The Role of Feedback in Visual Attention and Awareness

STEPHEN L. MACKNIK AND SUSANA MARTINEZ-CONDE

ABSTRACT The mammalian visual system includes numerous brain areas that are profusely interconnected. With few exceptions, these connections are reciprocal. Anatomical feedback connections in general outnumber feedforward connections, leading to widespread speculation that feedback connections play a critical role in visual awareness. However, evidence from physiological experiments suggests that feedback plays a modulatory role, rather than a driving role. Here we discuss theoretical constraints on the significance of feedback's anatomical numerical advantage, and we describe theoretical limits on feedback's potential physiological impact. These restrictions confine the potential role of feedback in visual awareness and rule out some extant models of visual awareness that require a fundamental role of feedback. We propose that the central role of feedback is to maintain visuospatial attention, rather than visual awareness. Our conclusions highlight the critical need for experiments and models of visual awareness that control for the effects of attention. As a matter of clarity in this chapter: by “visual awareness” or “visibility” we mean the conscious perception that a stimulus is visible. Thus, for the purposes of this discussion, we use the terms visual awareness, visibility, and consciousness interchangeably.

Anatomical observations of feedback in the visual system

The visual areas of the brain are interconnected in a complex pattern of feedforward, lateral, and feedback pathways (Felleman & Van Essen, 1991). Feedback connections are ubiquitous throughout the cortex, and subcortical regions in ascending hierarchical pathways also receive a large amount of feedback from cortical areas (Erisir, Van Horn, & Sherman, 1997; Fitzpatrick, Usrey, Schofield, & Einstein, 1994; Guillery, 1969; Sherman & Guillery, 2002).

Anatomy of feedback in the LGN Corticogeniculate input is the largest source of synaptic afferents to the lateral geniculate nucleus (LGN). Whereas retinal afferents only encompass 25% of the total number of inputs to LGN interneurons, 37% of the synaptic contacts come from the cortex. In the case of relay cells, the respective percentages are 12% versus 56% (Montero, 1991). Similar estimates have been calculated in the primate, and the general agreement is that the cortical-to-retinal input ratio is between 1:2 and 1:6 in both cats and primates (Erisir et al., 1997; Fitzpatrick et al., 1994; Guillery, 1969; Sherman & Guillery, 2002; Van Horn, Erisir, & Sherman, 2000). Boyapati and Henry (1984) concluded that feedback connections from the cat visual cortex to the LGN concentrated a larger fraction of fine axons than feedforward geniculocortical connections, presumably resulting in comparatively slower conduction speeds.

These and other considerations concerning the synaptic size, efficacy, and contribution of feedback connections underscore the potential mistake in assuming that a numerically larger number of inputs means that those inputs are functionally most important (Sherman & Guillery, 2002).

Anatomy of feedback in the primary visual cortex Cortical feedforward pathways usually project from the supragranular layers of visual areas early in the hierarchy (less than 10–15% of the connections may arise from deep layers) and terminate in layer 4 of areas later in the hierarchy. In contrast, feedback projections usually arise from the infragranular layers of later areas and terminate outside of layer 4 in the early areas (Barone, Batardiere, Knoblauch, & Kennedy, 2000; Felleman & Van Essen, 1991; Hilgetag, O’Neill, & Young, 1996a, 1996b; Maunsell & Van Essen, 1983).

Direct feedforward projections to primate area V1 (also called primary visual cortex, striate cortex, and Brodmann’s area 17) originate from the pulvinar, LGN, claustrum, nucleus paracentralis, raphe system, locus coeruleus, and nucleus basalis of Meynert (Blasdel & Lund, 1983; Doty, 1983; Fitzpatrick et al., 1994; Hendry & Yoshioka, 1994; Lachica & Casagrande, 1992; Ogren & Hendrickson, 1976; Perkel, Bullier, & Kennedy, 1986; Rezak & Benevento, 1979). Direct feedback projections from V1 extend to V2, V3, V5 or MT, MST, and FEF (Boussaoud, Ungerleider, & Desimone, 1990; Livingstone & Hubel, 1987; Lund, Lund, Hendrickson, Bunt, & Fuchs, 1975; Maunsell & Van Essen, 1983; Shipp & Zeki, 1989; Ungerleider & Desimone, 1986a, 1986b). Direct feedback projections to V1 originate from V2, V3, V4, V5 or MT, MST, FEF, LIP, and inferotemporal cortex (Barone et al., 2000; Perkel et al., 1986; Rockland, Saleem, & Tanaka, 1994; Shipp & Zeki,
Physiological observations of feedback in the visual system

Methodological Shortcomings in Physiological Studies of Feedback Some visual physiology studies have found that feedback connections between the secondary and primary visual cortices enhance or decrease neuronal responsiveness without fundamentally altering response specificity (Martinez-Conde et al., 1999) (see figure 81.1). These studies were conducted by microinjecting small amounts of neuronal modulators into area 18 of the cat while recording from the corresponding retinotopic position in area 17. This method is accurate in its assessment of feedback effects because it sequesters the source of neuronal enhancement and suppression to a small focol region that cannot directly affect the neuronal responses of the neurons being recorded in the area of interest. Thus the only possible cause of the response modulation in area 17 was the feedback connection from area 18. Another positive aspect of this technique is that the effects are fully reversible, which is not a feature shared by the ablation (Super & Lamme, 2007) and lesion methods.

Other studies have proposed a more significant physiological role for feedback in the visual system. However, these studies have generally used alternative methods such as ablation, cooling, transcranial magnetic stimulation, and direct pharmacological manipulation of the neurons being recorded. Such techniques are usually disadvantageous in that they are nonfocal, nonreversible, and/or may have unknown or poorly understood nonspecific effects on the physiological milieu of the neurons being directly recorded (such as by changing the pH, osmolarity, temperature, or other effects). Nonfocal and/or nonreversible techniques may also affect the vasculature feeding the targeted neurons or fibers of passage with known or unknown connectivity (either direct or indirect) to the targeted neurons. Thus the results obtained are more difficult to interpret, as the responses of the targeted neurons may have been affected in ways unrelated to any putative role of feedback.

Also, as we will discuss more fully in a later section, it is critical that physiological measurements of feedback, as they relate to awareness, be conducted with careful controls for the effects of attention, as well as its underlying circuits. This is a necessary precaution, as the physiological process of attention is differentiable from that of awareness (Koch & Tsuchiya, 2007). For a comprehensive discussion of this issue, please see Koch (chapter 79, this volume).

What Do We Mean by Feedback? For the purposes of this chapter, we restrict our definition of the word “feedback” to the long-range fibers that connect a higher brain area to a lower brain area, within an ascending sensory system. In this definition, the same information arrives to the same neural circuit at least twice: first as it feeds forward through the system and later again as it feeds back.

Other types of feedback loops in the brain are not discussed in this chapter. For instance, information may flow upward from one thalamic nucleus to the cortex (i.e., from the LGN to area V1) and then back down to a different thalamic nucleus (i.e., the pulvinar) (Rockland, 1996). In this case, it could be said that the thalamus as a whole sends information to, and receives feedback from, area V1 (as Rockland describes it). However, the feedforward and feedback projections are mediated by two separate thalamic nuclei. Thus this type of circuit does not meet the definition of feedback used here.

Here we will discuss specifically those reciprocal connections between visual areas of the geniculocortical pathway. Thus the feedback we will consider entails connections from neurons processing more complex visual information (and having more complex receptive fields) to neurons processing less complex visual information (which have simpler and less selective receptive fields) (figure 81.2).

This chapter aims to describe the powerful constraints on the functional role of feedback, even within such an ordinary and basic neural system. Figure 81.2A illustrates the basic connectivity between an area of the geniculocortical pathway and the next area up in the hierarchy (i.e., the LGN and area V1). The lower, simpler level of processing feeds forward to the higher level. There, information is further processed by neural circuits with more complex receptive fields. The higher, more complex level then feeds back information to the simpler level. If such a feedback connection is functionally effective, the receptive fields from the lower level will acquire the specificity and complexity that characterize the higher-level receptive fields (figure 81.2B). This physiological prediction should apply to any feedback pathways that are both engaged and significant in strength.

Physiology of Feedback in the LGN Corticogeniculate connections to the LGN are retinotopically organized, and
they preferentially end on LGN layers with the same ocular dominance as the cortical cells of origin (Murphy & Sillito, 1996). Although only 12–58% (Montero, 1991) of the inputs to geniculate cells are retinal in origin, these synapses drive the primary responses of geniculate relay cells, whereas feedback inputs play a modulatory role (Sherman & Guillery, 1998, 2002).

**Physiology of Feedback in the Primary Visual Cortex**

Cortico-cortical feedback connections are also retinotopically specific (Salin, Girard, Kennedy, & Bullier, 1992). For instance, there is a functional projection from area 18 to area 17 neurons with similar retinotopic locations (Bullier, McCourt, & Henry, 1988; Martinez-Conde et al., 1999; Salin et al., 1992; Salin, Kennedy, & Bullier, 1995). Girard, Hupe, and Bullier (2001) found that feedforward and feedback connections between areas V1 and V2 of the monkey have similarly rapid conduction speeds.

In the cat visual cortex, electrical stimulation from areas area 18 and area 19 demonstrated 50% of monosynaptic connections with superficial layers of area 17, in regions with similar functional properties, such as retinotopic location (Bullier et al., 1988). Mignard and Malpeli (1991) also found that inactivation of area 18 in the cat led to decreased responses in area 17. Martinez-Conde and colleagues (1999) found that focal reversible inactivation of area 18 produced suppressed or enhanced visual responses in area 17 neurons with a similar retinotopy. In most area 17 neurons, orientation bandwidths and other functional characteristics remained unaltered, suggesting that feedback from area 18 modulates area 17 responses without fundamentally altering their specificity.

In the squirrel monkey, Sandell and Schiller (1982) found that most area V1 cells decreased their visual responses when area V2 was reversibly cooled, although a few cells became more active. Orientation selectivity remained

---

**Figure 81.1** Reversible removal of feedback from area 18 to area 17 in the cat. (A) Orientation tuning curve of a cell from layers 2/3 of area 17. (B) Orientation tuning curve of the same cell during GABA application in area 18. Note that, although the firing rate of this area 17 neuron increases significantly, its orientation selectivity is virtually unchanged in the absence of feedback from area 18. Thus feedback modulates the magnitude of the neuronal responses but does not affect their functional specificity. (C) Orientation tuning curve of the same cell after area 18 blockade. (D) Solid line, control tuning curve of an area 18 cell recorded simultaneously. Dotted line, tuning curve of the same cell after blockade. Inset: receptive fields of both cells. For clarity, only the preblockade post-stimulus time histograms (PSTHs) are shown in D. The number of spikes for PSTHs of areas 17 and 18 are indicated at bottom of B and D, respectively. Bin size: 100 ms. Time base: 1 s. (Reprinted from Martinez-Conde et al., 1999.)
unchanged, although direction selectivity decreased in some instances. Bullier, Hupe, James, and Girard (1996) reported in the cynomologous monkey that, following GABA inactivation of area V2, V1 neurons showed decreased or unchanged responses in the center of the classical receptive field, but increased responses in the region surrounding it. These results were supported by subsequent findings in areas V1, V2, and V3 following area MT inactivation (Hupe et al., 1998). More recently, Angelucci and colleagues (Angelucci & Bressloff, 2006; Angelucci, Levitt, & Lund, 2002) have suggested that area V1 extraclassical receptive field properties arise from area V2 feedback.

In summary, physiological studies as a whole suggest that feedback connections in the visual system may play a modulatory role, rather than a driving role, in shaping the responses of hierarchically lower areas. This evidence agrees with the “no-strong-loops” hypothesis formulated by Crick and Koch (1998b). The no-strong-loops hypothesis proposes that all strong connections in the visual system are of the feedforward type. That is, “the visual cortex is basically a feedforward system that is modulated by feedback connections,” which is “not to say that such modulation may not be very important for many of its functions” (p. 248). Crick and Koch argued that “although neural nets can be constructed with feedback connections that form loops, they do not work satisfactorily if the excitatory feedback is too strong.” Similarly, if feedback connections formed “strong, directed loops” in the brain, the cortex would as a result “go into uncontrolled oscillations.” Therefore, the relative number of feedback versus feedforward anatomical connections to any given visual area may be misleading as to the respective roles of such connections. For instance, the fact that the cat LGN receives substantially larger numbers of synapses from the cortex than from the retina (Montero, 1991) does not necessarily mean that corticogeniculate connections are more important than retinogeniculate connections in determining the response characteristics of LGN neurons.

Although the role of feedback modulation in our visual perception remains unclear, one possibility is that feedback may be involved in attentional mechanisms (Martinez-Conde et al., 1999). We will discuss this idea more fully in the next section.

The role of feedback in attention

Based on the evidence we have reviewed, one potentially important role for feedback may be to carry attentional modulation signals. Other modulatory roles for feedback remain possible, but none are as clearly established. Thus it may be that the sole effect of all feedback connectivity is to facilitate and suppress attention. At first, given the massive amount of anatomical feedback versus feedforward connections, this possibility may seem unlikely. Indeed, the great extent of feedback connectivity suggests to some that feedback must have a large number of roles (Sherman & Guillery, 2002; Sillito & Jones, 1996). However, we will argue here that the great number of feedback connections may potentially be explained by the need for top-down attentional modulation alone. Ascending circuits in the visual system primarily form a labeled-line hierarchy, and so feedback connections necessarily require more wiring than feedforward connections to send back even the simplest signal.

To illustrate the logic of this argument, let us consider the anatomical connectivity between the LGN and V1 (figure 81.3). As previously described, LGN relay cells receive more numerous feedback connections from the cortex than feedforward inputs from the retina. However, because V1 receptive fields are orientation selective and LGN receptive fields are not, any functionally significant feedback from V1 to a given retinotopic location in the LGN must represent many, or all, orientations. (One should note that Vidyasagar & Urbas, 1982, found slight orientation biases in LGN receptive fields; these biases were much smaller than the strong orientation selectivity found in V1.) That is, for each unoriented feedforward connection from the LGN to V1, there must be many oriented feedback connections from V1 to the LGN, each with a different orientation, so that the sum of all feedback projections spans all the orientation space. Otherwise, if the orientation space of the feedback connections were not filled completely, LGN receptive fields would show a substantial orientation bias. Thus anatomical feedback connectivity must be large so as to represent the entire orientation space at each retinotopic location. However, because

**Figure 81.2** A generalized model of the effect of feedback in a hierarchy of simple to complex neural processing. (A) In a functional hierarchy, information processing becomes more complex as one ascends in the pathway. (B) When feedback is engaged, the lower levels of the hierarchy take on the more complex properties of the upper levels.
of their orientation selectivity, only a fraction of all feedback connections will be active at any given time, depending on the orientation of the visual stimulus, whereas the feedforward connections will be active irrespective of stimulus orientation. In summary, the massive feedback versus feedforward connectivity ratio can be misleading: this large ratio does not necessarily mean that feedback signals are more important or more physiologically relevant than feedforward signals. Because higher visual areas are more selective than lower visual areas, only a relatively small fraction of the feedback may be expected to be active at any given moment. Thus feedback connections may need to tile the entire space of receptive field properties of the higher level; otherwise, the feedback would impose high-level receptive field characteristics on the receptive fields of lower areas.

Figure 81.4 illustrates this idea in terms of the feedback from dichoptic to monoptic levels of the visual pathway. To be clear about the jargon: “monocular” means “with respect to a single eye,” and “monoptic” means either “monocular” or “not different between the two eyes.” “Binocular” means “with respect to both eyes,” and “dichoptic” means “different in the two eyes.” Thus subcortical levels of the visual system are monoptic (because the cells are monocular), whereas cortical visual areas that have binocular circuits may potentially process dichoptic information.

To summarize: because receptive fields in ascending pathways become more selective, larger, and more complex in their properties as one rises through the higher levels of the brain’s hierarchies, anatomical feedback connections must be more numerous than feedforward connections. Otherwise, the hierarchical nature of the visual system would be diminished (figure 81.2). Moreover, the numerical

Figure 81.3 A model of the effects of feedback from area V1 on an LGN neuron. Numerous V1-oriented cells must feed back to every feedforward LGN cell in order to account for the lack of significant orientation bias in LGN receptive fields.

Figure 81.4 The effects of feedback from dichoptic levels to monoptic levels of visual processing. (A) A general model of early visual binocular integration in the absence of feedback connections. (B) If significant feedback existed between dichoptic levels of processing and earlier monoptic levels, the earlier levels should acquire the properties of the dichoptic levels (i.e., they should become dichoptic by virtue of the feedback). (Reprinted from Macknik, 2006.)
advantage of feedback over feedforward connections should be expected even if there is just a single functional role for feedback (i.e., attentional modulation).

If we combine these ideas with the Crick and Koch’s no-strong-loops hypothesis and the physiological findings indicating that feedback plays a modulatory rather than a driving role, we may conclude that feedback inputs have more moderate physiological effects than feedforward inputs, despite being anatomically more numerous. This concept is supported by the known physiology: besides their lack of orientation selectivity, another feature that distinguishes LGN from V1 receptive fields is their smaller size (Allman, Miez, & McGuinness, 1983; Desimone, Schein, Moran, & Ungerleider, 1983; Kastner, Nothdurft, & Pigarev, 1999; Knierim & Van Essen, 1991; Zeki, 1978a, 1978b). If feedback connections from V1 to the LGN were functionally as strong as their feedforward counterparts, LGN receptive fields would be as large as V1 receptive fields, but they are not. That is, because LGN receptive fields are smaller than V1 receptive fields, feedback from V1 to the LGN must be weaker than the retinal inputs.

It follows from these ideas that when feedback is operational, some receptive field properties, such as size, which continues to increase throughout the visual hierarchy (Allman et al., 1983; Desimone et al., 1983; Kastner et al., 1999; Knierim & Van Essen, 1991; Zeki, 1978a, 1978b) will be fed back from higher to lower levels. Thus we may predict that, if attention is carried by feedback connections, earlier receptive fields should increase in size when attention is applied actively. This prediction has been confirmed experimentally (He, Cavanagh, & Intriligator, 1996; Williford & Maunsell, 2006). Chen and colleagues (2008) recently showed that attentional modulation of V1 neurons in the awake monkey is spatially specific: increasing task difficulty enhanced V1 neuronal firing rate at the focus of attention and suppressed it in surrounding regions, in support of Desimone and Duncan’s (1995) center-surround model of attention (figure 81.5). Moreover, response enhancement and suppression were mediated by distinct neuronal populations that differed in direction selectivity, spike width, interspike-interval distribution and contrast sensitivity. This finding suggested that attentional feedback facilitates and suppresses distinct populations of neurons in the primary visual cortex.

To conclude, feedback connections may potentially have no other function than to modulate (facilitate or suppress) feedforward signals as a function of attentional load.

**The role of visual masking, binocular rivalry, attention, and feedback in the study of visual awareness**

Let us assume that visual awareness is correlated to brain activity within specialized neural circuits, and that not all brain circuits maintain awareness. It follows that the neural activity that leads to reflexive or involuntary motor action may not correlate with awareness because it does not reside within awareness-causing neural circuits (Macknik & Martinez-Conde, 2009).

![Figure 81.5](image_url)

**Figure 81.5** V1 attentional response modulation in awake monkey single cells during hard and easy tasks. (A) Temporal structure of a trial. Two rhesus monkeys were trained to fixate on a small cross while covertly attending to a spatial location that was cued at the beginning of each trial. The cue was a thin red ring with a diameter that was threefold larger than the diameter of the neuronal receptive field (RF). Following the cue, drifting gratings were presented simultaneously at five different spatial locations for 1.5–3 s. Following a randomized period of time, one of the gratings changed color/luminance, and the animal was tasked with detecting the change by releasing a bar within 0.5 s. The attentional modulations were measured at the last cycle of the drifting grating before the color change. (B) The color change could be easy or hard to detect and could occur inside or outside of the receptive field. (C) The number of cells that were significantly modulated by attention (red) was much lower during the easy task (top) than during the hard task (bottom). A subset of 8 cells was significantly modulated by attention during both the easy and the hard task ($P < 0.05$). (D) An increase in task difficulty leads to an enhancement of V1 visual responses at the focus of attention and a suppression outside the focus. Difficulty-enhanced V1 neurons have poor directional selectivity and broad interspike interval distributions (sustained responses). Difficulty-suppressed V1 neurons are directional selective and have tight interspike interval distributions (transient responses). (Reprinted from Chen et al., 2008.)
Let us also propose that there is a “minimal set of neural conditions” necessary to achieve conscious visibility (see Chalmers, 2000, for an excellent review of this idea). Such conditions take the form of a specific type (or types) of neural activity within a subset of brain circuits. The minimal set of conditions will not be met if the correct circuits have the wrong type of activity (too much activity, too little activity, sustained activity when transient activity is required, etc). Moreover, if the correct type of activity occurs, but solely within circuits that do not maintain awareness, visibility will also fail. Finding the conditions in which visibility fails is critical to the research described here; although we do not yet know what the minimal set of conditions is, we can nevertheless systematically modify potentially important conditions (change neural circuits, modify levels of activity) and see if they result in stimulus invisibility. It so, the modified condition would be part, potentially, of the minimal set of neural conditions necessary to maintain visibility.

To establish the minimal set of conditions for visibility we need to answer at least four questions (Macknik, 2006). The questions and their (partial) answers are as follows:

1. What stimulus parameters are important to visibility?
   A. The spatiotemporal edges (also curves and corners) of stimuli are the most important parameters to stimulus visibility (Macknik, Martinez-Conde, & Haglund, 2000; Troncoso, Macknik, & Martinez-Conde, 2005; Troncoso et al., 2007).

2. What types of neural activity best maintain visibility (transient versus sustained firing, rate codes, bursts of spikes, etc.—that is, what is the neural code for visibility)?

3. What brain areas must be active to maintain visibility?
   A. Visual areas downstream of V2, lying within the occipital lobe, must be active to maintain visibility of simple unattended targets (Macknik, 2006; Macknik & Martinez-Conde, 2004a; Tse, Martinez-Conde, Schlegel, & Macknik, 2005).

4. What specific neural circuits within the relevant brain areas maintain visibility?
   A. The specific circuits that maintain visibility are currently unknown, but their responsivity is modulated by lateral inhibition (Macknik, 2006; Macknik & Livingstone, 1998; Macknik & Martinez-Conde, 2004a, 2004b; Macknik et al., 2000).

We must also determine the set of standards that will allow us to conclude that any given brain area, or neural circuit within an area, is responsible for generating a conscious
experience. Parker and Newsome developed a “list of idealized criteria that should be fulfilled if we are to claim that some neuron or set of neurons plays a critical role in the generation of a perceptual event” (Parker & Newsome, 1998, p. 230). If one replaces the words “perceptual event” with “conscious experience,” Parker and Newsome’s list can be used as an initial foundation for the neurophysiological requirements needed to establish whether any given neuron or brain circuit may be the neural substrate of awareness (Macknik & Martinez-Conde, 2007). Parker and Newsome’s list (pp. 230–231) follows:

1. The responses of the neurons and of the perceiving subject should be measured and analyzed in directly comparable ways.

2. The neurons in question should signal relevant information when the organism is carrying out the chosen perceptual task. Thus the neurons should have discernible features in their firing patterns in response to the different external stimuli that are presented to the observer during the task.

3. Differences in the firing patterns of some set of the candidate neurons to different external stimuli should be sufficiently reliable in a statistical sense to account for, and be reconciled with, the precision of the organism’s responses.

4. Fluctuations in the firing of some set of the candidate neurons to the repeated presentation of identical external stimuli should be predictive of the observer’s judgment on individual stimulus presentations.

5. Direct interference with the firing patterns of some set of the candidate neurons (e.g., by electrical or chemical stimulation) should lead to some form of measurable change in the perceptual responses of the subject at the moment that the relevant external stimulus is delivered.

6. The firing patterns of the neurons in question should not be affected by the particular form of the motor response that the observer uses to indicate his or her percept.

7. Temporary or permanent removal of all or part of the candidate set of neurons should lead to a measurable perceptual deficit, however slight or transient in nature.

However, visual circuits that may pass muster with Parker and Newsome’s guidelines may nevertheless fail to maintain awareness, as we shall explain. To guide the search for the neural correlates of consciousness (NCC), the minimal neuronal mechanisms jointly sufficient for a particular percept (Crick & Koch, 1995), some additional standards must be added.

The first additional standard concerns the use of illusions as the tool of choice to test whether a neuronal population or circuit may maintain awareness. Visual illusions, by definition, dissociate the subject’s perception of a stimulus from its physical reality. Thus visual illusions are powerful devices in the search for the NCC (Myerson, Miezin, & Allman, 1981), as they allow us to distinguish the neural responses to the physical stimulus from the neural responses that correlate to perception. Our brains ultimately construct our perceptual experience, rather than reconstruct the physical world (Macknik & Hagnud, 1999). Therefore an awareness-maintaining circuit should express activity that matches the conscious percept, irrespective of whether it matches the physical stimulus. Neurons (circuits, brain areas) that produce neural responses that fail to match the percept provide the most useful information because they can be ruled out, unambiguously, as part of the NCC. As a result, the search for the NCC can be focused to the remaining neural circuits. Conversely, neurons that do correlate with perception are not necessarily critical to awareness, as they may simply play a support role (among other possibilities) without causing awareness themselves.

The second additional standard derives from a major contribution of Crick and Koch’s: the distinction between explicit and implicit representations in the study of visual awareness (Crick & Koch, 1998a). In an explicit representation of a stimulus feature, there is a set of neurons that represents that feature without substantial further processing. In an implicit representation, the neuronal responses may account for certain elements of a given feature; however, the feature itself is not detected at that level. For instance, all visual information is implicitly encoded in the photoreceptors of the retina. The orientation of a stimulus, however, is not explicitly encoded until area V1, where orientation-selective neurons and functional orientation columns are first found. Crick and Koch propose that there is an explicit representation of every conscious percept.

Here we offer the following corollary to Crick and Koch’s idea of explicit representation: Before one can test a neuronal population or circuit for its role in the NCC, the specific neurons (or the population/circuit being tested) must be shown to explicitly process the test stimulus. That is, the neurons must respond to the test stimulus or show selective tuning to some range of features of the stimulus. This corollary constrains the design of neurophysiological experiments aimed to test the participation of specific neurons, circuits, and brain areas in the NCC. For instance, if one found that retinal responses do not correlate with auditory awareness, such a discovery would not carry great weight. The neurons in the eye do not process auditory information, and so it is not appropriate to test their correlation to auditory perception. However, this caveat also applies to more nuanced stimuli. What if V1 activity was tested for its correlation to the perception of faces versus houses? Faces and houses are visual stimuli, but V1 has never been shown to process faces or houses explicitly, despite the fact that visual information about faces and houses must implicitly be represented in V1. Therefore, one cannot test V1’s role in the NCC using houses versus faces and expect to come to any meaningful
Binocular rivalry has been used as a tool to assess the NCC, but it has generated controversy because of conflicting results (Macknik & Martinez-Conde, 2004a; Tse et al., 2005). Some human fMRI studies report that BOLD activity in V1 correlates with awareness of binocular rivalry percepts (Lee, Blake, & Heeger, 2005; Polonsky, Blake, Braun, & Heeger, 2000; Tong & Engel, 2001). In contrast, other human fMRI studies (Lumer, Friston, & Rees, 1998), as well as neuronal recording studies in nonhuman primates (Leopold & Logothetis, 1996), report that activity in area V1 does not correlate with visual awareness of binocular rivalry percepts. One possible reason for this discrepancy is that none of these studies determined that the visual areas tested contained the interocular suppression circuits necessary to mediate binocular rivalry. That is, since binocular rivalry is a process of interocular suppression, the neural circuits underlying the perception of binocular rivalry must be shown to produce interocular suppression—explicitly. Otherwise, it cannot be demonstrated that binocular rivalry is a valid stimulus for testing the NCC in those areas. Thus awareness studies using binocular rivalry are valid only in areas that have been shown to maintain interocular suppression. If binocular rivalry fails to modulate activity within a visual area, one cannot know, by using binocular rivalry alone, if the perceptual modulation failed because awareness is not maintained in that area or because the area does not have circuits that drive interocular suppression. This is more than just a theoretical possibility: as we will describe, we have shown in the human and the macaque monkey that the initial binocular neurons of the early visual system (areas V1 and V2) are binocular for excitation but monocular for inhibition. That is, they fail to process interocular suppression explicitly (Macknik & Martinez-Conde, 2004a; Tse

Figure 81.7 (A) Summary statistics of monoptic versus dichoptic masking responses in the LGN and area V1 of the macaque monkey. Monoptic (black bars) and dichoptic (white bars) masking magnitude as a function of cell type: LGN, V1 monocular, V1 binocular (nonresponsive to dichoptic masking), and V1 binocular (responsive to dichoptic masking) neurons. Inset shows the linear regression of dichoptic masking magnitude in V1 binocular neurons as a function of their degree of binocularity (all neurons plotted were significantly binocular as measured by their relative responses to monocular targets presented to the two eyes sequentially). BI of 0 indicates that the cells were monocular, while a BI of 1 means both eyes were equally dominant. (Reprinted from Macknik & Martinez-Conde, 2004a.) (B) Monoptic and dichoptic masking magnitude as a function of occipital retinotopic brain area in the human. Negative values indicate decreased visual masking (increased target visibility), whereas values ≥ 0 indicate increased masking (decreased target visibility). (Reprinted from Tse, Martinez-Conde, Schlegel, & Macknik, 2005.)
et al., 2005) (figure 81.7). There is no control condition with which to establish whether binocular rivalry fails because of a (mundane) lack of interocular suppression or (more interestingly) fails because of a lack of awareness maintaining circuits. One could address this issue by using binocular rivalry in tandem with a different stimulus to test for the explicit representation and strength of interocular suppression, such as visual masking stimuli. In visual masking, a monoptic form of the illusion is available, and so one can distinguish failures of interocular suppression from failures of visual awareness. But if one were to use masking stimuli to assess interocular suppression in a given visual area, then the role of such area in maintaining visibility and awareness would have also been established (in the absence of binocular rivalry stimulation), thus obviating the need for subsequent testing with binocular rivalry stimuli. Because one must rely on non-binocularly-rivalrous stimuli to determine the explicit representation and strength of interocular suppression in a given area, it is not possible to unambiguously interpret the neural correlates of perceptual state using binocular rivalry alone in any visual area, irrespective of the strength of the binocular rivalry response.

Our visual masking studies have shown that binocular neurons in areas V1 (the first stage in the visual hierarchy where information from the two eyes is combined) and V2 of humans and monkeys can integrate excitatory responses from the two eyes (Macknik & Martinez-Conde, 2004a; Tse et al., 2005) (figure 81.7). However, these same neurons do not express interocular suppression between the eyes. That is, binocular neurons in V1 are largely binocular for excitation while nevertheless being monocular for suppression (that is, input from one eye will not suppress the firing rate of a V1 or V2 cell that is primarily tuned for input from the opposite eye). Because most early binocular cells do not explicitly process interocular suppression, these neurons cannot explicitly process binocular rivalry. Further, binocular rivalry cannot distinguish between the role of interocular suppression and the role of awareness at any level of the visual system. Therefore no conclusions can be reached about the localization of the NCC to specific parts of the visual system based on binocular rivalry studies alone. If a given visual area does not correlate to binocular rivalry, it may simply mean that interocular suppression is not at play in that area, rather than that area is not maintaining awareness. However, these findings also beg the question of why some studies have found binocular rivalry modulation in low-level visual areas (Haynes, Driver, & Rees, 2005; Lee et al., 2005; Polonsky et al., 2000; Tong & Engel, 2001; Wunderlich, Schneider, & Kastner, 2005). One possible reason for this paradox is that these studies failed to control for the effects of attentional feedback, thus confounding apparent modulation to interocular suppression with attentional modulation. Because the subjects in these studies needed to attend to the binocular rivalry stimuli, attention itself, rather than binocular rivalry, may have produced the retinotopic activation found.

Monoptically and dichoptically presented visual-masking illusions (such as forward and backward masking and the standing wave) can differentiate between interocular suppression and awareness, and thus they are immune to these shortcomings. Therefore visual masking is an ideal illusion to isolate the NCC. Further, visual-masking illusions allow us to examine the brain’s response to the same physical target under varying levels of visibility (unlike in binocular rivalry, where one only measures which of two rivalrous percepts was dominant at any given time, without consideration of how visible that percept was). Thus by quantifying the perceptual and physiological effects of visible versus invisible (masked) targets we will determine many, if not all, of the conditions that cause visibility.

We propose that, to test for explicit processing in neuronal populations or circuits, one should use a visual illusion, such as visual masking, that can be presented in at least two modes of operation: one mode to ensure that the neural circuit in question is able to process the stimulus explicitly and another mode to test the correlation to awareness. In visual masking, the monoptic mode establishes that the neural circuit in consideration explicitly processes visual-masking stimuli, and then the dichoptic mode can be used to probe the NCC.

Choosing an appropriate stimulus that is processed at multiple levels of the visual system is key to localizing awareness. However, one must take care to control for other potential experimental confounds. Lamme and colleagues used visual-masking stimuli to examine the NCC and concluded that stimulus-derived late responses (e.g., after-discharges) are due to feedback from higher areas (Lamme, Zipser, & Spekreijse, 2002) and that this feedback is critical to maintaining awareness (figure 81.8). But if the late responses are due to feedback and not to feedforward circuits, their timing should be stable with respect to stimulus duration. That is, if late responses are due to feedback, target duration should not affect their latency, because the feedback would be driven by the target’s onset-response as it rises through the visual hierarchy (figure 81.9f). On the contrary, if the late responses are caused by the target’s termination in a feedforward fashion, then target duration would critically affect their latency (figure 81.9g). Figure 81.10 shows that, as the stimulus duration increases, so does the latency of the afterdischarge, against the predictions from Lamme’s feedback model (Macknik & Martinez-Conde, 2004b). The model is further ruled out on psychophysical grounds, as the perceptual strength of masking varies with target duration (Macknik & Livingstone, 1998).

Despite these arguments, Lamme’s group has maintained that late responses are due to feedback. In a recent study
Figure 81.8 Alternative model of visual backward masking and awareness that require recurrent feedback. (A) Population response strength in awake monkey V1 to a figure-present (thick line) and ground (no figure-present, thin line) stimulus. (B) Responses to figure and ground conditions in which the figure was seen (left) and not seen (right). Notice that in the not-seen trials the late response does not differ for figure and ground conditions. (Reprinted from Super, Spekreijse, & Lamme, 2001.) (C) Model suggesting that the visibility-correlated late response in panels A and B is due to recurrent feedback from higher level cortices (activated by the feedforward onset response). (Reprinted from Lamme, 2003.)

Figure 81.9 Predicted temporal dynamics of neuronal responses (as a function of stimulus duration) if the afterdischarge is (A) due to recurrent feedback driven by the stimulus onset or (B) driven by the termination of the stimulus in a feedforward manner.
they surgically removed the entire extrastriate visual cortex of a monkey (V3, V3A, V4, MT, MST, DP, LOP, LIPd, and 7a), a procedure which led to a reduction of area V1 late responses (Super & Lamme, 2007). However, surgical ablations are irreversible by definition: one cannot reverse the procedure to show that reinstating the ablated tissue cancels out the effect. Moreover, the surgical removal of the extrastriate cortex involves the resection of a large portion of the cerebral cortex, thus causing massive traumatic brain damage, including substantial damage to the cortical vascular systems as well as fibers of passage and nearby neural structures such as the optic radiations. Therefore it is unclear exactly what processes may or may not have been affected by such a drastic ablation.

Conclusions

We have reviewed the literature on the anatomy and physiology of feedback in the visual system and concluded that feedback connections may be the source of attentional facilitation and suppression, and that other proposed roles for feedback are not as clearly supported. We have also proposed that the large ratio of feedback to feedforward connections does not necessarily indicate a significant physiological role for feedback, but may instead be a requirement of any feedback pathway operating within a hierarchical neural system, such as the visual hierarchy. This statement would be true even if feedback subserves only a single role, such as top-down attentional modulation.

Finally, we have discussed the strengths of visual masking in the study of visual awareness, as compared to binocular rivalry, and have concluded that visual masking is a sound paradigm in awareness studies, whereas binocular rivalry has serious shortcomings as a tool to localize the NCC. Using visual masking as a tool, we have developed several new standards that must be met to determine the role of neural circuits, neurons, and brain areas in maintaining consciousness.

We have emphasized the need to control for the effects of attention as an important strategy in designing experiments that localize awareness. Attention can enhance or suppress the magnitude of neural responses to a given stimulus (Chen et al., 2008; Desimone & Duncan, 1995; McAdams & Maunsell, 1999; Moran & Desimone, 1985; Spitzer, Desimone, & Moran, 1988; Williford & Maunsell, 2006), and thus it may facilitate or suppress its perceptual awareness. However, attention is a distinct process from awareness itself (Koch & Tsuchiya, 2007; Merikle, 1980; Merikle & Joordens, 1997; Merikle, Smilek, & Eastwood, 2001). For instance, low-level bottom-up highly salient stimuli (such as flickering lights or loud noises) can lead to awareness and draw attention, even when the subject is actively attending to some other task, or not attending to anything (i.e., when the subject is asleep). It follows that experiments to isolate the NCC should control for the effects of attention.

Therefore, we add the following three standards for testing a neural circuit’s contribution to awareness to Parker and Newcombe’s list:

8. The candidate neurons should be tested with an illusion that allows one to dissociate the physical stimulus from its perception. If the candidate set of neurons is capable of
maintaining awareness, the neural responses should match the subjective percept, rather than the objective physical stimulus.

9. The candidate neurons must explicitly process the type of information or stimulus used to test them.

10. The responses of the neurons, and of the perceiving subject, should be measured with experimental controls for the effect of attention.

acknowledgments We thank Mona Stewart, Hector Rieiro, and Jorge Otero-Millan for their technical assistance. This study was funded by the Barrow Neurological Foundation, National Science Foundation, Arizona Biomedical Research Institute, and Science Foundation Arizona.

REFERENCES


