10

Cortical Dynamics of Attentive Object Recognition, Scene Understanding, and Decision Making

Stephen Grossberg

This chapter discusses recent neural models that clarify how cortical dynamics carry out various higher level processes, such as attentive object learning and recognition, scene understanding, and decision making. These models contribute to research areas in which George Sperling has made important, indeed often seminal and pioneering, contributions, as noted subsequently. They are part of a coordinated research program aimed at developing an increasingly comprehensive model of visual intelligence.

View-Invariant Object Learning and Recognition

One model proposes how spatial and object attention are coordinated by visual surface and boundary representations to control object category learning and recognition. These processes are spread over multiple stages of the What and Where cortical processing streams. This project clarifies the following basic issues: What is an object? How does the brain learn what an object is under both unsupervised and supervised learning conditions? In particular, how does the brain learn to bind multiple views of an object into a view-invariant categorical representation of a complex object while scanning its various parts with active eye movements?

To solve this problem, one needs to face the following dilemma squarely: Suppose that, as your eyes scan a scene, two successive eye movements focus on different parts of the same object part of the time, and on different objects the rest of the time. How does the brain avoid the problem of erroneously classifying views of different objects as belonging to a single object? One cannot say that the brain does this by knowing that some views belong together whereas others do not because this can happen even before the brain has a concept of what the

---

DOI: 10.1037/14135-011
Human Information Processing: Vision, Memory, and Attention, C. Chubb, B. A. Dosher, Z.-L. Lu, and R. M. Shiffrin (Editors)

object is. Indeed, such scanning eye movements may be used to learn the object concept in the first place.

In addition, we take for granted the fact that our eyes seem to scan a scene intelligently, as quantified by earlier investigators such as Yarbus (1961). Why do not our saccades jump haphazardly from one part of a scene to another? How does the brain know how to direct the eyes to explore an object's surface even before it has a concept of the object?

A cortical model called ARTSCAN (see Figure 10.1) is being developed to explain how the brain uses scanning saccadic eye movements to learn view-invariant object categories (Fazl, Grossberg, & Mingolla, 2009). ARTSCAN predicts how spatial and object attention work together to direct eye movements to explore object surfaces and to enable learning of view-invariant object categories from the multiple view categories that are thereby learned. Of course, the development of ingenious experiments and modeling of the spatial and temporal characteristics of attention is a major theme in George Sperling's work, such as Reeves and Sperling (1986; see also Grossberg & Stone, 1986), Shih and Sperling (2002), and Weichselgartner and Sperling (1987). The ARTSCAN model proposes, in addition, that interactions between spatial and object attention during visual *search* are part of the brain system for visual *object learning*.

In particular, ARTSCAN predicts that spatial attention uses an *attentional shroud,* or form-fitting distribution of spatial attention, that is derived through feedback interactions with an object's surface representation. Tyler and Kontsevich (1995) introduced the concept of an attentional shroud as an alternative to the perception of simultaneous transparency and provided psychophysical evidence that only one plane is seen at a time within the perceptual moment. This concept focused on object perception. ARTSCAN proposes that an attentional shroud also plays a fundamental role in regulating object learning.

Such a shroud is proposed to persist within the Where Stream during active scanning of an object with attentional shifts and eye movements. This claim raises the basic question: How can the shroud persist during active scanning of an object if the brain has not yet learned that there is an object there? ARTSCAN proposes how a *preattentively* formed surface representation leads to activation of a shroud even before the brain can recognize the surface as representing a particular object (see Figure 10.2). Such a shroud can be formed starting with either bottom-up or top-down signals. In the bottom-up route (see Figures 10.2a and 10.2b), a surface representation (e.g., in visual cortical area V4) directly activates a shroud that conforms its shape to that of the surface, in a spatial attention cortical area (e.g., posterior parietal cortex). The shroud, in turn, can topographically prime the surface representation via top-down feedback. A *surface-shroud resonance* can hereby develop.

In the top-down route, a volitionally controlled, local focus of spatial attention (an attentional spotlight) can send a top-down attentional signal to a surface representation. This spotlight of enhanced activation can then fill in across the entire surface, being contained only by the surface boundary (Grossberg & Mingolla, 1985a). Surface filling in generates a higher level of filled-in surface activation than did the bottom-up input to the surface alone. The filling in of such a top-down attentional spotlight can hereby have an effect on the total filled-in surface activity that is similar to that caused by a higher bottom-up

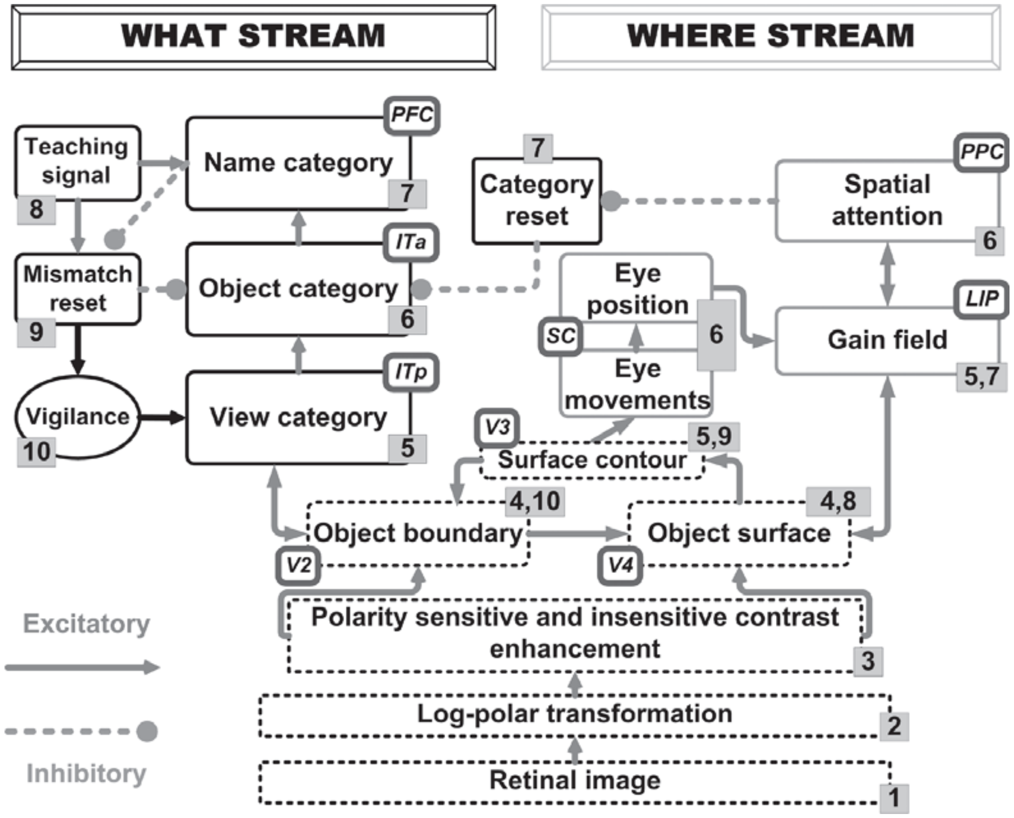

Figure 10.1. ARTSCAN model diagram. The Boundary and Surface processes have dashed borders and send input to both visual streams. The Where Stream modules have light gray borders, and the What Stream modules have black borders. The small white tabs with round edges next to each box represent the anatomical region in which the process is thought to occur. The numbers in the gray boxes next to each module box show the approximate order of first activation in that module after the retina receives an input. If there are two such numbers in a box, the second one represents the time that feedback reaches that module. Solid arrows represent excitatory connections, and dashed connections with a round head represent inhibitory ones. ITa = anterior part of inferotemporal cortex; ITp = posterior part of inferotemporal cortex; LIP = lateral intraparietal cortex; LGN = lateral geniculate nucleus; PFC = prefrontal cortex; SC = superior colliculus; V1 and V2 = primary and secondary visual areas; V3 and V4 = visual areas 3 and 4. See text for details. From "View-Invariant Object Category Learning, Recognition, and Search: How Spatial and Object Attention Are Coordinated Using Surface-Based Attentional Shrouds," by A. Fazl, S. Grossberg, and E. Mingolla, 2009, *Cognitive Psychology, 58,* p. 14. Copyright 2009 by Elsevier. Reprinted with permission.

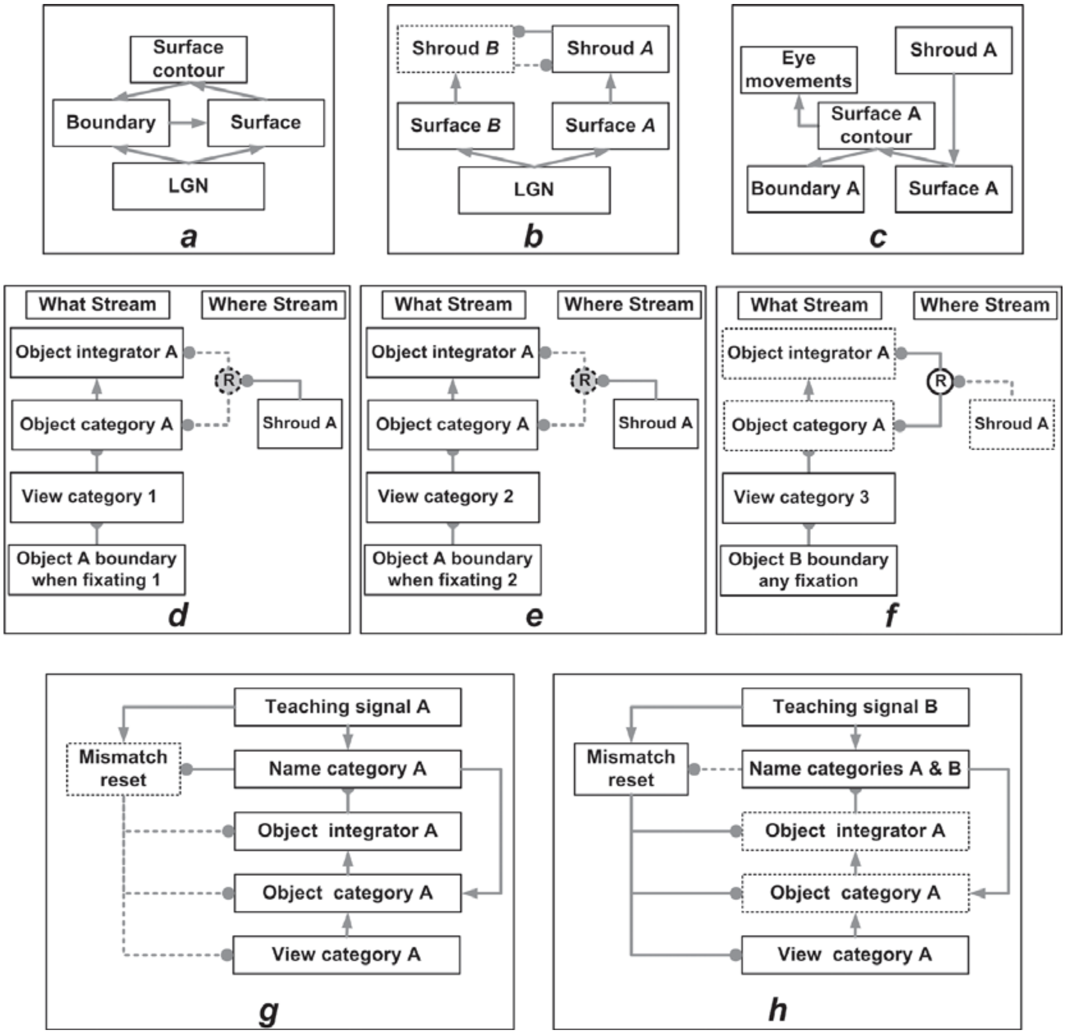

Figure 10.2. Schematic of ARTSCAN operations. (a–c): Where stream operations. (d–f): Unsupervised learning in ARTSCAN. (g–h): Supervised learning in ARTSCAN. (a) Preattentive boundary–surface interaction. The visual image represented in the lateral geniculate nucleus (LGN) input is processed by two cortical streams: boundaries and surfaces. Wherever there is a closed boundary on the boundary map, a surface will form as the result of gated diffusion on the surface map. These completed surfaces will in turn up-regulate their corresponding boundaries through feedback via surface contours. (b) Shroud formation. If there is more than one surface present, the competition between their representations on the spatial attention map results in a winner, called the attentional shroud. The coordinate transform between the retinotopic surface map and the head-centric spatial attention map in a gain field is not shown in these simplified diagrams. The attentional shroud enhances its corresponding surfaces through feedback. The circuit in panel c conveys this effect to surface contour and boundary maps. Surface contour feedback to the eye movement map increases the activity of all of the hotspots on an *attended* object, making them possible winners as the next saccade target. (d) The eyes fixate point 1 on object *A,* while the shroud has formed around that object. The feedback discussed in panels a through c has already down-regulated any other object boundary activities. This boundary activation excites View Category 1, object category neuron *A,* and its corresponding object integrator neuron. (e) If the eyes

stimulus contrast (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Reynolds & Desimone, 2003). The more highly active surface representation can reactivate the spatial attention region to define a surface-form-fitting spatial locus of spatial attention, that is, a shroud. Again, the shroud is defined by a surface-shroud resonance.

Any surface in a scene can potentially sustain an attentional shroud, and surface representations dynamically compete for spatial attention (Figure 10.2b). The winner of the competition at a given moment gains activity and becomes the shroud. It has been proposed that perceptual boundaries are amodal, or invisible, within the boundary processing stream through primary visual cortical area (V1) interblobs, secondary visual area (V2) pale stripes, and V4. Visible percepts have been predicted to be surface percepts within the surface-processing stream through V1 blobs, V2 thin stripes, and V4 (Grossberg, 1994; Grossberg & Mingolla, 1985a). Earlier modeling work has also predicted that "all conscious states are resonant states" (Grossberg, 1980, 1999b, p. 1). These two streams of modeling come together in the assertion that surface-shroud resonances support conscious percepts of visible visual features. This hypothesis clarifies such phenomena as the visual neglect that occurs when a parietal lesion is made.

As saccadic eye movements explore an object's surface, the surface-induced shroud modulates object learning in the What Stream by maintaining activity of an emerging view-invariant category representation while multiple view-specific representations are linked to it through learning (Figure 10.2d and 10.2e). Output from the shroud also helps to select the boundary and surface features to which eye movements will be directed (Figure 10.2c) via a surface contour process that is predicted to play a key role in three-dimensional figure–ground

---

(Continued.) move to fixation point 2 on the same object, the new object boundary map activity might activate a different view category neuron 2, but it will activate the same object category and integrator *A,* because the attentional shroud is still active around the object *A* and inhibits the category reset neuron shown as R. (f) If the attentional shroud collapses around object *A,* the eyes can look at a different object, and another view category neuron will become active. Collapse of the attentional shroud disinhibits the category reset neurons, which inhibits all neurons in the two object layers, so these view category neurons will no longer be associated with object category neuron *A.* (g) If ARTSCAN receives the name of the object it is visiting, for example, object *A,* by a teaching signal *A,* it will associate it with the active object category and integrator neurons at that time. The activated name category neuron also inhibits the mismatch reset neuron. (h) Incorrect recall of object *B*'s name. In the same scenario as in panel g, if the bottom-up input from object boundaries eventually excites name category *A,* but the teaching signal activates name *B,* both name category neurons *A* and *B* are activated, and, due to normalization by a shunting on-center off-surround network (see Figure 10.9), the activity of both will decrease to below a threshold such that none can inhibit the mismatch reset neuron anymore. This allows the teaching signal to activate the mismatch reset neuron and inhibit both object layers and stop learning. This also increases the vigilance parameters in the view category layer. From "View-Invariant Object Category Learning, Recognition, and Search: How Spatial and Object Attention Are Coordinated Using Surface-Based Attentional Shrouds," by A. Fazl, S. Grossberg, and E. Mingolla, 2009, *Cognitive Psychology, 58,* p. 14. Copyright 2009 by Elsevier. Reprinted with permission.

separation (Grossberg, 1994, 1997) and to be mediated via cortical area V3A (Nakamura & Colby, 2000, 2002).

The model postulates that an active shroud weakens through time due to self-inhibitory inputs at selected target locations ("inhibition of return"; Grossberg, 1978a, 1978b; Koch & Ullman, 1985), combined with chemical transmitters that habituate, or are depressed, in an activity-dependent way (Abbott, Sen, Varela, & Nelson, 1997; Francis, Grossberg, & Mingolla, 1994; Grossberg, 1968) and gate the signals that sustain the shroud. When an active shroud is weakened enough, it collapses and can no longer inhibit a reset signal (Figure 10.2f). When a reset signal is disinhibited, its tonic activity can transiently inhibit the active view-invariant object category in the What Stream, thereby preventing it from erroneously being linked to the view categories of subsequently foveated objects. Then a new shroud, corresponding to some other surface, is selected in the Where Stream, thereby inhibiting once again the reset signal, as a new object category is activated in the What Stream by the first view of the new object.

The model hereby predicts that a spatial attention shift causes a domain-independent transient reset signal that, in turn, causes a shift in categorization rules. Such a transient signal was recently reported in humans in medial superior parietal lobule by Chiu and Yantis (2009) using rapid event-related magnetic resonance imaging.

While a shroud remains active, several kinds of learning go on through time. The first kind is the bottom-up learning of a view-specific category, that is, a category that is activated by information about a specific view, or similar views, of the object (Figure 10.2d and 10.2e). The second kind of learning is top-down learning of an expectation that is activated by an active view-specific category. Such a learned expectation focuses attention on the pattern of *critical features* that are incorporated during learning into the prototype that is used to recognize variations of that particular object view. This type of top-down attention is *object attention* (Posner, 1980), and it is distinct from the *spatial attention* (Duncan, 1984) that the shroud embodies. Taken together, these bottom-up category learning and top-down expectation learning mechanisms have been predicted by adaptive resonance theory (ART) to enable the brain to rapidly learn, and stably remember, new object categories without experiencing catastrophic forgetting of previously learned object categories (Carpenter & Grossberg, 1993; Grossberg, 2003). Recent review articles summarize behavioral and neurobiological data that support all of the main ART predictions about how the brain does this, notably predictions about the role of top-down attention in regulating learning (Grossberg, 2003; Raizada & Grossberg, 2003).

These view-specific categories are, in turn, linked by associative learning to the emerging view-invariant object categories, the persistent activity of which is ensured by an active shroud (Figure 10.2d and 10.2e). In particular, even while a view-specific category is reset as the eyes move to foveate a new view of an object, the view-invariant category is protected from reset so that multiple view-specific categories of the same object can be associated with a single view-invariant object category. Surface-based spatial attention and object attention are hereby coordinated to learn view-invariant object categories and thereby to solve the view-to-object binding problem.

How does the brain move the eyes to scan different parts of an object before moving to scan other objects? Why do successive saccades not instead jump haphazardly around a scene? The ARTSCAN model traces this property to following basic design properties of the FACADE model of three-dimensional (3D) vision and figure–ground separation and its realization by laminar visual cortical circuits in the 3D LAMINART model (e.g., Cao & Grossberg, 2005; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000). In these models, perceptual boundaries both initiate and block the filling in of surface lightness and color via *boundary-to-surface* signals, as shown in Figure 10.2a. Figure 10.2a also contains a *surface-to-boundary feedback* pathway, via a *surface contour* process (Figure 10.2c). This process was introduced in the FACADE model to ensure perceptual *consistency*: It explains how, even though boundaries and surfaces form according to *complementary* computational rules, they give rise to a *consistent* visual percept (Grossberg, 1994, 2003). Surface-to-boundary feedback ensures consistency by using the surface contours of successfully filled-in surfaces to confirm and strengthen the boundaries that triggered filling-in of these surfaces and to inhibit boundaries that do not, such as boundaries that do not surround a surface region. The FACADE model predicts that the surface-to-boundary feedback mechanism also plays a key role in 3D figure–ground separation.

This boundary–surface feedback loop is proposed to work as follows: 3D boundary signals are topographically projected from V2 Interstripes to V2 Thin Stripes. If a boundary is closed, it can contain the filling in of lightness and color within it. If the boundary has a sufficiently big gap, surface lightness and color can dissipate through the gap. The surface contour process is sensitive to the contrasts at the border of a successfully filled-in surface within a closed boundary. This contrast-sensitive process is realized by an on-center, off-surround network that detects the contours of successfully filled-in surfaces. This is a shunting on-center, off-surround network across position and within depth.

The surface contour outputs from successfully filled-in surfaces use topographic excitatory signals to strengthen the boundaries that generated these surfaces, and inhibitory signals to weaken spurious boundaries at the same positions but farther depths. This is an on-center, off-surround network within position and across depth. This surface-to-boundary feedback is predicted to arise from V2 Thin Stripes and terminate in V2 Pale Stripes.

By eliminating these spurious boundaries, surface-to-boundary feedback enables occluding surfaces and partially occluded surfaces to be separated onto different depth planes and allows partially occluded boundaries and surfaces to be amodally completed behind their occluders. Such completed representations can then be more easily recognized in the inferotemporal cortex and beyond. This sort of interaction can, for example, explain the *proximity luminance covariation* that has been described by Dosher, Sperling, and Wurst (1986). In summary, the FACADE model predicts why and how contour-sensitive surface-to-boundary feedback helps to define an object by ensuring that the correct object boundaries and surfaces are consistently bound together to form a preattentive object representation.

ARTSCAN predicts that the *same process* that preattentively defines and segregates objects in depth also regulates attentive learning of an object category. It does this by inducing sequences of scanning saccadic eye movements on an object surface, the spatial attentional shroud of which is active at any given time. Here is how that is proposed to happen.

First, the lateral geniculate nucleus (LGN) generates preattentive bottom-up inputs in response to the surfaces in a scene (Figure 10.2a). The surfaces, in turn, attempt to topographically activate spatial attention to form surface-fitting attentional shrouds (Figure 10.2b). The spatial attention network contains long-range inhibitory interactions that tend to select the strongest shroud and inhibit weaker ones, other things being equal (Figure 10.2b). The winning shroud sends topographic feedback to its generative surface, thereby activating it further (Figure 10.2c). Thus, as noted earlier, ARTSCAN predicts that surface representations receive both contrastive bottom-up inputs from areas such as the LGN and top-down spatial attention inputs from areas like the parietal cortex. Recent data support the view that attention can, in fact, increase the perceived brightness of a surface (Carrasco et al., 2000; Reynolds & Desimone, 2003) and connections from parietal areas to prestriate visual cortex are known (Cavada & Goldman-Rakic, 1989, 1991; Distler, Boussaoud, Desimone, & Ungerleider, 1993; Webster, Bachevalier, & Ungerleider, 1994).

ARTSCAN predicts that this feedback plays an important role in object learning. In particular, when the winning surface has its activation enhanced by top-down spatial attention, its contrast relative to its surround increases. As a result, *its surface contour signals increase.* As summarized in Figure 10.2c, stronger surface contour signals generate stronger eye movement target commands to the saccadic eye movement system to direct scanning eye movements to the object surface whose shroud is active. ARTSCAN hereby predicts how the views that the eyes happen to look at tend to belong to the same object surface (Theeuwes, Mathot, & Kingstone, 2010) while its spatial attentional shroud is on, and these are the views that will be learned.

Cortical area V3A is one possible brain area where such surface contour signals may get converted into eye movement target signals. Studies show that V3A is concerned with relative disparity (Backus, Fleet, Parker, & Heeger, 2001), gaze (Galletti & Battaglini, 1989), saccades (Caplovitz & Tse, 2007; Nakamura & Colby, 2000, 2002), and prehensile hand movements (Nakamura et al., 2001). This prediction clarifies how the process which helps to define separable object representations, also helps to guide our eyes to look at these object representations.

The ARTSCAN model learns with 98.1% accuracy on a letter database with letters that vary in size, position, and orientation. It does this while achieving a compression factor of 430 in the number of its category representations, compared with what would be required to learn the database without the view-invariant categories. The model also simulates reaction times (RTs) in human data about object-based attention: RTs are faster when responding to the noncued end of an attended object compared with a location outside the object, and slower engagement of attention to a new object occurs if attention has to first be disengaged from another object first (Brown & Denney, 2007; Egly, Driver, & Rafal, 1994).

ARTSCAN also provides new insights into the functional role of the *cortical magnification factor,* namely, the transformation from retinal to cortical representation is space-variant (Daniel & Whitteridge, 1961; Drasdo, 1977; Schwartz, 1977), a fact that cannot be ignored when analyzing how the eyes explore the world. These intrinsic variations in the input to the cortex, created by the combination of eye movements and cortical magnification on the same stationary object, are often larger than the extrinsic variations due to rigid transformations of the object itself. The ARTSCAN learning processes described here enable object category neurons to tolerate both *extrinsic* rigid object transformations due to movements of an object relative to an observer and *intrinsic* variations due to saccades that explore different parts of the same object.

Volitional Control of Spatial Attention Permits Multiple Shrouds to Coexist: Homologs With Working Memory Storage and Visual Imagery and Fantasy

Spatial attention need not form a shroud around only one object. Spatial attention can form over more than one object (Downing, 1988; Eriksen & Yeh, 1985; LaBerge & Brown, 1989; McMains & Somers, 2005; Pylyshyn & Storm, 1988; Yantis, 1992). I predict that the inhibitory gain that determines the strength of inhibition across shrouds (Figure 10.2b) is under volitional control by the basal ganglia. Weaker inhibition allows more than one shroud to exist at a time. One possible target of such volitional control is the inhibitory interneurons in cortical Layer 4. Volitional control of the balance between cortical excitation and inhibition is predicted to be a general brain mechanism that expresses itself behaviorally in strikingly different ways.

For example, in the visual cortex, top-down expectations provide attentional modulation of bottom-up inputs. They do so via a top-down, *modulatory* on-center, off-surround network that has its effect on Layer 4 cells (see Grossberg, 1999a, 2003, for supportive experimental data). The modulatory on-center can sensitize target cells to respond more vigorously and synchronously to attended visual feature combinations. If increasing volitional gain inhibits the inhibitory interneurons, it can convert the modulatory on-center into one that can drive suprathreshold activation of its target cells via a top-down expectation. This volitional mechanism has been predicted to enable top-down expectations to generate suprathreshold conscious percepts of visual imagery and fantasy (Grossberg, 2000b). When phasic volitional control over visual imagery and fantasy is replaced by tonic hyperactivity, hallucinations can occur, as in schizophrenia.

The same sort of volitional control is predicted in ventrolateral prefrontal cortex to control when a sequence of items is stored in a cognitive working memory (Grossberg & Pearson, 2008). These predictions suggest how homologous mechanisms within laminar neocortical circuits can carry out different functions: allocation of spatial attention in the parietal cortex, visual imagery and fantasy in the visual cortex, and working memory storage in the prefrontal cortex.

Scene Understanding: Gist, Texture, and Multiple-Scale Shrouds

Attentional shrouds can be allocated on multiple spatial scales to enable scene understanding to occur. This clarifies how humans may rapidly recognize a scene by shifting spatial attention from global gist to more local properties of textures and objects. The ARTSCENE neural model (Grossberg & Huang, 2009) proposes how to learn and predict scene identity incrementally by learning a scene's gist as a large-scale texture category and how to improve scenic prediction by sequentially scrutinizing finer textures of a scene that are selected by attentional shrouds (Figure 10.3). Scene identity is predicted via a learned mapping from multiple-scale gist and texture category activations. Tested on the real scene images from Oliva and Torralba (2001), this gist-plus-texture system can distinguish natural scene categories up to 92% correct, outperforms alternative models in the literature that use biologically implausible computations, and outperforms systems that use either gist or texture information alone. It was sufficient to combine texture information from the three largest annotated regions in an image to achieve good scene recognition. These texture measures correlate highly with scene identity, and the correlation strength is proportional to the texture size. On average, in the scenic database that was used to benchmark the model, three principal textures together constitute 92.7% of the total area of a landscape image in the data set and appear much more salient than small objects and textures.

The ARTSCENE model is consistent with several studies in global-to-local visual processing (e.g., Navon, 1977; Schyns & Oliva, 1994) and with the fact that human viewers can detect a named object in a scene within approximately 150 ms, that is, less than the average fixation time (~300 ms; Potter, 1975). The ARTSCENE computation of texture builds on circuits that combine multiple stages of oriented filtering, rectification, and lateral inhibition that were introduced in Grossberg and Mingolla (1985b) and then adapted by other authors (e.g., Chubb, Sperling, & Solomon, 1989; Solomon, Sperling, & Chubb, 1993; Sutter, Beck, & Graham, 1989).

Motion-Defined Objects, Formotion Interactions, and Eye Movement Control

Both the ARTSCAN and ARTSCENE models involve What cortical processing stream mechanisms of object category learning and recognition, and how they interact with Where cortical processing stream mechanisms of spatial attention and action. Both models consider the processing of object forms, which may be stationary or may move and thereby change their visible views with respect to an observer. However, neither model analyzes the related problems of motion-defined objects, how form and motion information interact to generate object percepts, and how the eye movements that enable tracking of moving objects to occur are controlled. Neural models of these processes have been developed with various colleagues over the years, and integrating of these remains a goal of current research.

Models of how smooth pursuit and saccadic eye movements are controlled have become ever more detailed in the characterization of their underlying brain processes and their explanatory and predictive range (e.g., Gancarz & Grossberg, 1998; Grossberg & Kuperstein, 1986, 1989; Grossberg, Roberts, Aguilar, & Bullock, 1997; Pack, Grossberg, & Mingolla, 2001; Srihasam, Bullock, & Grossberg, 2009).

The same is true about how motion-defined objects are perceived and how form and motion interact to generate object percepts and thereby to overcome

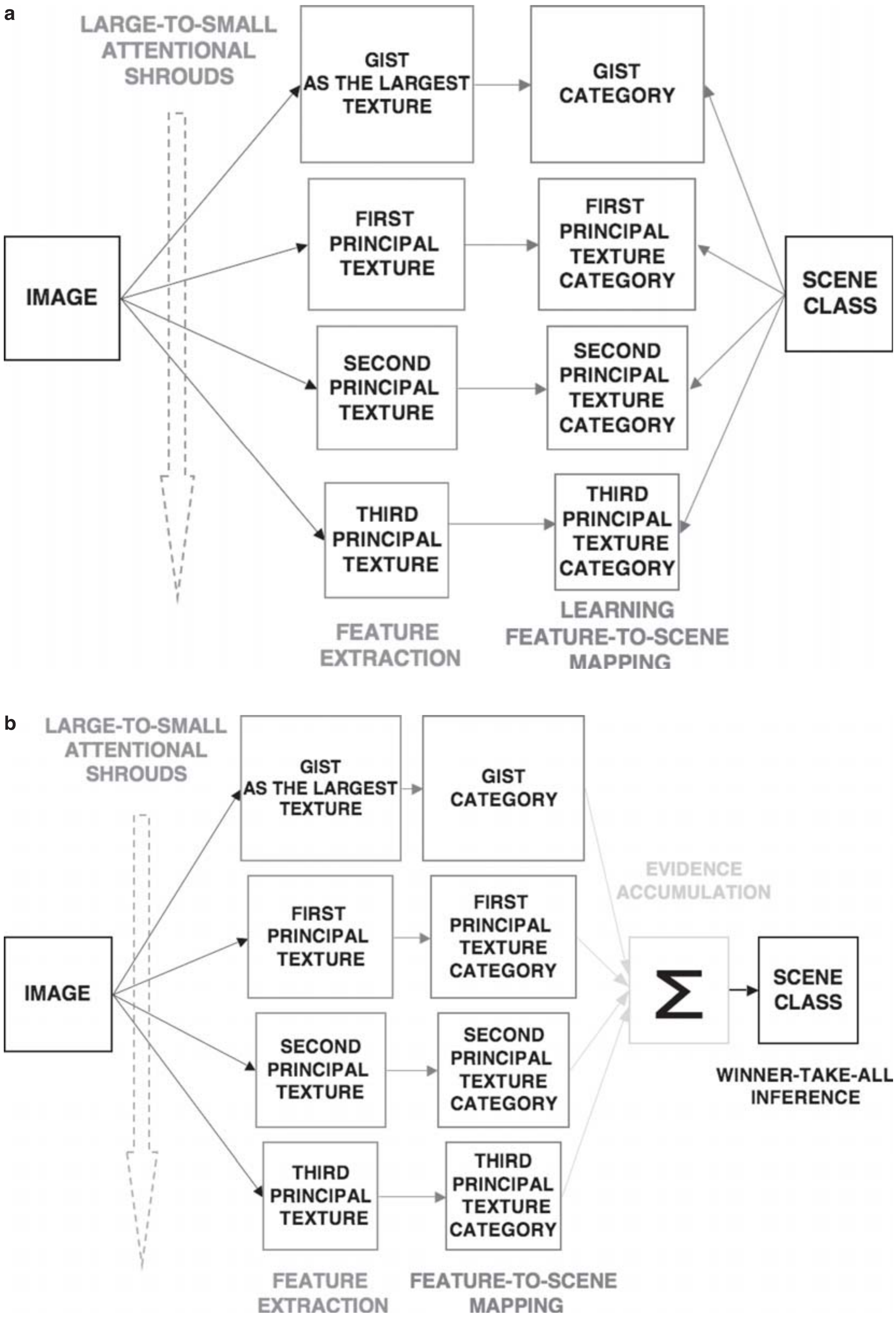

Figure 10.3. ARTSCENE model. (a) ARTSCENE training mode. (b) ARTSCENE testing mode. From “ARTSCENE: A Neural System for Natural Scene Classification,” by S. Grossberg & T.-R. Huang, 2009, *Journal of Vision, 9,* p. 4. doi:10.1167/9.4.6 Copyright 2009 by Association for Research in Vision and Ophthalmology. Reprinted with permission.

the complementary deficiencies (Grossberg, 2000a) of the What and Where cortical processing streams, taken separately, in computing effective representations to define and track moving-objects-in-depth. To understand such processes, one needs to answer the following question: How do visual form and motion processes cooperate to compute object motion when each process separately is insufficient? Consider, for example, a deer moving behind a bush. Here the partially occluded fragments of motion signals available to an observer must be coherently grouped into the motion of a single object. A 3D FORMOTION model has been developed to answer this question (Baloch & Grossberg, 1997; Baloch, Grossberg, Mingolla, & Nogueira, 1999; Berzhanskaya, Grossberg, & Mingolla, 2007; Chey, Grossberg, & Mingolla, 1997, 1998; Grossberg, Mingolla, & Viswanathan, 2001; Grossberg & Rudd, 1989, 1992) and thereby to propose a solution of the global *aperture problem.* Wallach (1935) first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: no matter what the real direction of motion may be, the perceived direction is perpendicular to the orientation of the line, that is, the *normal component* of motion. The aperture problem is faced by any localized neural motion sensor, such as a neuron in the early visual pathway, which responds to a local contour moving through an aperture-like receptive field. In contrast, a moving dot, line end, or corner provides unambiguous information about an object's true motion direction (Shimojo, Silverman, & Nakayama, 1989). The 3D FORMOTION model proposes how such moving visual features activate cells in the brain that compute *feature-tracking* signals which can disambiguate an object's true direction of motion.

In its current form, the 3D FORMOTION model comprises five important functional interactions involving the brain's form and motion systems that address such situations. In the first, 3D boundary representations, in which figures are separated from their backgrounds, are formed in cortical area V2, just as in Figure 10.2a. These depth-selective V2 boundaries select motion signals at the appropriate depths in middle temporal cortical area (MT) via V2-to-MT signals. In the second, motion signals in MT disambiguate locally incomplete or ambiguous boundary signals in V2 via MT-to-V1-to-V2 feedback. The third functional property concerns resolution of the aperture problem along straight moving contours by propagating the influence of unambiguous motion signals generated at contour terminators or corners. Here is where sparse feature-tracking signals, for example, from line ends, are amplified to overwhelm numerically superior ambiguous motion signals along line segment interiors. In the fourth, a spatially anisotropic motion grouping process takes place across perceptual space via feedback between MT and medial superior temporal cortical area (MST) to integrate veridical feature tracking and ambiguous motion signals to determine a global object motion percept. The fifth property can also use the MT–MST feedback loop to convey an attentional priming signal from higher brain areas back to V1 and V2. Baloch et al. (1999) have proposed how such form-motion, or *formotion,* interactions provide a neurobiological account of the type of *third-order motion* that George Sperling and his colleagues (e.g., Lu & Sperling, 1995a, 1995b) have so brilliantly articulated.

The 3D FORMOTION model has been used to simulate a wide range of perceptual data about motion perception, most of them without any other explanation at the present time. Rather than review these explanations, I would like to illustrate how a model can explain many data other than the ones that were used to discover its underlying design principles and brain mechanisms.

Temporal Dynamics of Decision Making During Motion Perception in the Visual Cortex

One such database concerns perceptual decision making, as exhibited by saccadic eye movements, during visual motion perception. Speed and accuracy of perceptual decisions are known from many types of experiments to covary with certainty in the input (Gold & Shadlen, 2007; Luce, 1986) and have been modeled by using statistical models. Recently, neurophysiologic data have been collected that correlate with the rate of evidence accumulation in parietal and frontal cortical “decision neurons.” Such data challenge us to go beyond statistical models to discover how the brain accomplishes such decision making. A biophysically realistic model of interactions within and between retina/LGN and cortical areas V1, MT, MST, and lateral intraparietal (LIP), gated by basal ganglia, has been developed to quantitatively simulate dynamic properties of decision making in response to ambiguous visual motion stimuli (Grossberg & Pilly, 2008). This Motion Decision, or MODE, model clarifies how the brain circuits that were characterized in the 3D FORMOTION model to solve the aperture problem can also contribute to making probabilistic decisions in real time.

Some scientists have claimed that perception and decision-making can be described using Bayesian inference, or related general statistical ideas, that estimate the optimal interpretation of the stimulus given priors and likelihoods. However, such concepts do not disclose the neocortical mechanisms that enable perception and make decisions. The MODE model explains behavioral and neurophysiological decision-making data without an appeal to Bayesian concepts and, unlike other existing models of these data, generates perceptual representations and choice dynamics in response to the moving experimental visual stimuli. Quantitative model simulations include the time course of LIP neuronal dynamics, as well as behavioral accuracy and reaction time properties, during both correct and error trials at different levels of input ambiguity in both fixed duration (FD) and RT tasks. Model MT–MST interactions compute the global direction of random dot motion stimuli, whereas model LIP computes the stochastic perceptual decision that leads to a saccadic eye movement.

A valuable paradigm for studying decision making, which links psychophysics and neurophysiology, has been developed by Newsome, Shadlen, and colleagues (Roitman & Shadlen, 2002; Shadlen & Newsome, 2001). This research studies how brain dynamics in the LIP area relate to saccadic behavior of monkeys (percentage accuracy, RT) that are based on discriminating the motion direction of random dot motion stimuli at various degrees of coherence.

In these experiments, two kinds of tasks were used: FD and RT tasks. Macaques were trained to discriminate net motion direction and report it via a saccade. Random dot motion displays, covering a 5° diameter aperture centered at the fixation point on a computer monitor, were used to control motion coherence—namely, the fraction of dots moving nonrandomly in a particular direction from one frame to the next in each of the three interleaved sequences. Varying the motion coherence provided a quantitative way to manipulate the ambiguity of directional information that a monkey could use to make a saccadic eye movement to a peripheral choice target in the judged motion direction, and thus the task difficulty. More coherence resulted in better accuracy and faster responses.

In the FD task (Roitman & Shadlen, 2002; Shadlen & Newsome, 2001), monkeys viewed the moving dots for a fixed duration of 1 s and then made a saccade to the target in the judged direction after a variable delay. In the RT task (Roitman & Shadlen, 2002), monkeys had theoretically unlimited viewing time and were trained to report their decision as soon as the motion direction was determined. The RT task allowed measurement of how long it took the monkey to make a decision, which was defined as the time from the onset of the motion until when the monkey initiated a saccade. The two monkeys in the Roitman & Shadlen (2002, p. 9476) experiment were shaped during RT task training to initiate the choice saccade within "approximately 1 s" after the dots turn on. In each RT task trial, the monkeys had to wait for a minimum of about 1 s (one monkey: 800 ms; the other: 1,200 ms) after motion onset to receive a reward, however rapidly they responded. Human subjects in a similar RT task usually respond around 1 s from motion onset for the weakest coherence without any speed instruction (Palmer, Huk, & Shadlen, 2005, p. 385).

Neurophysiological recordings were done in LIP while the monkeys performed these tasks. The recorded neurons had receptive fields that encompassed just one target and did not include the circular aperture in which the moving dots were displayed. Additionally, they were among those that showed sustained activity during the delay period in a memory-guided saccade task. It was found that the speed and accuracy of perceptual decisions covaried strongly with the rate of evidence accumulation in LIP cells.

Figure 10.4 summarizes the MODE model processing stages. All the stages through MSTv are part of the 3D FORMOTION model. The new processes include saccadic target selection in cortical area LIP and gating of LIP cell responses by the basal ganglia. As noted earlier, the 3D FORMOTION model proposes a solution of the aperture problem. The LIP processing stage embodies a solution of the *noise-saturation problem* (Grossberg, 1973, 1980), which is faced by all neurons because their activations fluctuate within a small interval of possible values: How does a network of cells, or cell populations, remain sensitive to the spatially distributed pattern of their inputs as they vary greatly in size through time? Cells that interact in shunting recurrent on-center, off-surround networks are capable of solving this problem. A special case of such networks shows how the most highly activated cell, or cell population, is selected to make a decision.

In the Roitman and Shadlen (2002) and Shadlen and Newsome (2001) experiments, a fraction of the dots move randomly through time and another fraction move in a fixed direction. Local directional processes can be fooled when there are multiple dots in each frame, some dots move incoherently, or independent random dot motion sequences are interleaved through time. The directional transient cells (Figure 10.4) generate local directional signals between any two dots that occur with an appropriate spatiotemporal displacement.

Each processing stage in Figure 10.4 has been used to explain neurophysiological and anatomical data in the articles that introduced the 3D FORMOTION model. Here one comparison worth noting is between the elaborated Reichardt (1961) detectors that van Santen and Sperling (1985) put to good use in explaining various motion data, and the directional transient cells used in the 3D FORMOTION model. In the 3D FORMOTION model, directional transient cells build on the motion-veto directional circuit that was proposed by Barlow and Levick (1965). Grossberg, Mingolla, and Viswanathan (2001) elaborated that model to propose how directional selectivity could be ensured in response to targets moving with variable speeds. This was accomplished by a circuit in which nondirectional transient cells send signals to inhibitory directional interneurons and directional transient cells, and the inhibitory interneurons interact with each other and with the directional transient cells. This predicted interaction is consistent with recent retinal data concerning how bipolar cells interact with inhibitory starburst amacrine cells and direction-selective ganglion cells, and how starburst cells interact with each other and with ganglion cells (Fried, Münch, & Werblin, 2002). The possible role of starburst cell inhibitory interneurons in ensuring directional selectivity at a wide range of target speeds has not yet been tested.

The directional short-range filter (Figure 10.4) integrates directional evidence from any active directional transient cells that occur within its directionally selective receptive field. Typically some directions will be amplified more than others by the short-range filter. However, lower motion coherence, higher dot density, and more interleaving of stimulus frames in the definition of the random dot stimuli increase the probability that incorrect directional signals will be generated in the short-range filter and thereby reduce the impact of correct local groupings in determining a clear motion directional percept. Apart from being few in number, these correct directional signals also have a short life span because a new set of signal dots are chosen every frame. Therefore, the motion stream must somehow enable a relatively sparse set of short-lived and correct feature tracking signals to gradually discount the more numerous incorrect local directional groupings. This problem is in many ways similar to the aperture problem, even though individual moving dots do not cause an aperture problem.

Spatial and opponent directional competition selectively strengthen feature-tracking signals, weaken ambiguous motion signals, and create speed-sensitive receptive fields (Figure 10.4). These cells are predicted to be a Where cortical stream analog of the simple cells for form perception that are well known to occur in the What cortical stream. A directional long-range filter (Figure 10.4) then gives rise to true directional cells by pooling output signals with the same, or nearly the same, directional preference over multiple orientations, opposite contrast polarities, both eyes, and a larger spatial scale. The cells that receive inputs from the directional long-range filter are also predicted to be depth-selective. They are predicted to occur in cortical area MT and to be analogous to complex cells in the form pathway. The properties of sensitivity to speed, as well as pooling across multiple orientations, opposite contrast polarities, both eyes, and a range of depths, are not needed in our current simulations of direction discrimination.

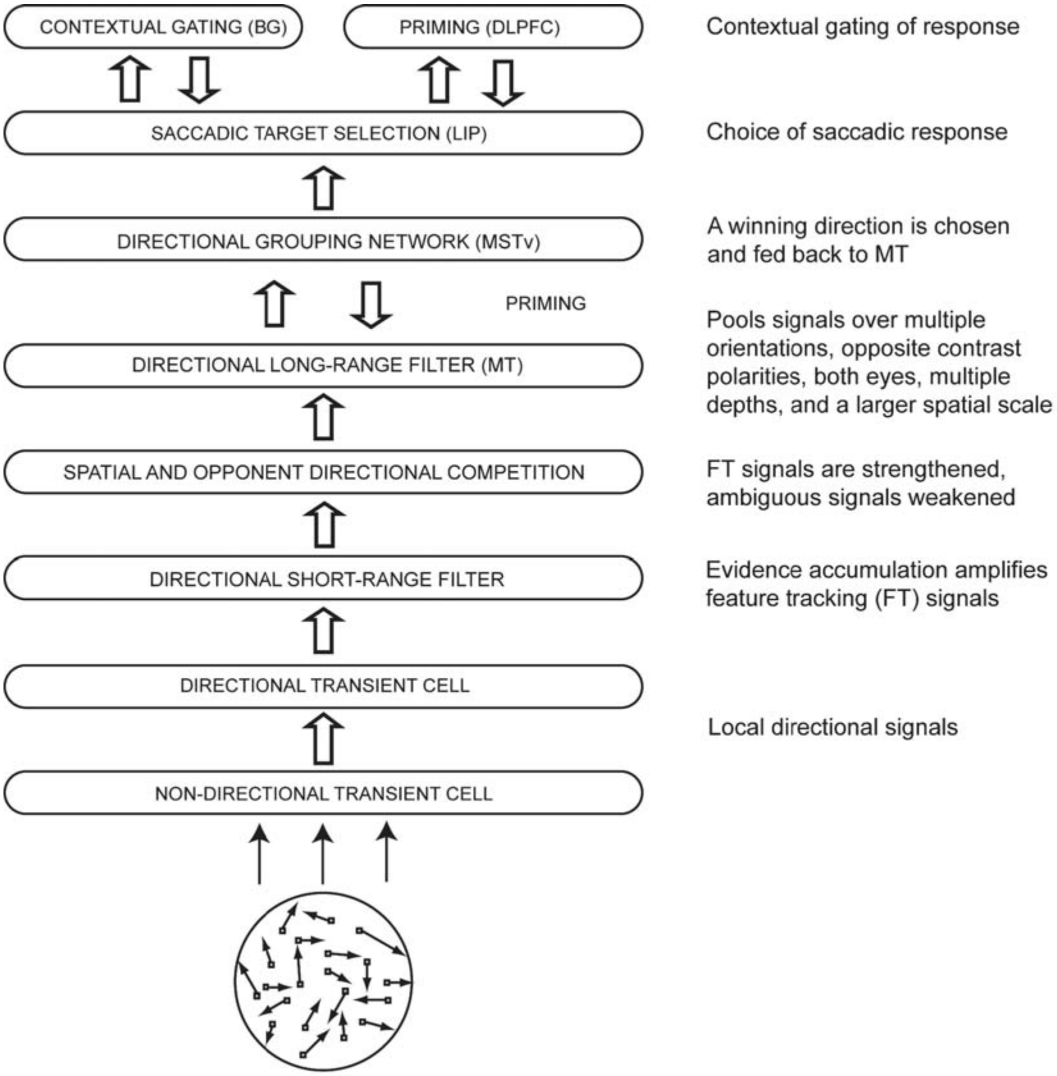

Figure 10.4. Retina/LGN–V1–MT–MST–LIP–BG model processing stages. The first two model stages (change-sensitive receptors and directional transient cells) in Retina/LGN compute local directional signals in response to image luminance increments or decrements. These signals feed into the directional short-range filter in V1, which accumulates evidence for motion in a given direction and thereby amplifies the feature tracking (FT) signals. Spatial and opponent directional competitions selectively strengthen FT signals, weaken ambiguous motion signals, and create speed-sensitive receptive fields. A long-range filter gives rise to true directional cells by pooling output signals with the same directional preference that survive the competitive stages. Within each direction, these signals are pooled over multiple orientations, opposite contrast polarities, both eyes, multiple depths, and a larger spatial scale. Pooling across multiple depths, both eyes, and opposite contrast polarities is not needed in the current simulations. These directional cells activate a directional grouping network, proposed to occur within cortical area MST, within which directions compete to determine a local winner. Enhanced feature tracking signals typically win this competition over ambiguous motion signals. Model MST cells that encode the winning direction feed back to model MT cells via a top-down spatial filter, where they indirectly boost directionally consistent cell activities by suppressing inconsistent activities over the spatial region to which they project. This

Figure 10.5 schematizes the LIP shunting recurrent on-center, off-surround network, also called a *recurrent competitive field,* that converts a spatially distributed, representation of motion into a directional saccadic decision, or command. Variants of this model, since its origin in Grossberg (1968) and Sperling and Sondhi (1968), as well as its mathematical analysis in Grossberg (1973), have been used by later authors to simulate the dynamics of perceptual or motor decisions in both deterministic models (e.g., Brown, Bullock, & Grossberg, 2004; Chey et al., 1997; Francis & Grossberg, 1996a, 1996b; Francis, Grossberg, & Mingolla, 1994) and stochastic models (Boardman, Grossberg, Myers, & Cohen, 1999; Cisek, 2006; Grossberg, Boardman, & Cohen, 1997; Usher & McClelland, 2001). Even a deterministic recurrent competitive field typically describes the temporal evolution of population mean activities and cell firing frequencies. It is only when properties that depend on the variance of cell firing become rate-limiting that explicit noise terms add explanatory power. Figures 10.6 through 10.8 illustrate the model's ability to simulate these data about decision making in response to motion stimuli, using the motion stimuli themselves as model inputs.

Concluding Remarks: Toward Solving the Mind–Body Problem

The several model examples that have been reviewed here are a small sample of those now being developed (see http://cns.bu.edu/~steve) for quantitatively explaining and predicting how brain mechanisms give rise to mental functions. These models embody new design principles that have been discovered by trying to explain psychological data as the result of an individual adapting autonomously to a changing world moment-by-moment in real time. The models have provided an ever-improving horizontal and a vertical understanding of behavioral and brain organization and dynamics: *horizontal* by showing how

---

(Continued.) feedback accomplishes directional and depthful motion capture within that region, whereas ambiguous or incoherent motion directions are suppressed. This shift in the spatial locus of unambiguous feature tracking signals continues to propagate across space as the MT-to-MST feedback process cycles through time. It is the action of this feedback loop that is predicted to solve the aperture problem and to generate a representation of global object direction and speed (Grossberg et al., 2001). This feedback loop between MT and MST is what regulates the temporal dynamics of decision making in the neurophysiological and behavioral data. The feedback loop needs more time to capture incoherent motion signals when there are more of them and cannot achieve as high a level of asymptotic response magnitude when more of them compete with the emerging winning direction. The motion information from MST is passed onto LIP, which converts it into an eye movement command, whereby the monkey reports its decision via a saccade. BG = basal ganglia; LIP = lateral intraparietal cortex; LGN = lateral geniculate nucleus; MST = medial superior temporal cortical area; MT = middle temporal cortical area; PFC = prefrontal cortex; SC = superior colliculus; V1 = primary visual area; V4 = visual area 4. From "Temporal Dynamics of Decision-Making During Motion Perception in the Visual Cortex," by S. Grossberg & P. Pilly, 2008, *Vision Research, 48,* p. 1347. Copyright 2008 by Elsevier. Reprinted with permission.

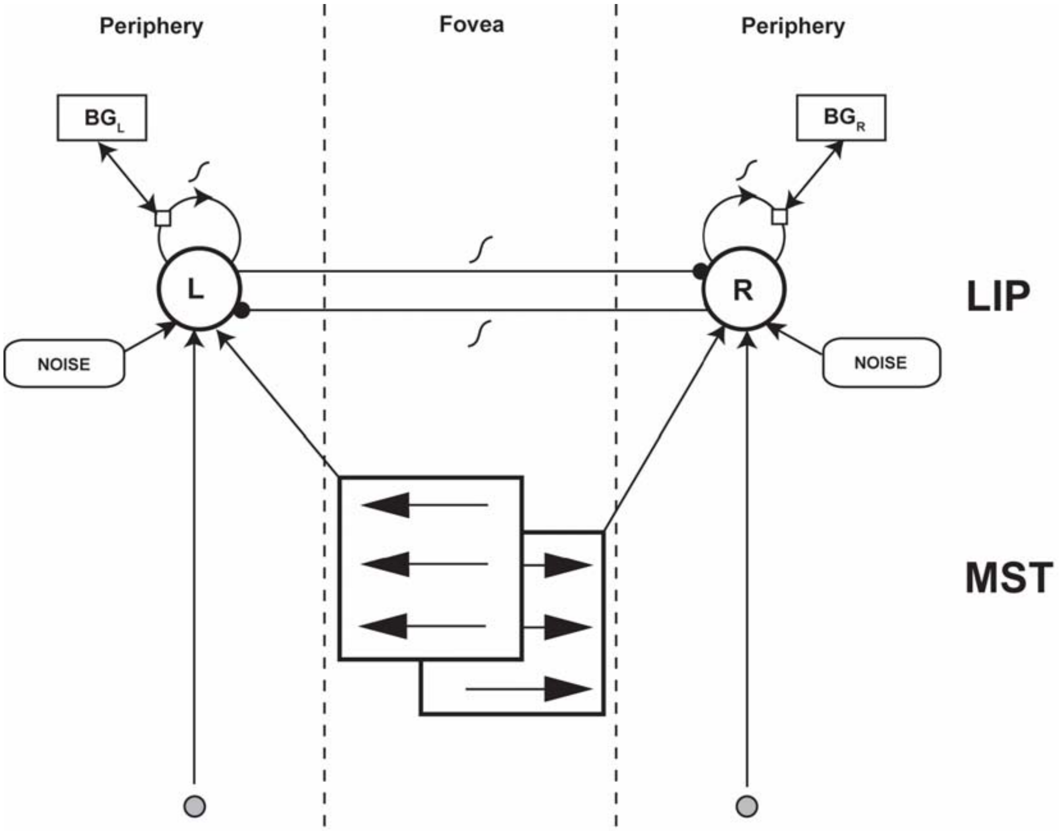

Figure 10.5. Lateral intraparietal cortex (LIP) decision circuit. The activities of cells in the recurrent competitive field predict the direction of the randomly moving dots. Each cell receives two bottom-up inputs: one from the form stream due to the presence of the peripheral choice target (indicated by a small circle) in the response field, and the other from the foveal medial superior temporal cortical area (MST) pool of the motion stream tuned to its preferred direction. The latter sensory input comes from outside the response field. This nonclassical neural connection is thought to result from training on the stereotypical task. Also, each cell self-excites itself and is inhibited by other competing cells in the field via different sigmoidal signal functions, which are indicated on the respective connections. The gain of the recurrent self-excitation is regulated by the basal ganglia. Internal noise processes to each cell help to simulate the probabilistic nature of perceptual decisions. BG = basal ganglia. From "Temporal Dynamics of Decision-Making During Motion Perception in the Visual Cortex," by S. Grossberg & P. Pilly, 2008, *Vision Research, 48,* p. 1351. Copyright 2008 by Elsevier. Reprinted with permission.

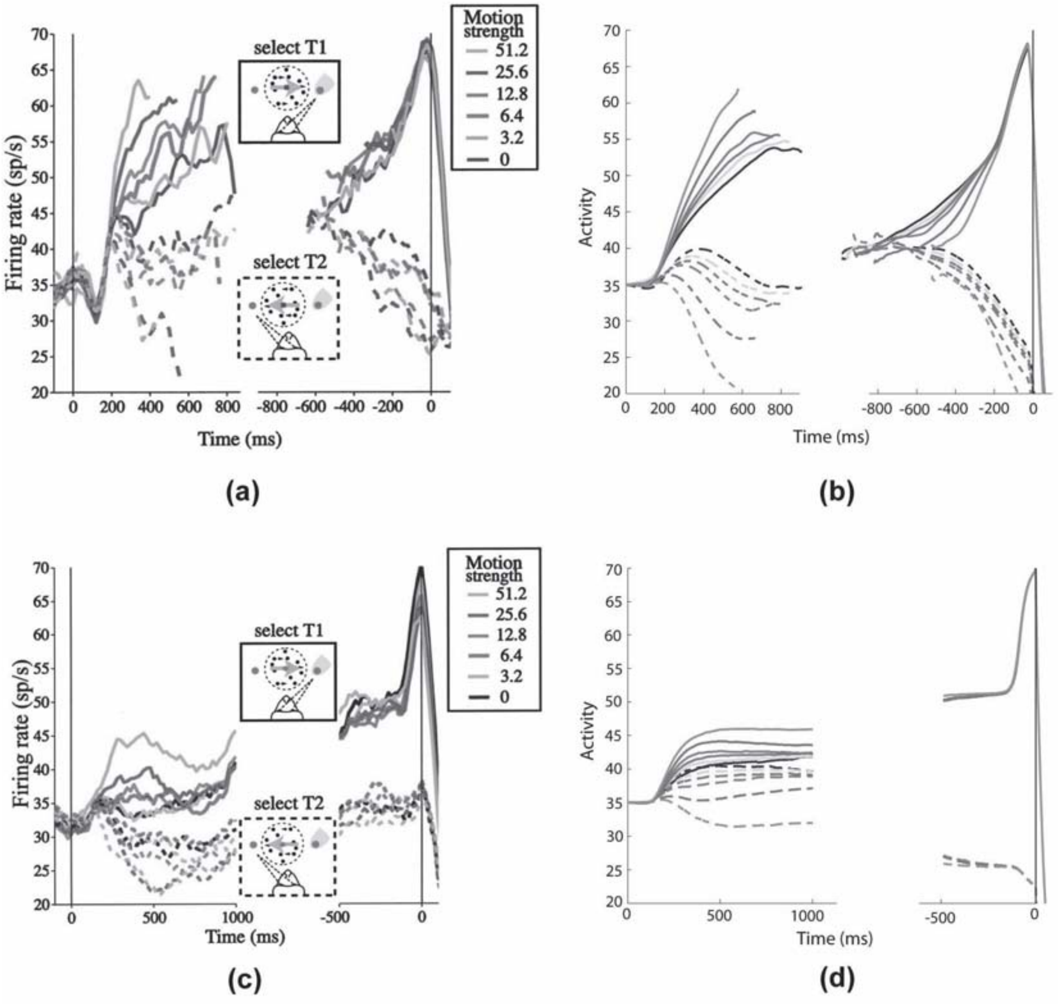

Figure 10.6. Temporal dynamics of lateral intraparietal cortex (LIP) neuronal responses during the fixed duration (FD) and reaction time (RT) tasks. (a) Average responses of a population of 54 LIP neurons among correct trials during the RT task (Roitman & Shadlen, 2002). The left part of the plot is time-aligned to the motion onset, includes activity only up to the median RT, and excludes any activity within 100 ms backward from saccade initiation (which roughly corresponds to presaccadic enhancement in firing). The right part of the plot is time-aligned to the saccade initiation and excludes any activity within 200 ms forward from motion onset (which corresponds to initial transient pause in firing). (b) Model simulations replicate LIP cell recordings during the RT task. In both data and simulations for the RT task, the average responses were smoothed with a 60-ms running mean. (c) Average responses of a population of 38 LIP neurons among correct trials during the 2002 FD task (Roitman & Shadlen, 2002) during both the motion viewing period (1 s) and a part (0.5 s) of the delay period before the saccade is made. (d) Model simulations mimic LIP cell recordings during the 2002 FD task.

(continued on next page)

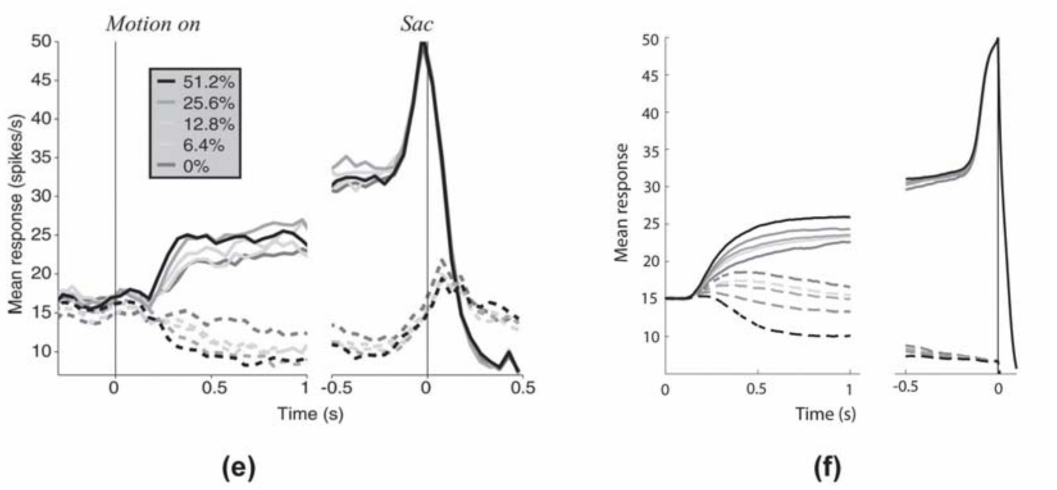

Figure 10.6. *(Continued)*. (e) Average responses of a population of 104 LIP neurons among correct trials during the 2001 FD task (Shadlen & Newsome, 2001), during both the motion viewing period (1 s) and a part (0.5 s) of the delay period before the saccade is made. (f) Model simulations emulate LIP cell recordings during the 2001 FD task. In panels a through f, solid and dashed curves correspond to trials in which the monkey correctly chose the right target (T_{in}) and the left target (T_{out}), respectively. Cell dynamics (rate of rise or decline in both tasks and response magnitude in FD task) reflect the incoming sensory ambiguity and the perceptual decision (solid: T_{in} choices, dashed: T_{out} choices). Note that for 0% coherence, even though there is no correct choice per se, the average LIP response rose or declined depending on whether the monkey chose T_{in} or T_{out}, respectively. From "Temporal Dynamics of Decision-Making During Motion Perception in the Visual Cortex," by S. Grossberg & P. Pilly, 2008, *Vision Research, 48,* p. 1352. Copyright 2008 by Elsevier. Reprinted with permission.

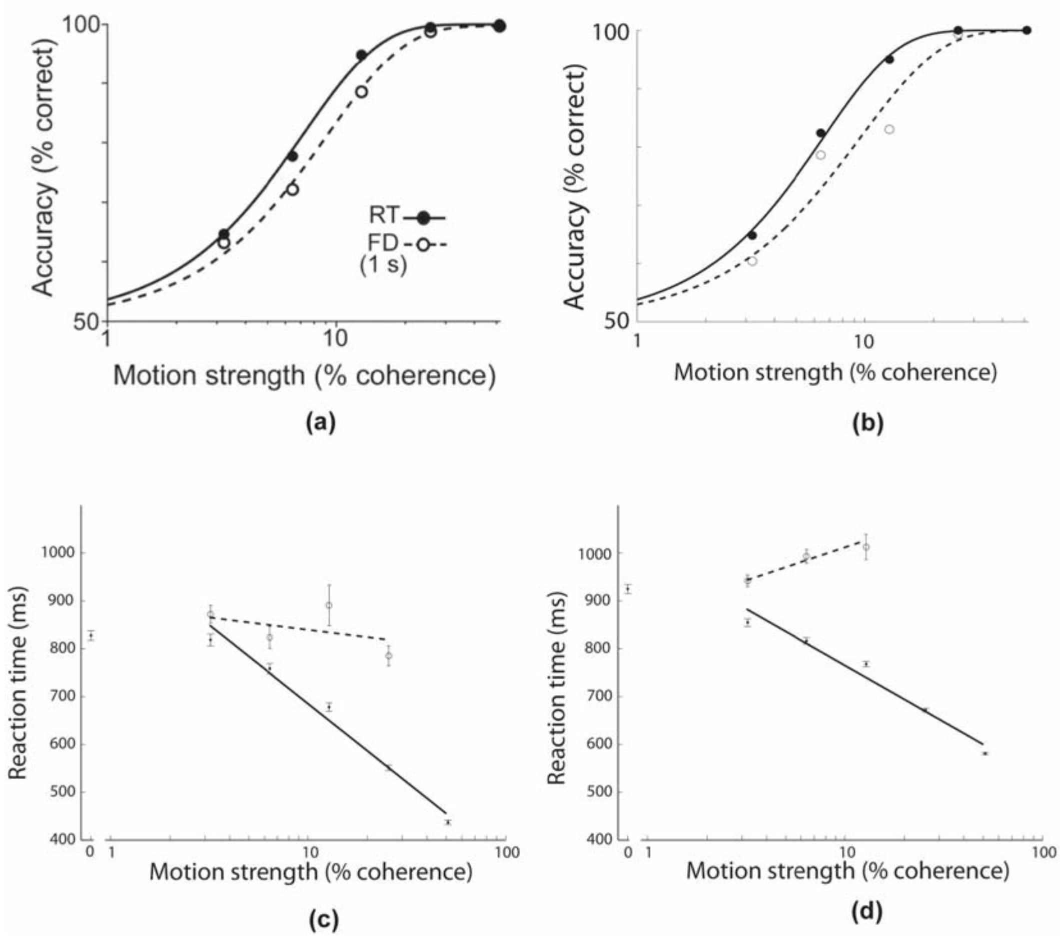

Figure 10.7. Psychometric and chronometric data during the fixed duration (FD) and reaction time (RT) tasks (Roitman & Shadlen, 2002). (a) Accuracy data (% correct) as a function of motion coherence (% certainty) is fit using a cumulative Weibull distribution function for both FD and RT tasks (Figure 6B in Mazurek et al., 2003). The ability to discriminate motion direction depends on the stimulus strength. Accuracy in the RT task is slightly better than that in the FD task for lower coherence levels. (b) Model simulations emulate these data. Solid curve corresponds to the RT task, and dashed curve to the 1-s FD task. Number of trials is 500. (c) Reaction time data (milliseconds) as a function of motion coherence (% certainty) is linear fit using a weighted, least squares estimate (as per the convention in Figure 3B of Roitman & Shadlen, 2002). The plot is prepared from the data for T_{out} (left target) choices in Table 2 of Roitman and Shadlen (2002). Data for T_{in} (right target) choices gives a similar plot. Less ambiguity implies a faster decision. Solid line corresponds to correct trials and dashed line to error trials. Error bars shown are standard errors of mean (SEM). SEMs decrease with coherence on correct trials but increase with coherence on error trials. Moreover, error trials have relatively higher SEMs. (d) Model simulations emulate the RT data on both correct and error trials. The behavior of SEMs with respect to coherence and correctness of trials is captured in the simulations. Number of trials is 500. In panels a through d, the abscissa is in the $\log_{10}$ scale. From "Temporal Dynamics of Decision-Making During Motion Perception in the Visual Cortex," by S. Grossberg & P. Pilly, 2008, *Vision Research, 48,* p. 1353. Copyright 2008 by Elsevier. Reprinted with permission.

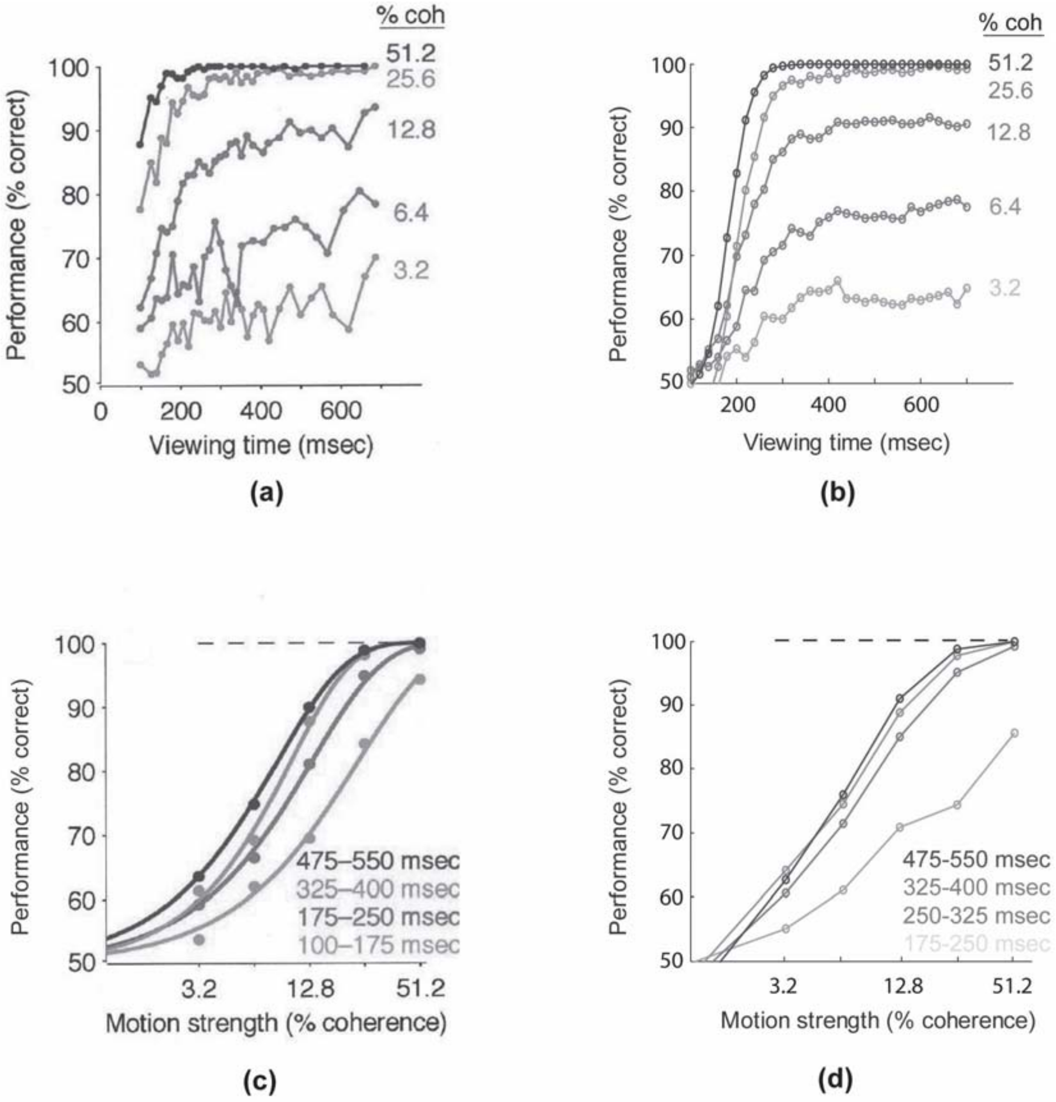

Figure 10.8. Influence of viewing duration on performance at various coherences in the fixed duration (FD) task paradigm. (a) Data from Gold and Shadlen (2003) show that the more time the dots are observed, the better the performance. This effect saturates at every coherence level. (b) Model simulations (2001 FD task) reproduce this influence of viewing time. (c) The psychometric function as a function of duration ranges. More viewing time tends to shift the psychometric function to the left, thus reducing the discrimination threshold. (d) The simulated psychometric functions capture these data trends. coh = coherence. From “Temporal Dynamics of Decision-Making During Motion Perception in the Visual Cortex,” by S. Grossberg & P. Pilly, 2008, *Vision Research, 48,* p. 1354. Copyright 2008 by Elsevier. Reprinted with permission.

multiple brain regions cooperate to overcome the complementary deficiencies of individual regions acting alone (Grossberg, 2000a); *vertical* by explaining how multiple levels of organization, including behavior, neuroanatomy, neurophysiology, biophysics, and biochemistry, work together to control adaptive behaviors (e.g., Fiala, Grossberg, & Bullock, 1996; Grossberg & Versace, 2008). Such models illustrate how the classical mind–body problem is finally getting solved during our time.

References

Abbott, L. F., Sen, K., Varela, J. A., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science, 275,* 221–224. doi:10.1126/science.275.5297.221

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86,* 2054–2068.

Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formotion dynamics. *Vision Research, 37,* 3037–3059. doi:10.1016/S0042-6989(97)00103-X

Baloch, A. A., Grossberg, S., Mingolla, E., & Nogueira, C. A. M. (1999). A neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 16,* 953–978. doi:10.1364/JOSAA.16.000953

Barlow, H. B., & Levick, W. R. (1965). The mechanism of directionally selective units in the rabbit's retina. *The Journal of Physiology, 178,* 477–504.

Berzhanskaya, J., Grossberg, S., & Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision, 20,* 337–395. doi:10.1163/156856807780919000

Boardman, I., Grossberg, S., Myers, C., & Cohen, M. (1999). Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception & Psychophysics, 61,* 1477–1500. doi:10.3758/BF03213112

Brown, J. M., & Denney, H. I. (2007). Shifting attention into and out of objects: Evaluating the processes underlying the object advantage. *Perception & Psychophysics, 69,* 606–618. doi:10.3758/BF03193918

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks, 17,* 471–510. doi:10.1016/j.neunet.2003.08.006

Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision, 18,* 515–578. doi:10.1163/156856805774406756

Caplovitz, G. P., & Tse, P. U. (2007). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cerebral Cortex, 17,* 1179–1189. doi:10.1093/cercor/bhl029

Carpenter, G. A., & Grossberg, S. (1993). Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences, 16,* 131–137. doi:10.1016/0166-2236(93)90118-6

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vision Research, 40,* 1203–1215. doi:10.1016/S0042-6989(00)00024-9

Cavada, C., & Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: II. Evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *The Journal of Comparative Neurology, 287,* 422–445. doi:10.1002/cne.902870403

Cavada, C., & Goldman-Rakic, P. S. (1991). Topographic segregation of corticostriatal projections from posterior parietal subdivisions in the macaque monkey. *Neuroscience, 42,* 683–696. doi:10.1016/0306-4522(91)90037-O

Chey, J., Grossberg, S., & Mingolla, E. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America A: Optics, Image Science, & Vision, 14,* 2570–2594. doi:10.1364/JOSAA.14.002570

Chey, J., Grossberg, S., & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research, 38,* 2769–2786. doi:10.1016/S0042-6989(97)00372-6

Chiu, Y.-C., & Yantis, S. (2009). A domain-independent source of cognitive control for task sets: Shifting spatial attention and switching categorization rules. *The Journal of Neuroscience, 29,* 3930–3938. doi:10.1523/JNEUROSCI.5737-08.2009

Chubb, C., Sperling, G., & Solomon, J. A. (1989). Texture interactions determine perceived contrast. *Proceedings of the National Academy of Sciences of the United States of America, 86,* 9631–9635. doi:10.1073/pnas.86.23.9631

Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *The Journal of Neuroscience, 26,* 9761–9770. doi:10.1523/JNEUROSCI.5605-05.2006

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology, 159,* 203–221.

Distler, C., Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in macaque monkeys. *The Journal of Comparative Neurology, 334,* 125–150. doi:10.1002/cne.903340111

Dosher, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research, 26,* 973–990. doi:10.1016/0042-6989(86)90154-9

Downing, C. J. (1988). Expectancy and visual-spatial attention: effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 188–202. doi:10.1037/0096-1523.14.2.188

Drasdo, N. (1977, April 7). The neural representation of visual space [Letter]. *Nature, 266,* 554–556. doi:10.1038/266554a0

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113,* 501–517. doi:10.1037/0096-3445.113.4.501

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123,* 161–177. doi:10.1037/0096-3445.123.2.161

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583–597. doi:10.1037/0096-1523.11.5.583

Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology, 58,* 1–48. doi:10.1016/j.cogpsych.2008.05.001

Fiala, J. C., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye blink response. *The Journal of Neuroscience, 16,* 3760–3774.

Francis, G., & Grossberg, S. (1996a). Cortical dynamics of boundary segmentation and reset: Persistence, afterimages, and residual traces. *Perception, 25,* 543–567. doi:10.1068/p250543

Francis, G., & Grossberg, S. (1996b). Cortical dynamics of form and motion integration: Persistence, apparent motion and illusory contours. *Vision Research, 36,* 149–173. doi:10.1016/0042-6989(95)00052-2

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research, 34,* 1089–1104. doi:10.1016/0042-6989(94)90012-4

Fried, S. I., Münch, T. A., & Werblin, F. S. (2002, November 28). Mechanisms and circuitry underlying directional selectivity in the retina [Letter]. *Nature, 420,* 411–414. doi:10.1038/nature01179

Galletti, C., & Battaglini, P. P. (1989). Gaze-dependent visual neurons in area V3A of monkey prestriate cortex. *The Journal of Neuroscience, 9,* 1112–1125.

Gancarz, G., & Grossberg, G. (1998). A neural model of the saccade generator in the reticular formation. *Neural Networks, 11,* 1159–1174. doi:10.1016/S0893-6080(98)00096-3

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30,* 535–574. doi:10.1146/annurev.neuro.29.051605.113038

Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences of the United States of America, 60,* 758–765. doi:10.1073/pnas.60.3.758

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 213–257.

Grossberg, S. (1978a). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17,* 199–219. doi:10.1016/0022-2496(78)90016-0

Grossberg, S. (1978b). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York, NY: Academic Press.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87,* 1–51. doi:10.1037/0033-295X.87.1.1

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics, 55,* 48–120. doi:10.3758/BF03206880

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review, 104,* 618–658. doi:10.1037/0033-295X.104.3.618

Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision, 12,* 163–185. doi:10.1163/156856899X00102

Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition: An International Journal, 8,* 1–44. doi:10.1006/ccog.1998.0372

Grossberg, S. (2000a). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences, 4,* 233–246. doi:10.1016/S1364-6613(00)01464-9

Grossberg, S. (2000b). How hallucinations may arise from brain mechanisms of learning, attention, and volition. Invited article for the *Journal of the International Neuropsychological Society, 6,* 579-588.

Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews, 2,* 47–76. doi:10.1177/1534582303002001003

Grossberg, S., Boardman, I., & Cohen, C. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 481–503. doi:10.1037/0096-1523.23.2.481

Grossberg, S., & Huang, T.-R. (2009). ARTSCENE: A neural system for natural scene classification. *Journal of Vision, 9,* 1–19. doi:10.1167/9.4.6

Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements.* Amsterdam, North Holland: Pergamon Press.

Grossberg, S., & Kuperstein, M. (1986/1989). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements* (2nd ed.). Amsterdam, North Holland: Pergamon Press.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211. doi:10.1037/0033-295X.92.2.173

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38,* 141–171. doi:10.3758/BF03198851

Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research, 41,* 2521–2553. doi:10.1016/S0042-6989(01)00131-6

Grossberg, S., & Pearson, L. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review, 115,* 677–732. doi:10.1037/a0012618

Grossberg, S., & Pilly, P. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research, 48,* 1345–1373. doi:10.1016/j.visres.2008.02.019

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *The Journal of Neuroscience, 17,* 9706–9725.

Grossberg, S., & Rudd, M. (1989). A neural architecture for visual motion perception: Group and element apparent motion. *Neural Networks, 2,* 421–450. doi:10.1016/0893-6080(89)90042-7

Grossberg, S., & Rudd, M. E. (1992). Cortical dynamics of visual motion perception: Short-range and long-range apparent motion. *Psychological Review, 99,* 78–121. doi:10.1037/0033-295X.99.1.78

Grossberg, S., & Stone, G. O. (1986). Neural dynamics of attention switching and temporal order information in short-term memory. *Memory & Cognition, 14,* 451–468. doi:10.3758/BF03202517

Grossberg, S., & Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research, 1218,* 278–312. doi:10.1016/j.brainres.2008.04.024

Grossberg, S., & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research, 45,* 1725–1743. doi:10.1016/j.visres.2005.01.006

Kelly, F., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics, 62,* 1596–1618. doi:10.3758/BF03212158

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Toward the underlying neural circuitry. *Human Neurobiology, 4,* 219–227.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124. doi:10.1037/0033-295X.96.1.101

Lu, Z.-L., & Sperling, G. (1995a). Attention-generated apparent motion. *Nature, 377,* 237–239.

Lu, Z.-L., & Sperling, G. (1995b). The functional architecture of human visual motion perception. *Vision Research, 35,* 2697–2722. doi:10.1016/0042-6989(95)00025-U

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York, NY: Oxford University.

McMains, S. A., & Somers, D. C. (2005). Processing efficiency of divided spatial attention mechanisms in human visual cortex. *The Journal of Neuroscience, 25,* 9444–9448. doi:10.1523/JNEUROSCI.2647-05.2005

Nakamura, K., & Colby, C. L. (2000). Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *Journal of Neurophysiology, 84,* 677–692.

Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 4026–4031. doi:10.1073/pnas.052379899

Nakamura, H., Kuroda, T., Wakita, M., Kusunoki, M., Kato, A., Mikami, A., . . . Itoh, K. (2001). From three-dimensional space vision to prehensile hand movements: The lateral intraparietal area links the area V3A and the anterior intraparietal area in macaques. *The Journal of Neuroscience, 21,* 8174–8187.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383. doi:10.1016/0010-0285(77)90012-3

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175. doi:10.1023/A:1011139631724

Pack, C., Grossberg, S., & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience, 13,* 102–120. doi:10.1162/089892901564207

Palmer, J., Huk, A., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5,* 376–404. doi:10.1167/5.5.1

Posner, M. I. (1980). Orienting of attention. *The Quarterly Journal of Experimental Psychology, 32,* 3–25. doi:10.1080/00335558008248231

Potter, M. C. (1975, March 14). Meaning in visual search. *Science, 187,* 965–966. doi:10.1126/science.1145183

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179–197. doi:10.1163/156856888X00122

Raizada, R. D., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex, 13,* 100–113. doi:10.1093/cercor/13.1.100

Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review, 93,* 180–206. doi:10.1037/0033-295X.93.2.180

Reichardt, W. (1961). Autocorrection, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303–317). New York, NY: Wiley.

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron, 37,* 853–863. doi:10.1016/S0896-6273(03)00097-7

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of Neuroscience, 22,* 9475–9489.

Schwartz, E. L. (1977). Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biological Cybernetics, 25,* 181–194. doi:10.1007/BF01885636

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195–200. doi:10.1111/j.1467-9280.1994.tb00500.x

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology, 86,* 1916–1936.

Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109,* 260–305. doi:10.1037/0033-295X.109.2.260

Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research, 29,* 619–626. doi:10.1016/0042-6989(89)90047-3

Solomon, J. A., Sperling, G., & Chubb, C. (1993). The lateral inhibition of perceived contrast is indifferent to on-center/off-center segregation, but specific to orientation. *Vision Research, 33,* 2671–2683. doi:10.1016/0042-6989(93)90227-N

Sperling, G., & Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *Journal of the Optical Society of America, 58,* 1133–1145. doi:10.1364/JOSA.58.001133

Srihasam, K., Bullock, D., & Grossberg, S. (2009). Target selection by frontal cortex during coordinated saccadic and smooth pursuit eye movements. *Journal of Cognitive Neuroscience, 21,* 1611–1627. doi:10.1162/jocn.2009.21139

Sutter, A., Beck, J., & Graham, N. (1989). Contrast and spatial variables in texture segregation: Testing a simple spatial-frequency channel model. *Perception & Psychophysics, 46,* 312–332. doi:10.3758/BF03204985

Theeuwes, J., Mathot, S., & Kingstone, A. (2010). Object-based eye movements: The eyes prefer to stay within the same object. *Attention, Perception, & Psychophysics, 72,* 597–601. doi:10.3758/APP.72.3.597

Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: Stereoattention and surface perception in depth reconstruction. *Perception, 24,* 127–153. doi:10.1068/p240127

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review, 108,* 550–592. doi:10.1037/0033-295X.108.3.550

van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichard detectors. *Journal of the Optical Society of America A: Optics and Image Science, 2,* 300–321. doi:10.1364/JOSAA.2.000300

Wallach, H. (1935). On the visually perceived direction of motion. *Psychologische Forschung, 20,* 325–380. doi:10.1007/BF02409790

Webster, M. J., Bachevalier, J., & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex, 4,* 470–483. doi:10.1093/cercor/4.5.470

Weichselgartner, E., & Sperling, G. (1987). Dynamics of automatic and controlled visual attention. *Science, 238,* 778–780. doi:10.1126/science.3672124

Yantis, S. (1992). Multielement visual tracking: Attention and perceptual organization. *Cognitive Psychology, 24,* 295–340. doi:10.1016/0010-0285(92)90010-Y

Yarbus, A. L. (1961). Eye movements during the examination of complicated objects. *Biofizika, 6,* 52–56.