeed, evidence already exists for correlations between awareness and neural activity in at least two other extrastriate regions, which I discuss next.

2.1.2. Neural correlates of awareness of motion in MT/MST

Area MT/MST is a cortical region known to be involved in the processing of visual motion information in both monkeys and humans (Tootell, Reppas, Kwong et al., 1995). Several fMRI studies have shown strong correlations between neural activity in MT/MST and the perceptual experience of visual motion, unconfounded from stimulus motion. These studies make use of the motion aftereffect, in which adaptation to a stimulus with a constant direction of motion leads to a subsequent illusory percept of motion in the opposite direction. Tootell, Reppas, Dale et al. (1995) found that activity in MT/MST persisted for a longer period following adaptation to a motion stimulus with constant direction than following adaptation to a stimulus that changed direction frequently, consistent with perceptual reports of the subjects that a motion aftereffect was seen in the former but not the latter case.

Two subsequent studies made use of the fact that no motion percept occurs if the motion adaptation period is followed by a period in complete darkness. Instead, the aftereffect can be 'stored' for some period of time, producing a percept of motion

only later when a (stationary) stimulus is presented. Culham et al. (1999) demonstrated that activity in MT/MST was low during the storage period, but increased when a stationary stimulus subsequently appeared, exactly tracking the subjects' report of their experience of visual motion. He, Cohen, and Hu (1998) used a different design that exploited the spatial specificity of the motion aftereffect. After a long adaptation period, the investigators caused the aftereffect to alternately appear and disappear by having the subjects move their eyes so as to place a stationary stimulus either inside or outside the adapted region. The signal in MT/MST closely tracked the percept of motion. By unconfounding motion aftereffect storage from the experience of the motion aftereffect, these two studies strengthen the evidence that the neural signal in MT/MST is correlated with the percept of motion.

In a single-unit study of MT in awake behaving monkeys, Bradley, Chang, and Andersen (1998) showed another situation in which the activity of neurons in MT is correlated with changes in awareness that occur in the absence of changes in the stimulus. They used displays in which two sets of interleaved dots (each in a different stereo depth plane) move in opposite directions, producing a percept of a rotating cylinder. When the same display is viewed without stereo information a rotating cylinder is still perceived, but the percept is bistable, oscillating from one state in which one direction of motion is perceived in front and the opposite direction in back, to the other state in which the assignment of motion directions to depth planes is reversed. Some cells in MT preferred motion in one direction in the front plane and the opposite direction of motion in the back plane in unambiguous stereo displays. Of these, half (34/68) responded differently when the monkey reported different percepts in the ambiguous 2D versions of the same displays. Most of these cells (27/34) showed a higher response when the neuron's preferred pattern was perceived. This finding shows that activity in some cells within area MT in the macaque is correlated with the content of awareness.

2.1.3. Perceiving masked objects and letter stimuli

Another cortical area where correlates of awareness have been demonstrated very recently is the 'lateral occipital complex' (LOC), a large region in the ventral visual pathway that responds more strongly to images of objects, whether familiar or novel, than to scrambled images in which the structure of those objects is not discernable (Kanwisher, Woods, Iacoboni, & Mazziotta, 1996; Malach et al., 1995). Does this region play a role in awareness of object identity? Grill-Spector, Kushnir, Itzhak, and Malach (2000) presented photographs of familiar objects for 40 ms (followed by a 460 ms mask), 120 ms (followed by a 380 ms mask), or 500 ms unmasked. The subjects' accuracy in identifying the objects was measured separately for each presentation duration. Grill-Spector et al. also measured the corresponding response in the LOC using fMRI. For each stimulus presentation duration the investigators compared the response to objects followed by masks with the response to control stimuli with the same timing parameters but in which a different mask was presented in the place of the object (i.e. a mask followed by a different mask). The unmasked 500 ms object exposure was used to derive the maximal fMRI response and maximal

behavioral performance. Thus, both accuracy and fMRI response in the two shorter duration conditions could be plotted as a percentage of this maximal response, a clever technique enabling fMRI and behavioral functions to be directly compared. Grill-Spector et al. found strikingly similar functions relating object recognition performance to stimulus duration and relating the MR response in the LOC to stimulus duration. On the other hand, because this correlation was derived from comparisons across different stimulus durations, Grill-Spector et al. carried out a further test for a correlation between behavioral and MR response when the stimulus conditions were identical. They trained subjects to recognize briefly-presented objects, and demonstrated that the improvement in behavioral performance after training was paralleled by an increase in the MR signal in the LOC to these images after training (compared to before). Overall, across trained and untrained conditions and across exposure durations, the correlation between object recognition performance and MR signal in the LOC was very high, and indeed higher than in other regions of cortex that were sensitive to object structure.

Several other related results have also been reported recently. Bar et al. (in press) found a strong correlation between degree of success in object recognition and MR signal intensity in a region of the fusiform gyrus about 1 cm anterior to the FFA. In a similar vein, Kleinschmidt, Buchel, Huton, and Frackowiak (1998) presented a letter in a random dot pattern background and gradually ramped the clarity of the letter up and then down by varying the density of dots making up the letter. A hysteresis effect was found for both perception and MR signal intensity in the region of the LOC in which both the subject's performance and the neural responses were higher for a given intermediate level of stimulus information when the letter clarity was being ramped down compared to when it was being ramped up. That is, for these intermediate levels of stimulus clarity the probability of letter recognition and the neural activity in object-related areas were both higher if the subject had already seen the letter clearly than if they had not. Finally, Rees, Russell, Frith, and Driver (1999) displayed stimuli in which line drawing pictures of familiar objects overlapped spatially with letter strings that were either real words or non-word consonant strings. When subjects directed their attention to the letter stimuli, Rees et al. found a stronger MR response in several cortical areas to real words compared to non-words. More importantly, this differential response to words versus non-words was abolished when subjects directed their attention to the pictures, consistent with subjects' inability to report the identity of words presented in such displays. Thus, the subjective impression that words are not recognized when unattended was mirrored by the loss of the neural signature of word recognition in this condition.

All of these findings show impressive correlations between the ability to identify an object, letter, or word, and the strength of the neural signal in the relevant cortical area. However, one thing these studies do not yet clearly address is the precise aspect of the stimulus information that is correlated with awareness, which could range from detection of something (rather than nothing), to a mid-level analysis of the shape (or orthography, for the Rees et al., 1999 study) of the item, to an appreciation of the high-level meaning of the stimulus in question. Because awareness of each of these kinds of information is likely to be highly correlated in the studies described

above, the observed neural correlations could reflect awareness of information at any (or all) of these levels.

2.1.4. *Attention, imagery, etc.*

Other phenomena that affect the contents of our perceptual awareness include attention, mental imagery, and changing states of consciousness. For each of these phenomena, neural signals have been shown to covary with perceptual awareness. As described above for the Rees et al. (1999) study, simply focusing visual attention on different aspects of an unchanging stimulus has a strong effect on the content and intensity of perceptual awareness. Closely following the effect of attention on subjective experience, numerous studies using single-unit recordings (Desimone & Duncan, 1995), ERPs (Luck & Girelli, 1998), and brain imaging (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990; O'Craven, Rosen, Kwong, Treisman, & Savoy, 1997) have shown clear modulations of sensory responses by attention, even for a constant stimulus, and even in primary visual cortex (see Kanwisher & Wojciulik, 2000 for a review). A rather different manipulation of perceptual awareness occurs during mental imagery, in which no stimulus is present at all. Selective activation of MT/MST has been reported during mental imagery of motion (Goebel, Khorram-Sefat, Muckli, Hacker, & Singer, 1998), and selective activation of the FFA and PPA has been reported (O'Craven & Kanwisher, in press) for face and place imagery, respectively. In each of these cases, the activations during mental imagery are weaker than the corresponding stimulus activations.¹ Finally, a recent fMRI study has shown that the response of auditory and language cortex to speech stimuli disappears soon after sleep onset (McDermott, 1996), consistent with the subjective experience that auditory awareness largely ceases at sleep onset.

2.1.5. *Microstimulation*

The studies described above show that across a wide range of manipulations in which the contents of perceptual awareness vary but the stimulus does not, neural signals exist that follow closely in step with subjective experience. But are these patterns of neural activity sufficient to cause the corresponding percept? Evidence bearing on these questions is scarce, but one technique is particularly informative here. Salzman, Britten, and Newsome (1990) showed that when a monkey performs a motion direction discrimination task, its response can be biased by microstimulation of a small region within cortical area MT where cells respond preferentially to a given direction of motion. Such findings provide unusually strong evidence for the causal connection between neural activity in a given

¹ This result was anticipated by Hume, who commented on the relationship between percepts and ideas/images as follows: "The difference betwixt these consists in the degrees of force and liveliness, with which they strike upon the mind... [Perceptions] enter with most force and violence... By ideas I mean the faint images of these in thinking and reasoning."

extrastriate area and the resulting perceptual experience. Although we cannot exactly ask the monkey what it experiences when electrically stimulated, its performance in a perceptual discrimination task seems a reasonable proxy for such a report. Further, a consistent picture is provided by the few studies of cortical microstimulation in humans where we can ask the subject what they experience. Puce, Allison, and McCarthy (1999) measured responses from electrodes implanted subdurally (for the purposes of presurgical mapping) in ventral extrastriate areas in epileptic patients. Face-selective responses were sometimes found in fusiform electrode sites, and in several cases subsequent stimulation through the same site produced a percept of a face or a face part (see also Penfield & Perot, 1963; Vignal, Chauvel, & Halgren, 2000). These results suggest that neural activity in particular locations within extrastriate cortex can cause specific subjective perceptual experiences, strengthening the evidence for a causal connection between neural activity and awareness (but see Section 3 below).

2.2. *Brain loci of the neural correlates of perceptual awareness*

The multiplicity of cortical loci where correlations with awareness have been found provides some evidence against one of the oldest ideas about consciousness, that the contents of awareness are represented in a single unitary system (Schacter, McAndrews, & Moscovitch, 1988), variously described as a stage (Taine, quoted in Ellenberger, 1970), workspace (Baars, 1988), ‘Cartesian theater’ (criticized by Dennett, 1991), or cave wall (Plato). Instead, the data described above seem more consistent with a view in which the contents of current awareness can be represented in many different neural structures. However, one could still argue that the neural correlates described above are not in fact the actual representations that constitute the conscious percept, but merely information that is likely to make it onto the (as-yet-undiscovered) screen of awareness, so the possibility of such a unitary awareness system is not definitively ruled out by these data.

In contrast to the idea of a unitary and content-general Cartesian theater of awareness, the data summarized above fit more naturally with the following simple hypothesis: *the neural correlates of awareness of a given perceptual attribute are found in the very neural structure that perceptually analyzes that attribute*. This hypothesis accommodates the fact that perceptual awareness is not simply a matter of knowing whether a stimulus was or was not presented, but is a much more multifaceted phenomenon. There are as many ways to be aware of a stimulus as there are kinds of information to register about that stimulus. Thus, perceptual awareness might involve any aspect of the stimulus, from its simple presence (as opposed to absence), to the presence or nature of one or more of its perceptual attributes, to the category of object present in the image, to a fine-grained recognition of the particular exemplar of that category, to the ‘gist’ of a complex scene. Decentralizing the neural correlates of awareness to the processors where this information is extracted provides a straightforward account of why some aspects of a stimulus can be consciously perceived while other attributes are not.

Are there any constraints at all on the neuroanatomical loci that can participate

in awareness? Surely events can occur on the retina, for example, without our becoming aware of them. Which neural systems are likely to hold the contents of awareness and which are not? One possibility is that neural representations become more correlated with awareness at later stages of perceptual processing, as Logothetis (1998) found for the macaque in the case of binocular rivalry. However, a recent fMRI study of binocular rivalry in humans found substantial correlations between visual awareness and neural representations in human V1 (Polonsky, Blank, Braun, & Heeger, 2000), so it is not clear that the neural correlates of human visual awareness will behave in the same way, increasing as one ascends the visual system. Another common speculation is that the contents of awareness are represented only in the cortex, not in subcortical structures. A third possibility is that the ventral (occipitotemporal) visual pathway holds the contents of awareness whereas the dorsal (occipitoparietal) pathway is more involved in a variety of unconscious computations underlying visuomotor coordination (Milner & Goodale, 1995). Below I propose the related but somewhat different hypothesis that the neural correlates of the *contents* of visual awareness are represented in the ventral pathway, whereas the neural correlates of more general-purpose *content-independent* processes associated with awareness (attention, binding, etc.) are found primarily in the dorsal pathway. However, this hypothesis is highly speculative and indeed is already known to have at least one exception: the correlations between awareness of visual motion and activity in area MT (a dorsal pathway area) already described in Section 2.1.2.

3. Mere correlation, or causal connection?

As Section 2 of this article makes clear, neural correlates of perceptual experience, an exotic and elusive quarry just a few years ago, have suddenly become almost commonplace findings. Specific neural populations have been found in which neural activity is strongly correlated with subjective experiences of faces, places, objects, and motion. But what exactly do these findings tell us about perceptual awareness? Any deep scientific understanding requires getting beyond mere correlations, to a deeper understanding of the causal structure of the underlying phenomena. In the case of the relationship between neural activity and perceptual awareness, what we really want to know is not what patterns of neural activity are correlated with perceptual awareness, but rather what patterns of neural activity are *necessary* and/or *sufficient* for perceptual awareness.

The causal relationship between a particular pattern of neural activity (e.g. in the FFA) and the corresponding state of perceptual awareness (e.g. of a face) can be evaluated by considering the situations represented by each of the four cells in Fig. 1. The findings described in Section 2 of this paper include many cases in which the relevant pattern of neural activity and corresponding state of awareness are either both present (the lower right cell) or both absent (the upper left cell). These examples are consistent with a strong causal connection between the relevant neural activity and the relevant state of awareness. The findings from microstimu-

	Perceptual Awareness -	Perceptual Awareness +
Neural Activity -	√	Potential Evidence Against Necessity
Neural Activity +	Potential Evidence Against Sufficiency	√

Fig. 1. The possible combinations of a particular pattern of neural activity or its absence, and a corresponding state of perceptual awareness or its absence, and the evidence each case can provide about the causal relationship of the pattern of neural activity to the perceptual state.

lation described in Section 2.1.5 are particularly strong evidence for a causal connection. However, it is the other two cells in this figure that are potentially more informative, as it is only these cases that can in principle provide evidence against a strong causal connection (or identity) between a particular pattern of neural activity and a particular state of perceptual awareness. Specifically, a situation in which a given pattern of neural activity is absent but the relevant state of perceptual awareness is present (i.e. the upper right cell in Fig. 1) would imply that the pattern of neural activity in question is not necessary for that state of awareness. And conversely a situation in which a given pattern of neural activity is present but the relevant state of perceptual awareness is absent (i.e. the lower left cell in Fig. 1) would imply that the pattern of neural activity in question is not sufficient for that state of awareness. Evidence for either of these two situations would therefore refute a strong claim that the neural activity in question is causally related to or identical to the perceptual state in question.

Consider first the question of necessity. If a condition were ever found in which a subject is aware of a face yet a strong response were not found in their FFA, that would show that activity in the FFA is not necessary for awareness of faces (modulo the sensitivity of the measurement technique). I know of little convincing evidence of this kind. However, proving the null hypothesis is notoriously problematic, all the more so when the physiological signal being monitored is very noisy (as in the case of fMRI). This is therefore a particularly difficult condition to test. One possible approach is to turn to neuropsychology, to ask whether awareness of

a given perceptual attribute is ever found in the complete absence of the relevant cortical structure. For example, if one lacked an MT, would awareness of visual motion be obliterated? Evidence from one patient suggests that it would be (Zihl, von Cramon, & Mai, 1983). Are the FFA and PPA necessary for awareness of faces and places? Some evidence suggests that patients who lack an FFA can perceive faces as faces, but are very impaired at identifying the individual whose face they are looking at (de Gelder & Kanwisher, 1999). One might therefore argue that the FFA is necessary for awareness of facial identity, though perhaps not for the awareness of faces at all. This kind of investigation has the potential to be very useful in determining the particular cortical regions that are necessary for a subject to experience a particular state of perceptual awareness.

Evidence against the sufficiency of a particular pattern of neural activity for a particular perceptual state would come from a situation in which that neural activity occurs (e.g. activation of the FFA for faces) yet the expected perceptual state (e.g. awareness of faces) does not. Insofar as the relevant neural signal was sufficiently selective, such a case would also provide a demonstration of perception without awareness, a question of interest in its own right. More importantly, any such demonstration that perceptual representations can be decoupled from awareness would set the stage for a research program directed toward determining what else is necessary for perceptual awareness beyond the mere existence of a perceptual representation. We therefore consider the evidence for perceptual representations without awareness in some detail in the next section.

3.1. Evidence for activation of perception representations in the absence of awareness

A long tradition of research in experimental psychology has provided considerable evidence that stimuli can affect behavioral responses even when they are not consciously perceived (Sidis, 1898, reviewed in Merikle, Smilek, & Eastwood in this volume). Another fascinating line of work has demonstrated many cases in which perceptual awareness can be decoupled from perceptual processing in neuropsychological patients (Driver & Vuilleumier, this volume; Milner & Rugg, 1992). Here we will focus on the evidence from on-line measures of neural activity.

If a stimulus is so faint as to be completely invisible, can it nonetheless lead to activation of visual cortex? In a recent study by Tootell, Hadjikhani, and Somers (1999), subjects were scanned with fMRI while they viewed stimuli in which periods of dynamic visual gratings alternated with periods in which a uniform gray field of equal mean luminance was displayed. The grating stimuli were displayed with several different levels of contrast in different scans. Two important results were obtained from this study. First, for all visual areas scanned (including V1, V2, V3, VP, V3A, V4v, and MT/MST), activity increased monotonically with stimulus contrast. Second, for the lowest contrast tested, although at the end of the scan the subjects reported having seen nothing but a uniform field for the entire

scan,² all retinotopic visual areas tested showed significantly stronger activation to the invisible gratings than to the uniform gray field. These results demonstrate a clear neural response to a stimulus that apparently did not enter awareness. Thus, several different stages of the visual hierarchy, from V1 to V4v, can be activated by stimuli that the subject is not aware of.

Can such visual activations outside of awareness be found for even higher levels of processing? Whalen et al. (1998) asked whether the response of the amygdala to angry compared to happy faces would be found even when subjects were unaware of any emotional expression in the faces at all. They scanned subjects who viewed a series of brief (33 ms) presentations of emotionally expressive faces, each of which was immediately followed by a 167 ms presentation of a neutral face. The neutral faces masked the preceding emotionally expressive faces such that emotional expressions were rarely perceived and at the end of the experiment eight of the ten subjects reported never having seen an emotionally expressive face at all in the entire experiment. Nonetheless, a significant activation of the amygdala was found for the epochs in which masked angry faces were presented, compared to masked happy faces. Thus, even such subtle and high-level visual information as the emotional expression of a face can be represented neurally without the subject reporting any awareness of that information.

Are visual responses to emotional stimuli ‘special’, or can neural representations of other kinds of high-level information be found for stimuli that are not consciously perceived? In a recent study, Rees et al. (2000) (see also Driver & Vuilleumier, this volume) scanned a patient with right parietal damage and extinction, which is the failure to perceive stimuli presented in the contralesional or ‘bad’ field when a competing stimulus is presented simultaneously in the ipsilesional or ‘good’ field. Of interest was the finding that an independently-defined face-selective region in the fusiform gyrus of this patient showed activations for faces that were at least as strong when the faces were not consciously perceived (i.e. in the bilateral presentation extinction condition) as when they were (in the unilateral presentation condition). These activations, though statistically weak, appear to be stimulus-selective as they were not found for house stimuli in the same region.

Is there any evidence that even semantic information can be neurally represented without awareness? Luck, Vogel, and Shapiro (1996) also measured the neural response to an unseen stimulus, but they used a perceptual phenomenon called the ‘attentional blink’ (Raymond, Shapiro, & Arnell, 1992). In the attentional blink,

² Note that this study (as well as the study by Whalen et al. (1998) described next) used a ‘subjective’ measure of lack of awareness, rather than an ‘objective’ measure. That is, the subjects simply said they did not see anything, but were not required to do a forced-choice discrimination task. One might argue that the finding would be stronger if an objective (forced-choice) measure were used, because we don’t know what criterion the subject used to decide they ‘did not see’ something. However, the subjective measure is closer to the intuitive notion of lack of awareness. The choice of definitions could lead to different results if subjects show above-chance performance on a forced-choice task while reporting zero awareness of the stimulus. To insist that we take their performance rather than their subjective report as the index of awareness assumes that any correct performance is consciously mediated, an assumption that is unlikely to be valid. Given this problem, it is most useful to have both measures of awareness.

subjects view two successive masked target stimuli, separated by a temporal interval of variable duration. If subjects must carry out a task on the first target, then their ability to detect the second target falls dramatically for inter-stimulus intervals of 100–400 ms. However, the second target is accurately detected at shorter or longer intervals, or if subjects need not carry out any task on the first target. Thus, the requirement to analyze the first target leads to a drop in awareness of the second target. Luck et al. presented a rapid sequence of symbol strings to subjects, and asked them to report two targets from each sequence. One string in each sequence was a row of identical digits, and subjects had to report whether the digits were odd or even. The second target was a word, and subjects had to report whether the word was related or unrelated to a context word presented just before the sequence began. In different conditions, zero, two, or six items appeared between the digit string and the word. Consistent with prior findings on the attentional blink, performance identifying the digit string was high for all conditions, but accuracy on the word task was much lower for the intermediate lag than for the zero-lag or six-lag conditions. While subjects performed this task their scalp ERPs were measured. The amplitude of the N400, which is found for words that do not fit semantically in the context compared to words that do, was just as great for unrelated word targets for the intermediate lag (when conscious report of those words was very low) as for the other two lags (when overt report of the words was high). Thus, even though subjects failed to recognize the word targets on most of the intermediate-lag trials, their N400 response to the meaning of the word was undiminished compared to the other lags. This study therefore demonstrates that a neural correlate of accessing word meaning is unaffected by whether the word reaches awareness or does not.

One common intuition is that we can only respond overtly to a stimulus if that stimulus has been consciously perceived. But does the preparation of a motor response to a stimulus in fact require awareness of the stimulus responded to? A study by Dehaene et al. (1998) suggests that it may not. These researchers presented number words to subjects very briefly, followed by a mask, under conditions in which subjects were at chance in discriminating their presence versus absence, and at discriminating the words from nonsense strings. Immediately after the masked number word prime, a suprathreshold target digit was displayed, and subjects had to report whether it was greater or less than five. Behavioral responses to the target digit were slower when the correct response to the suprathreshold target was inconsistent with the response that would have been required for the preceding unseen prime word, compared to when the prime was consistent with the target. This result demonstrates that even though subjects had no task to carry out on the prime word, and even though they were not aware of it, they nonetheless processed it to a high level. To obtain this effect the prime word must have been processed at least to the level of representing the meaning (i.e. the magnitude) of the named number. But was this information processed to an even higher level? To answer this question Dehaene et al. measured both scalp ERPs and fMRI responses from motor cortex from the subjects while they carried out the task. As expected, both measures demonstrated clear responses in motor

cortex in the hemisphere contralateral to the hand the subject used to respond on that trial. However, more important was the finding that motor cortex activation was also seen contralateral to the hand that would have produced the correct response to the unseen prime word. Of course, the motor responses to the unseen prime word were smaller in magnitude than those to the suprathreshold target digit. Nonetheless, the fact that a specific effect was found to the prime word in motor cortex demonstrates that processing of an unseen target can proceed all the way to the preparation of a motor response. Similar findings using ERPs were also reported by Eimer and Schlaghecken (1998).

In sum, specific neural responses to unseen stimuli have been observed at a variety of levels from early visual processing in retinotopic cortex to the extraction of structural or emotional information from faces, to accessing the meanings of words and even the preparation of a motor response.

4. What is the difference between a conscious perceptual representation and an unconscious one?

The data summarized in Section 3.1 show that perceptual representations can be activated in the absence of awareness of those representations. Evidently, activation of these representations is not sufficient for awareness. What else is needed? Put another way, what is the difference between a perceptual representation that is consciously experienced and one that is not?

4.1. The activation strength hypothesis

Probably the simplest hypothesis that has been offered in answer to this question, sometimes called the ‘quality of representation’ (Farah, 1994) or ‘activation’ (Baars, 1988; Palmer, 1999) hypothesis, is this: *the more active a given neural representation, the stronger its representation in awareness*. This hypothesis is congenial to the fact that perceptual awareness is not generally an all-or-none affair, but a graded phenomenon which admits many shades of gray. This insight forms the basis of signal detection theory (Green & Swets, 1966), which posits a continuum in the possible amounts of perceptual information that may be extracted from a stimulus. This continuum is then divided into two response categories by a somewhat arbitrary threshold that the subject must impose when forced to make a binary decision about the stimulus. Where exactly the subject places the threshold on that continuum is determined by numerous factors such as the instruction and payoff matrix given to the subject by the experimenter. Thus, the fact that we can obligate subjects to produce a binary response should not fool us into thinking that their internal state itself is binary or that there is anything important or fixed about the particular threshold the subject uses. Indeed, anyone who has been a subject in a psychophysical experiment will be familiar with the uncomfortable feeling of having to force an unclear and inchoate perceptual experience into one of a small number of discrete response categories. The activation hypothesis holds that this continuum of degrees

of perceptual awareness is encoded neurally as the strength (or ‘quality’) of the underlying neural representation.

What do the data reviewed in Section 3.1 have to say concerning the activation strength hypothesis? It would be unsurprising if the function relating activation strength to awareness were not linear, but instead contained a threshold. At the lower end of the curve the strength of the neural representation might be greater than zero but the level of awareness might not. Thus, some cases of neural representations outside of awareness might be explained in terms of subthreshold activations that are strong enough to be detected by ERP or fMRI sensors, but not strong enough to result in awareness. However, this account does not work well for cases in which the strength of the neural signal is very similar when a given stimulus is consciously perceived and when it is not. Both the Luck et al. (1996) study and the Rees et al. (2000) study appear to be cases in which the neural signal is about as strong in the conscious as the non-conscious conditions; Driver and Vuilleumier (this volume) discuss parallel cases in which behavioral markers are just as strong for the conscious as the non-conscious cases. Thus, preliminary indications are that although the activation strength hypothesis may be partly true, it is incomplete. This in turn implies that awareness is dependent on something other than the strength of a given perceptual representation. What other factors might be important? Next I consider two more possibilities.

4.2. The informational access hypothesis

One line of thinking suggests that awareness of perceptual information requires not only a strong representation of the contents of awareness, but access to that information by other parts of the mind/brain (Baars, 1988). The idea that access to the relevant representations is a substantial constraint on perceptual awareness makes sense given the known functional architecture of the mind and brain. First, human neuroanatomy is characterized by wide variation in the degree of connectivity between different brain areas. While some neural path exists that connects any two parts of the brain, these paths will vary greatly in strength and directness. Second, at a functional/cognitive level, one of the key principles underlying the concept of the modularity of the human mind is ‘informational encapsulation’, the idea that there are substantial constraints on the access to intermediate representations computed within each functional module (Fodor, 1983). Thus, it would not be surprising if perceptual representations existed that failed to enter awareness, not because they were not ‘strong’ enough, but instead because other parts of the mind could not gain access to them.

Third, to appreciate the idea that the mere existence of a representation is not likely to be sufficient for awareness, consider the following thought experiment. Suppose cortical area MT was surgically removed from a human brain. Suppose further that its interconnections remained intact, and it was kept functional in a dish for some period of time despite the lack of input and output connections to the rest of the brain. Now suppose that a region within MT was microstimulated as described in Section 2.1.5, a manipulation that apparently produces a conscious percept when carried out in an intact animal or person. Surely awareness of motion would not

occur for an isolated MT in a dish. (Who would see the motion?!) Thus, common sense suggests that perceptual awareness probably requires not only a strong neural representation in a particular cortical area, but access to that representation by at least some other parts of the system.

But who or what must have access to a given representation for it to reach awareness? According to a common intuition about perceptual awareness (e.g. Baars, 1988), if you perceive something, then you can report on it through any output system (speech, button presses, drawing, American Sign Language, etc.). Perceptual information that could be reported through only one output system and not through another just would not fit with most people's concept of a true conscious percept.³ Thus, conscious access to perceptual information seems to imply access to most or all output systems. On the other hand, few would argue that perceptual awareness would be affected if temporary paralysis made overt report impossible, so access by output systems per se does not seem necessary for perceptual awareness. Instead, it seems that a core part of the idea of awareness is that not only effector systems, but indeed most parts of the mind have access to the information in question. Thus, in agreement with Baars (1988), it seems reasonable to hypothesize that *awareness of a particular element of perceptual information must entail not just a strong enough neural representation of that information, but also access to that information by most of the rest of the mind/brain.*

How might a given piece of perceptual information become accessible to most of the mind/brain? A unitary 'conscious awareness system' (Schacter et al., 1988) or 'global workspace' (Baars, 1988) that enabled information to be widely 'broadcast' could in principle accomplish this goal. The idea that the contents of awareness must be represented in a distinct neural locus has been criticized on the grounds that it implies a homunculus that must then look at the information so represented (Dennett, 1991). However, there is no need to posit such a mystical entity. The brain could in principle have a discrete locus where the contents of awareness are represented for the same reason that airlines have hub cities: to facilitate the most efficient transfer of information (or people) between any two points in a large space of possible destinations and points of departure. However, because the format of representations is very different in the different modules of the mind/brain, an important problem for any such unitary system would be how it could have the representational power to accommodate inputs from all of the different modules that would send information to it. In any event, I know of no evidence for a discrete neural structure that has the properties that would be required of a unitary system for awareness. Further, as summarized in Section 2, currently available data suggest that the contents of awareness are represented not in a single neural locus but in multiple different cortical areas.

One might think of the global workspace not as a neuroanatomically localized

³ Of course if one output system is damaged (e.g. in the case of aphasia) such that perceptual information could not be communicated through that output system, but could still be communicated through all other remaining intact output systems (drawing, button presses, etc.), this would be consistent with the intuition about awareness put forth here.

system, but instead as some kind of functional state of the brain. For example, on the Desimone and Duncan (1995) ‘interactive competition’ model, competitive interactions across cortical areas result in domination of perceptual representations by properties of a single object. This competition can be biased by either bottom-up factors (e.g. stimulus salience) or top-down factors (e.g. endogenous attention). In either case the net result is that the various properties of an object, represented in distinct cortical regions, enhance each other and suppress the representation of competing objects. On this view, attention and awareness are global properties of the entire perceptual system that span multiple cortical areas. Although Desimone and Duncan (1995) offer no mechanism to explain how different cortical areas come to represent attributes of the same object, there is some evidence that this in fact occurs (O’Craven, Downing, & Kanwisher, 1999). To the extent that mechanisms exist that can cause disparate cortical areas to represent perceptual information about the same object, one might expect that the same mechanisms could also cause that information to be widely available to much of the rest of the system. Synchronous firing of neurons across cortical areas could play some role in this process (Singer, 2000), though a full account would have to explain how the synchrony is established and how it is interpreted by subsequent stages of processing.

4.2.1. Changing access to perceptual information

Limits on conscious access to perceptual information may not be immutable. In the most extreme case, brain damage may disrupt neural pathways such that perceptual information represented in one neural structure no longer is accessed by other parts of the system. However, dissociations of perception and awareness are abundant in the neuropsychology literature (Farah, 1994; Milner & Rugg, 1992), and disconnections may not be sufficient to explain all of them. Another possibility is brain damage may disrupt a global state of integration of the entire brain, thereby affecting access even to information represented in sites remote from the damage.

Conscious access to perceptual information may also change over time even in undamaged brains. First, cognitive systems may become more integrated over the normal course of development in infancy and childhood, such that each modular component of the mind gains greater access to information represented in other modules. Indeed, Spelke, Vishton, and Von Hofsten (1995) have argued that “In adults, distinct systems of knowledge may work together, such that a wide range of distinct beliefs can jointly influence our thinking and deliberate action... In infancy, distinct knowledge systems may be less interconnected.”

A second situation in which information access and awareness may change in normal brains occurs in perceptual learning. It is a common experience of subjects in psychophysical tasks that as one improves at the task, one becomes aware of stimuli that one did not at first perceive. Perhaps what changes with practice is not simply the quality or strength of the underlying perceptual information, but the ability to ‘find’ or ‘read out’ that information by other parts of the system. Several studies have shown that the inclusion of a few suprathreshold trials in a perceptual learning procedure can lead to a sudden drop in the threshold for a perceptual task

(Rubin, Nakayama, & Shapley, 1997), as if the stronger signals available in these suprathreshold trials ‘show’ the subject where the relevant representations can be found in the nervous system. To the extent that this (highly speculative) ‘access’ interpretation of perceptual learning is true, then two strong predictions follow. First, in cases where perceptual learning occurs, it should be possible to demonstrate that the relevant perceptual information was actually present (though not consciously perceived) before the learning occurred. Second, in cases where perception without awareness has been demonstrated, it should be possible with sufficient training to become aware of the originally unconscious information. While these strong predictions may ultimately be shown to be wrong, the point being raised here is that perceptual learning may be mediated in part by changes in access to the relevant information, and not only by changes in the quality of the information accessed.

4.3. *The type-token hypothesis*

The many striking recent findings that relate neural activity to awareness are certainly thought provoking. However, there is no a priori reason to suppose that the *neural* correlates of awareness are any more likely to result in a deep understanding of perceptual awareness than are the *cognitive* correlates of awareness. Indeed, the behavioral literature has already independently led to the idea discussed above that a strong representation of a given perceptual attribute is not sufficient for awareness of that attribute, but that other processes must be involved.

In earlier papers (Kanwisher, 1987, 1991) I suggested that awareness of a particular perceptual attribute requires not only activation of a representation of that attribute, but also *individuation of that perceptual information as a distinct event*. Perceptual experience is made up not of free-floating perceptual features (e.g. redness, motion to the left), but instead of discrete objects that appear in particular spatial locations and at specific times (Kahneman & Treisman, 1984; Kanwisher, 1991; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). Thus, activated perceptual attributes must become associated with representations of specific objects and/or events in order to be experienced as fully fledged conscious percepts. In the terminology of Marcel (1983), conscious perception requires the attribution of perceptual information to a spatiotemporal ‘source’.

The gist of this idea is best explained by describing a typical subjective experience that occurs in experiments from this research tradition. You are seated in front of a computer monitor, and asked to view a very rapidly-presented sequence of words flashing on the screen. You are then asked to report the identities of the words just presented. But all you saw was a bunch of letters and patterns flash by so quickly that you have no idea what words were presented. If pressed to guess, you are left in an uncomfortable situation. Of course you can think up words to guess at random (and come to think of it the word ‘tiger’ would be as good a guess as any). But the exercise seems absurd and indeed intrusive. Given that you did not see any words, why should you tell the experimenter that the word ‘tiger’ just

popped into your mind? ‘Tiger’ is simply a random thought, not a percept. And what right does this experimenter have to the contents of your thoughts? But then, obligated to guess, you just say ‘tiger’ rather than bothering to make up anything else. Then to your amazement the experimenter tells you that that’s right, and ‘tiger’ was indeed one of the words in the sequence you just viewed.

What’s going on here? According to the token individuation hypothesis (Kanwisher, 1987), when perception is pushed beyond its processing capacity by very rapid presentation of stimuli, perceptual attributes (‘types’) can be activated without necessarily becoming linked to an episodic representation of a distinct perceptual object or event (a ‘token’). Because the activated type (e.g. the word ‘tiger’) does not get attributed to a specific external source (e.g. the flash of light at position x,y on the screen at time t), it feels subjectively more like a thought than a percept. This decoupling of type activation from token individuation occurs in numerous demonstrations of masked priming (Marcel, 1983), and is particularly strong in perceptual phenomena such as repetition blindness (Chun & Cavanagh, 1997; Kanwisher, 1987) and the attentional blink (Raymond et al., 1992). The experiment by Luck et al. (1996) described in Section 3.1 above provides evidence that the meaning of a ‘blinked’ word is activated even when the subject is unaware of the word. Indeed, Luck et al.’s evidence suggests that activation of the meaning of the word is no weaker when it is blinked than when it is not, consistent with our hypothesis that awareness is not merely a function of the strength of activation of the relevant information. The account proposed here is that the further necessary prerequisite for awareness that fails to occur in the attentional blink (and repetition blindness, masked priming, and presumably other cases of perception without awareness) is the binding of activated perceptual attributes with a representation that specifies the time and place that the word appeared (i.e. a ‘token’).⁴

What exactly is this process of binding activated types to individuated token? Some evidence (Kanwisher, 1991) suggests that it is the same process that is necessary for conjoining visual features (Treisman & Gelade, 1980). Visual attention is necessary for this binding to occur (Treisman & Gelade, 1980), and hence also for visual awareness. Thus, token individuation and visual attention are likely to be closely linked (if not identical) concepts, and they are likely to involve similar or identical neural substrates. Indeed, extensive evidence suggests that damage to similar structures in the parietal lobe leads to disorders of attention and awareness, explicit feature binding (Friedman-Hill, Robertson, & Treisman, 1995; Wojciulik & Kanwisher, 1998, 1999), and the linking of activated types to individuated perceptual tokens (Baylis, Driver, & Rafal, 1993).

Thus, neural activity in specific regions within the ventral pathway is apparently correlated with the *content* of perceptual awareness, whereas neural activity in the dorsal pathway may be correlated instead with the occurrence of perceptual awareness in a completely content-independent fashion. Interestingly, Driver and Vuilleumier (this volume) arrive at a very similar conclusion based on largely

⁴ See Mel and Fiser (2000) for suggestions on how object recognition may be possible without bottom-up feature binding, as implied by the type-token hypothesis.

independent evidence from that considered in this article. Further consistent with this suggestion, recent studies have provided evidence for content-independent activations of parietal structures during both the engagement of visual attention (Wojciulik & Kanwisher, 1999) and during changes in perceptual awareness (Lumer, Friston, & Rees, 1998). Although extensive evidence is not yet available, I will hazard a conjecture that (i) the same cognitive and neural mechanisms are involved in explicit feature binding, perceptual awareness, visual attention, and token individuation, and (ii) each of these processes will require interactions with the ventral pathway, where the relevant perceptual contents are represented. It may take a relatively long time in perceptual terms (between 100 and 200 ms) for these interactions to get established in a stable fashion for each percept. When this process is prevented or incomplete the subject may experience either a complete lack of awareness of the stimulus, or fleeting awareness followed by rapid forgetting (Potter, 1993).

5. Conclusions.

FMRI and ERPs have enabled us to peer into the human brain and observe the neural signatures of the contents of awareness, the shadows on the cave wall of the mind. Although the evidence described above sheds little light on the really difficult question of why awareness feels like anything, it does provide preliminary answers to a number of more scientifically tractable questions. Neural correlates of the contents of perceptual awareness can be found in many different cortical areas, from V1 to MT and the face area. I hypothesize that the contents of awareness are not represented in a single unitary consciousness system, but rather that each conscious perceptual content is represented in the same set of neurons that analyze that perceptual information in the first place. Further, there is now fairly compelling evidence from several different techniques showing that perception without awareness is possible. Thus, a strong neural representation in a given cortical area is not sufficient for awareness of the information so represented, raising the question of which perceptual information will reach awareness. I speculate that in order for a focal neural representation to reach awareness it may have to be accessible to other parts of the brain. Finally, I suggest that a conscious percept is not simply a disorganized soup of activated visual attributes, but rather a spatiotemporally structured representation in which visual attributes are associated with particular objects and events. The construction of a fully conscious percept may involve interactions between domain-specific systems for representing the contents of awareness (primarily in the ventral visual pathway) and domain-general systems (primarily in the dorsal pathway) for organizing those contents into structured percepts.

Acknowledgements

I thank the following people for very useful discussions and comments on the manuscript: Moshe Bar, Ned Block, Francis Crick, Dan Dennett, Russell Epstein,

Kalanit Grill-Spector, Christof Koch, Ken Nakayama, Molly Potter, John Rubin, Miles Shuman, and Frank Tong. This work was supported by a Human Frontiers grant and NIH grant 59150 to N.K.

References

- Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex*, *5*, 415–430.
- Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.
- Bar, M., Tootell, R. B. H., Schacter, D.L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B.R., & Dale, A. M. (in press). Cortical mechanisms specific to explicit visual object recognition. *Neuron*.
- Baylis, G., Driver, J., & Rafal, R. D. (1993). Visual extinction and stimulus repetition. *Journal of Cognitive Neuroscience*, *5*, 453–466.
- Blake, R., Yu, K., Lokey, M., & Norman, H. (1998). Binocular rivalry and motion perception. *Journal of Cognitive Neuroscience*, *10*, 46–60.
- Bradley, D. C., Chang, G. C., & Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, *392*, 714–717.
- Chalmers, D. (1995). *The conscious mind: in search of a fundamental theory*. Oxford: Oxford University Press.
- Chun, M. M., & Cavanagh, P. (1997). Seeing two as one: linking apparent motion and repetition blindness. *Psychological Science*, *8*, 74–79.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, *248* (4962), 1556–1559.
- Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375*, 121–123.
- Culham, J. C., Dukelow, S. P., Vilis, T., Hassard, F. A., Gati, J. S., Menon, R. S., & Goodale, M. A. (1999). Recovery of fMRI activation in motion area MT following storage of the motion aftereffect. *Journal of Neurophysiology*, *81*, 388–393.
- de Gelder, B., & Kanwisher, N. (1999). Absence of a fusiform face area in a prosopagnosic patient. *NeuroImage*, *9*, S604.
- Dehaene, S., Naccache, L., Le Clec, H. G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P. F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*, 597–600.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Dennett, D.C. (1991). *Consciousness explained*. Boston: Little, Brown and Company.
- DuTour, E. -F. (1763). Discussion d'une question d'optique. *Memoire de mathematique et de physique presentes par divers savants* (Vol. 4, pp. 499–511). Paris: Academie des Sciences.
- Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: behavioral and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1737–1747.
- Ellenberger, H.F. (1970). *The discovery of the unconscious: the history and evolution of dynamic psychiatry*. New York: Basic Books.
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*, 115–125.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.
- Farah, M. J. (1994). Visual perception and visual awareness after brain damage: a tutorial overview. In C. Umiltà & M. Moscovitch, (Eds.) *Attention and performance*, XV. Cambridge, MA: MIT Press.
- Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

- Fried, I., MacDonald, K., & Wilson, C. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron*, *18*, 753–765.
- Friedman-Hill, S. R., Robertson, L. C., & Treisman, A. (1995). Parietal contributions to visual feature binding: evidence from a patient with bilateral lesions. *Science*, *269*, 853–855.
- Goebel, R., Khorram-Sefat, D., Muckli, L., Hacker, H., & Singer, W. (1998). The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *European Journal of Neuroscience*, *10* (5), 1563–1573.
- Green, D. M., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Grill-Spector, K., Kushnir, T., Itzhak, Y., & Malach, R. (2000). The dynamics of object-selective activation correlate with recognition performance in humans. *Nature Neuroscience*, *3*, 837–843.
- He, S., Cohen, E. R., & Hu, X. (1998). Close correlation between activity in brain area MT/V5 and the perception of a visual motion aftereffect. *Current Biology*, *5*, 1215–1218.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences USA*, *96*, 9379–9384.
- Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.
- Kanwisher, N. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, *27*, 117–143.
- Kanwisher, N. (1991). Repetition blindness and illusory conjunctions: errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 404–421.
- Kanwisher, N., Downing, P., Epstein, R., & Kourtzi, Z. (in press). Functional neuroimaging of human visual recognition. In Kingstone, A., & Cabeza, R. (Eds.), *The handbook on functional neuroimaging*. Cambridge, MA: MIT Press.
- Kanwisher, N., McDermott, J., & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Kanwisher, N., & Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Neuroscience Reviews*. Manuscript submitted for publication.
- Kanwisher, N., Woods, R., Iacoboni, M., & Mazziotta, J. (1996). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, *9*, 133–142.
- Kleinschmidt, A., Buechel, C., Huton, C., & Frackowiak, R. S. J. (1998). Hysteresis effects in figure-ground segmentation. *NeuroImage*, *7*, S356.
- Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: changing views in perception. *Trends in Cognitive Sciences*, *3*, 254–264.
- Logothetis, N. K. (1998). Single units and conscious vision. *Proceedings of the Royal Society of London, Series B*, *353*, 1801–1818.
- Luck, S. J., & Girelli, M. (1998). Electrophysiological approaches to the study of selective attention in the human brain. In R. Parasuraman (Ed.), *The attentive brain* (pp. 71–94). Cambridge, MA: MIT Press.
- Luck, S. J., Vogel, E.K. & Shapiro K.L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*, 616–618.
- Lumer, E. D., Friston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*, 1930–1934.
- Malach, R., Reppas, J. B., Benson, R. B., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Brady, T. J., Rosen, B. R., & Tootell, R. B. H. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences USA*, *92*, 8135–8138.
- Marcel, A. J. (1983). Conscious and unconscious perception: an approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15*, 238–300.
- McCarthy, G., Puce, A., Gore, J., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, *9*, 605–610.
- McDermott, J. (1996, April). *Sleep induced changes in auditory processing: an fMRI study*. Poster presented at the 3rd annual meeting of the Cognitive Neuroscience Society, San Francisco, CA.

- Mel, B. W., & Fiser, J. (2000). Minimizing binding errors using learned conjunctive features. *Neural Computation*, 12, 731–762.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Milner, A. D., & Rugg, M. D. (1992). *The neuropsychology of consciousness*. London: Academic Press.
- Nagel, T. (1974). What is it like to be a bat? *Mortal questions* (pp. 165–180). Cambridge, MA: Cambridge University Press.
- O'Craven, K., & Kanwisher, N. (in press). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*.
- O'Craven, K. M., Rosen, B. R., Kwong, K. K., Treisman, A., & Savoy, R. L. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron*, 18, 591–598.
- O'Craven, K., Downing, P., & Kanwisher, N. (1999). fMRI Evidence for objects as the units of attentional selection. *Nature*, 401, 584–587.
- Palmer, S. (1999). *Vision science*. Cambridge, MA: MIT Press.
- Penfield, W., & Perot, P. (1963). The brain's record of auditory and visual experience. *Brain*, 86, 595–696.
- Polonsky, A., Blank, R., Braun, J., & Heeger, D. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. Manuscript submitted for publication.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156–161.
- Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III. Effects of top-down processing on face-specific potentials. *Cerebral Cortex*, 9, 445–458.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, 18, 849–860.
- Rees, G., Russell, C., Frith, C. D., & Driver, J. (1999). Inattention blindness versus inattention amnesia for fixated but ignored words. *Science*, 286, 2504–2507.
- 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*, 123, 1624–1633.
- Rubin, N., Nakayama, K., & Shapley, R. (1997). Abrupt learning and retinal size specificity in illusory-contour perception. *Current Biology*, 7, 461–467.
- Salzman, C.D., Britten, K.H., & Newsome, W.T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, 346, 174–177.
- Schacter, D. L., McAndrews, M. P., & Moscovitch, M. (1988). Access to consciousness: dissociations between implicit and explicit knowledge in neuropsychological syndromes. In L. Weiskrantz (Ed.), *Thought without language*. Oxford: Oxford University Press.
- Singer, W. (2000). Response synchronization: a universal coding strategy for the definition of relations. In M. Gazzaniga, *The new cognitive neurosciences*. Cambridge, MA: MIT Press.
- Spelke, E., Vishton, P., & Von Hofsten, C. (1995). Object perception, object-directed action, and physical knowledge in infancy. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 165–179). Cambridge, MA: MIT Press.
- Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21, 753–759.
- Tootell, R. B. H., Hadjikhani, N., & Somers, D. C. (1999). *fMRI reveals subthreshold activation in human visual cortex: implications for consciousness*. Talk presented at the 29th annual meeting of the Society for Neuroscience, Miami Beach, FL.
- Tootell, R. B., Reppas, J. B., Dale, A. M., Look, R. B., Sereno, M. I., Malach, R., Brady, T. J., & Rosen, B. R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature*, 375, 139–141.
- Tootell, R. B. H., Reppas, J. B., Kwong, K. K., Malach, R., Brady, T., Rosen, B., & Belliveau, J. (1995). Functional analysis of human MT/V5 and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, 15 (4), 3215–3230.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141.

- Whalen, P.J., Rauch, S.L., Etcoff, N.L., McInerney, S.C., Lee, M.B., & Jenike, M.A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience*, *18*, 411–418.
- Vignat, J. P., Chauvel, P., & Halgren, E. (2000). Localized face-processing by the human prefrontal cortex: 1. Stimulation-evoked hallucinations of faces. In N. Kanwisher, & M. Moscovitch (Eds.), *The cognitive neuroscience of face processing* (pp. 281–292). East Sussex: Psychology Press.
- von Helmholtz, H. (1962). *Helmholtz's treatise on physiological optics* (J. P. C. Southall, Trans.). New York: Dover. (Original work published 1866)
- Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement visual attention. *Neuron*, *4*, 747–764.
- Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, *79*, 1574–1578.
- Wolfe, J. M. (1986). Stereopsis and binocular rivalry. *Psychology Review*, *93*, 269–282.
- Zihl, J., von Cramon, D., & Mai, N. (1983). Selective disturbance of movement vision after bilateral brain damage. *Brain*, *106*, 313–340.