Psychological Review
2005, Vol. 112, No. 2, 291–328

Copyright 2005 by the American Psychological Association
0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.2.291

A Neural Theory of Visual Attention: Bridging Cognition and Neurophysiology

Claus Bundesen, Thomas Habekost, and Søren Kyllingsbæk
University of Copenhagen

A neural theory of visual attention (NTVA) is presented. NTVA is a neural interpretation of C. Bundesen's (1990) theory of visual attention (TVA). In NTVA, visual processing capacity is distributed across stimuli by dynamic remapping of receptive fields of cortical cells such that more processing resources (cells) are devoted to behaviorally important objects than to less important ones. By use of the same basic equations used in TVA, NTVA accounts for a wide range of known attentional effects in human performance (reaction times and error rates) and a wide range of effects observed in firing rates of single cells in the primate visual system. NTVA provides a mathematical framework to unify the 2 fields of research—formulas bridging cognition and neurophysiology.

This article presents a neural theory of visual attention: NTVA. The theory proposes a close link between attentional function at the behavioral and at the cellular level. By use of the same basic equations used in TVA (theory of visual attention; Bundesen, 1990), the theory accounts for both a large portion of the attentional effects reported in the psychological literature and a large portion of the attentional effects demonstrated in individual neurons. NTVA thus provides a mathematical framework to unify these two fields of research.

NTVA is a further development of TVA (Bundesen, 1990). TVA is a formal, computational theory that accounts for a wide range of attentional effects in mind and behavior reported in the psychological literature. At the heart of TVA are two equations, and NTVA is a neural interpretation of these equations. The equations jointly describe two mechanisms of attentional selection: *filtering* (selection of objects) and *pigeonholing* (selection of features). In NTVA, filtering affects the number of cells (cortical neurons) in which an object is represented, whereas pigeonholing is a multiplicative scaling of the level of activation in cells coding for particular features (see Figure 1). The total activation representing a visual categorization of the form "object *x* has feature *i*" is directly proportional to both the number of neurons representing the categorization (which is controlled by filtering) and the level of activation of the individual neurons representing the categorization (controlled by pigeonholing), and Equation 1 of TVA essentially expresses this fact.

Filtering is done in such a way that the number of cells in which an object is represented increases with the behavioral importance of the object (parallel processing with differential allocation of resources). More specifically, the probability that a cortical neuron represents a particular object within its classical receptive field (RF) equals the attentional weight of the object divided by the sum of the attentional weights across all objects in the RF.

Equation 2 of TVA describes how attentional weights are computed, and logically this computation must occur before processing resources (cells) can be distributed in accordance with the weights. Accordingly, in NTVA, a normal perceptual cycle consists of two waves: a wave of unselective processing followed by a wave of selective processing. During the first wave, cortical processing resources are distributed at random (unselectively) across the visual field. At the end of the first wave, an attentional weight has been computed for each object in the visual field and stored in a saliency map. The weights are used for reallocation of attention (visual processing capacity) by dynamic remapping of RFs of cortical neurons such that the number of neurons allocated to an object increases with the attentional weight of the object. Hence, during the second wave, cortical processing is selective in the sense that the amount of processing resources (number of neurons) allocated to an object depends on the attentional weight of the object. Because more processing resources are devoted to behaviorally important objects than to less important ones, the important objects are more likely to become encoded into visual short-term memory (VSTM). The VSTM system is conceived as a (*K*-winners-take-all) feedback mechanism that sustains activity in the neurons that have won the attentional competition.

This article contains three main sections. In the first main section, we review the equations describing the basic mechanisms of selection in TVA and summarize how TVA has been applied to a broad range of findings on human performance in visual recognition and attention tasks. The neural interpretation of TVA, NTVA, is presented in the second main section. We first develop the neural interpretation of Equation 1 of TVA in general terms. We then present a set of simple networks for performing the attentional operations of NTVA. We show how the computations of the networks correspond to the original equations of TVA, and

Claus Bundesen, Thomas Habekost, and Søren Kyllingsbæk, Center for Visual Cognition, Department of Psychology, University of Copenhagen, Copenhagen, Denmark.

This research was supported by grants from the Danish Research Council for the Humanities and the Carlsberg Foundation. We thank Kyle Cave, John Duncan, Glyn Humphreys, Axel Larsen, Gordon Logan, Tom Palmeri, Antonino Raffone, Jeff Schall, and Werner Schneider for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Claus Bundesen, Center for Visual Cognition, Department of Psychology, University of Copenhagen, Njalsgade 90, DK-2300 Copenhagen S, Denmark. E-mail: bundesen@psy.ku.dk

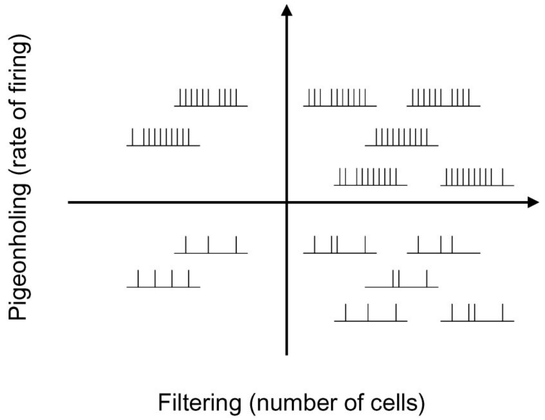

Figure 1. Attentional selection in NTVA (neural theory of visual attention): combined effects of *filtering* (selection of objects) and *pigeonholing* (selection of features) on the set of cortical spike trains representing a particular visual categorization of the form "object x has feature i." The four conditions (quadrants) correspond to the factorial combinations of two levels of filtering (weak vs. strong support to object x) and two levels of pigeonholing (weak vs. strong support to feature i). Filtering changes the number of cortical neurons in which an object is represented. Pigeonholing changes the rate of firing of cortical neurons coding for a particular feature.

we sketch where such networks may be localized in the primate brain. However, NTVA is not a connectionist model, and it does not depend critically on a specific anatomical localization of the proposed computations. NTVA is a fairly general neurophysiological interpretation of TVA, and the particular networks that we present may be regarded as merely proofs of the existence of simple and biologically plausible neural networks implementing the attentional operations implied by NTVA.

In the third main section, we apply NTVA to a wide range of findings from single-cell studies of attentional effects on visual representations in primates. The findings include attentional effects with multiple stimuli in the RF of the recorded neuron, effects with a single stimulus in the RF, and effects on baseline firing. By far the strongest changes of a cell's firing rate occur when multiple objects are present in the classical RF. Under these conditions, a general finding has been that attention to one of the objects modulates the firing rate either up or down, depending on the cell's sensory preference for the attended object. As detailed in the *Attentional Effects With Multiple Stimuli in the RF* section, this finding is readily explained by NTVA's notion of filtering on the basis of attentional weights. A second typical effect of attention is a modest modulation of firing rates with a single stimulus in the RF. In the *Attentional Effects With a Single Stimulus in the RF* section, this finding is explained by pigeonholing or, in some cases, by the presence of stimuli other than the one defined by the experimenter. A third common effect of attention is an increase in a cell's baseline firing rate when a target is expected to appear in its RF. In the *Attentional Effects on Baseline Firing* section, this effect is explained as the neural correlate of a more or less schematic mental image of the anticipated stimulus (a representation in VSTM). Our review spans the major findings of the research area: attentional effects with single versus multiple stimuli in the RF, spatial and nonspatial attentional effects, interactions with luminance contrast, multiplicative modulations of firing rates, and baseline shifts. We consider studies of visual areas V1, V2, V4, and inferotemporal (IT) cortex, as well as the middle temporal visual area (MT), medial superior temporal area (MST), and prefrontal (PF) cortex, and we investigate how attentional effects depend on the position in the cortical processing hierarchy.

Formal TVA

Basic Assumptions

In TVA, both visual identification and selection of objects in the visual field consist in making visual categorizations. A visual categorization has the form "object x has feature i" or, equivalently, "object x belongs to category i." Here, object x is a perceptual unit in the visual field, feature i is a visual feature (e.g., a certain color, shape, movement, or spatial position), and category i is a visual category (the class of all objects that have feature i).

A visual categorization is made by immediate perception if and when the categorization is encoded into VSTM. If and when the visual categorization is made that x belongs to i, object x is said (a) to be selected and (b) to be identified as a member of category i. Thus, an object is said to be selected if, and only if, it is identified as a member of one or another category. Similarly, an object is said to be represented in VSTM if, and only if, some categorization of the object is represented in VSTM.

Once a visual categorization of an object completes processing, the categorization enters VSTM, provided that memory space for the categorization is available in VSTM. The capacity of VSTM is limited to K different objects. Space is available for a new categorization of object x if object x is already represented in the store (with one or another categorization) or if fewer than K objects are represented in the store (cf. Luck & Vogel, 1997). There is no room for a categorization of object x if VSTM is filled up with other objects.

Rate Equation

Clearing of VSTM effectively starts a race among objects in the visual field to become encoded into VSTM. An object is encoded in VSTM if and when any categorization of the object is encoded in VSTM, so each object x is represented in the encoding race by all possible categorizations of the object. The rate, $v(x, i)$, at which a particular visual categorization, "x belongs to i," is encoded into VSTM is given by Equation 1 of TVA,

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z \in S} w_z}, \tag{1}$$

where $\eta(x, i)$ is the strength of the sensory evidence that x belongs to category i, β_i is a perceptual decision bias associated with category i ($0 \le \beta_i \le 1$), and

$$w_x / \sum_{z \in S} w_z$$

is the relative attentional weight of object x (i.e., the weight of object x, w_x, divided by the sum of weights across all objects in the visual field, S).

Bundesen (1990) interpreted $v(x, i)$ as the rate of point events generated by a Poisson process and assumed that a single point event suffices for encoding of a categorization into VSTM. (Here, a point event may be interpreted as a single neural spike or as a single volley of r or more synchronized neural spikes.) In this case, $v(x, i)$ equals the hazard function of the event that the categorization "x belongs to i" completes processing at a certain point in time (i.e., the probability density that the event occurs at time t, given that it has not occurred before time t). Logan (1996, 2002) also interpreted $v(x, i)$ as the event rate of a Poisson process but explored the implications of assuming that, depending on threshold settings, many point events from the Poisson process representing a particular categorization may be needed for encoding of the categorization into VSTM.

Weight Equation

The attentional weights in Equation 1 are derived from *pertinence* values. Every visual category is supposed to have a certain pertinence. The pertinence of a category is a measure of the current importance of attending to objects that belong to the category. The weight of an object x in the visual field is given by Equation 2 of TVA,

$$w_x = \sum_{j \in R} \eta(x,j)\,\pi_j, \qquad (2)$$

where R is the set of all visual categories, $\eta(x, j)$ is the strength of the sensory evidence that object x belongs to category j, and π_j is the pertinence of category j. By Equation 2, the attentional weight of an object is a weighted sum of pertinence values. The pertinence of a given category enters the sum with a weight equal to the strength of the sensory evidence that the object belongs to the category.

By Equations 1 and 2, $v(x, i)$ is a function of η, β, and π values. When these values are determined, Poisson events corresponding to different perceptual categorizations are mutually independent (cf. Bundesen, Kyllingsbæk, & Larsen, 2003).

Mechanisms of Selection

Equations 1 and 2 describe two mechanisms of selection: a mechanism for selection of objects (filtering) and a mechanism for selection of categories (pigeonholing). The filtering mechanism is represented by pertinence values and attentional weights. As an example, if selection of red objects is wanted, the pertinence of *red* should be high. Equation 2 implies that when *red* has a high pertinence, red objects get high attentional weights. Accordingly, by Equation 1, processing of red objects is fast, so red objects are likely to win the processing race and be encoded into VSTM.

The pigeonholing mechanism is represented by perceptual decision-bias parameters. Pertinence values determine which objects are selected, but perceptual decision-bias parameters determine how the objects are categorized. If particular types of categorizations are desired, decision-bias parameters of the relevant categories should be high. By Equation 1, then, the desired types of categorizations are likely to be made.

Consider how filtering and pigeonholing can be combined. For example, consider partial report of red digits from a mixture of red and black digits. A sufficient strategy for performing the task is as follows. To select the red objects, the pertinence value of the visual category *red* is set high, but other pertinence values are kept low. The effect is to speed up the processing of all types of categorizations of red objects. To perceive the identity of the red digits rather than other attributes of the objects, 10 perceptual decision-bias parameters (1 for each type of digit) are set high, but other perceptual decision-bias parameters are kept low. The effect is to speed up the processing of categorizations with respect to digit types. The combined effect of the adjustments of pertinence and decision-bias parameters is to speed up the processing of categorizations of red objects with respect to digit types in relation to any other categorizations.

The above example demonstrates the power of the mechanisms of selection contained in TVA. When the selection system is coupled to a sensory system that supplies appropriate η values, and when pertinence and decision-bias parameters have been set, both filtering and pigeonholing are accomplished by a race between visual categorization processes whose rate parameters are determined through the simple algebraic operations of Equations 1 and 2. Thus, the theory yields a computational account of selective attention in vision.

Applications

TVA has been applied to findings from a broad range of paradigms concerned with single-stimulus identification and selection from multiobject displays. In addition, TVA has been applied in clinical neuropsychological research, and the scope of the theory has been extended to memory and executive functions.

Single-Stimulus Identification

For single-stimulus identification, TVA provides a simple derivation of the biased-choice model of Luce (1963; see Bundesen, 1990; see also Bundesen, 1993). The biased-choice model has been successful in explaining many experimental findings on effects of visual discriminability and bias (see, e.g., Townsend & Ashby, 1982; Townsend & Landon, 1982).

Selection From Multiobject Displays

Bundesen (1990) applied TVA to many findings on selection from multiobject displays. The findings included effects of (a) object integrality in selective report (e.g., Duncan, 1984; see also Bundesen, 1991; Bundesen et al., 2003), (b) number and spatial position of targets in studies of divided attention (Posner, Nissen, & Ogden, 1978; Sperling, 1960, 1967; van der Heijden, La Heij, & Boer, 1983), (c) selection criterion and number of distractors in studies of focused attention (Bundesen & Pedersen, 1983; Estes & Taylor, 1964; Treisman & Gelade, 1980; Treisman & Gormican, 1988; see also Bricolo, Gianesini, Fanini, Bundesen, & Chelazzi, 2002), (d) joint effects of numbers of targets and distractors in partial report (Bundesen, Pedersen, & Larsen, 1984; Bundesen, Shibuya, & Larsen, 1985; Shibuya & Bundesen, 1988; see also Shibuya, 1991, 1993), and (e) consistent practice in search (W. Schneider & Fisk, 1982; see also Kyllingsbæk, Schneider, & Bundesen, 2001).

Attention Deficits After Brain Damage

Recently, the principles of TVA have been applied in the study of attention deficits after brain damage. Duncan et al. (1999)

showed how analysis in terms of parameters defined by TVA enables a very specific measurement of attentional deficits in visual neglect patients. TVA-based assessment has also been used in case studies of simultanagnosia (Duncan et al., 2003) and subclinical attention deficits (Habekost & Bundesen, 2003). Currently, research groups in Cambridge, England, Munich, Germany, and Copenhagen, Denmark, are extending these investigations to other patient groups (e.g., groups suffering from cortical and subcortical strokes, Huntington's disease, schizophrenia, depression).

Other Cognitive Domains

Logan (1996) proposed an extension of TVA, the CODE theory of visual attention (CTVA), which combines TVA with the contour detector theory of perceptual grouping by proximity (van Oeffelen & Vos, 1982, 1983). CTVA explains a wide range of spatial effects in visual attention (see Logan, 1996; Logan & Bundesen, 1996; see also Bundesen, 1998a, 1998b).

Logan and Gordon (2001) extended CTVA into a theory of executive control in dual-task situations that accounts for crosstalk, set-switching costs, and concurrence costs as well as dual-task interference. The theory, ECTVA, assumes that executive processes control subordinate processes by manipulating their parameters. TVA is used as the theory of subordinate processes, so a task set is defined as a set of TVA parameters that is sufficient to configure TVA to perform a task. Set switching is viewed as a change in one or more of these parameters, and the time taken to change a task set is assumed to depend on the number of parameters to be changed (Logan & Gordon, 2001; see also Logan & Bundesen, 2003, 2004).

Recently, Logan (2002) proposed an instance theory of attention and memory (ITAM) that combines ECTVA with the exemplar-based random-walk model of categorization (Nosofsky & Palmeri, 1997). The exemplar-based random-walk model itself is a combination of Nosofsky's (1986) generalized context model of categorization and Logan's (1988) instance theory of automaticity. In its integration of theories of attention, categorization, and memory, the development of ITAM seems to be an important step toward a unified account of visual cognition.

NTVA

NTVA gives the equations of TVA an interpretation at the level of individual neurons. A typical neuron in the visual system is assumed, first, to be specialized to represent a single feature and, second, to respond to the properties of only one object at any given time. Formally, if the neuron represents the categorization "*x* has feature *i*" at one time and the categorization "*y* has feature *j*" at another time, then *x* may differ from *y*, but *i* must equal *j*. That is, a neuron can represent different objects at different times, but—learning and development aside—it always represents the same feature *i*. Neurons representing feature *i* are called *feature*-i *neurons*. Feature *i* can be a more or less simple physical feature or a microfeature in a distributed representation (cf. Hinton, McClelland, & Rumelhart, 1986; Page, 2000), and a feature-*i* neuron may be broadly sensitive to feature *i*'s degree of presence rather than being sharply tuned to feature *i*.

The object selection of the neuron occurs by dynamic remapping of the cell's RF such that the functional RF contracts around the selected object. The remapping is done such that the probability that the neuron comes to represent a particular object equals the attentional weight of the object divided by the sum of the attentional weights of all objects in the classical RF. Equation 2 of TVA describes the way in which attentional weights are computed.

Let the *activation* of a neuron (at a certain point in time) by the appearance of an object in its RF be the increase in firing rate (spikes per second) above a baseline rate representing the undriven activity of the neuron. If the baseline rate is zero, the activation is just the firing rate. Independent of the distribution of neurons among objects (filtering), the activation in neurons representing particular features (e.g., the set of feature-*i* neurons or the set of feature-*j* neurons) is scaled up or down (pigeonholing). Equation 1 of TVA,

$$v(x,i) = \eta(x,i)\beta_i \frac{w_x}{\sum_{z\in S} w_z},$$

describes the combined effects of filtering and pigeonholing on the total activation of the population of neurons representing the categorization "object *x* has feature *i*." The *v* value on the left-hand side of the equation is the total activation of the neurons that represent the categorization at the level of processing where objects compete for entrance into VSTM. At this level of processing, the classical RFs of neurons are so large that each one covers the entire visual field. On the right-hand side of the equation, both β_i and

$$w_x/\sum_{z\in S} w_z$$

range between 0 and 1, so $\eta(x, i)$ equals the highest possible value of $v(x, i)$. Thus, $\eta(x, i)$ equals the total activation of the set of all feature-*i* neurons when every feature-*i* neuron represents object *x* (say, *x* is the only object in the visual field) and the featural bias in favor of *i* is maximal (i.e., $\beta_i = 1$). When the proportion of feature-*i* neurons representing object *x*,

$$w_x/\sum_{z\in S} w_z,$$

is less than 1, the total activation representing the categorization "object *x* has feature *i*" is scaled down by multiplication with the factor

$$w_x/\sum_{z\in S} w_z$$

on the right-hand side of the equation. The total activation representing the categorization depends not only on the number of neurons representing the categorization but also on the level of activation of the individual neurons representing the categorization. The bias parameter β_i is a scale factor that multiplies activations of all individual feature-*i* neurons, so the total activation representing the categorization "object *x* has feature *i*" is also directly proportional to β_i. Thus, in the neural interpretation we propose for Equation 1, the total activation representing the categorization "object *x* has feature *i*" is directly proportional to both the number of neurons representing the categorization and the level of activation of the individual neurons representing the categorization. The number of neurons is controlled by

$$w_x / \sum_{z \in S} w_z$$

(filtering), whereas the activation of the individual neurons is controlled by β_i (pigeonholing).

NTVA does not depend critically on a particular anatomical localization of the proposed computations. However, we suggest some plausible ways in which visual processing may be distributed across the human brain. One possibility is illustrated in Figure 2. Bias (β) and pertinence (π) values are not generated within the visual system but, rather, derive from higher order areas in frontal and parietal cortex and, directly or indirectly, from the limbic system. The parameter settings are transmitted via projections to the visual system (cf. Corbetta & Shulman, 2002; Kanwisher & Wojciulik, 2000; Kastner & Ungerleider, 2000; Mesulam, 1981, 1990; see also Logan & Gordon, 2001; O'Reilly, Braver, & Cohen, 1999).

Below, we present a set of simple networks for performing the attentional operations of NTVA. The principle of race-based attentional competition is implemented by use of winner-take-all neural networks, and the VSTM system is conceived as a (*K*-winners-take-all) feedback mechanism that sustains activity in the neurons that have won the attentional competition. We show how

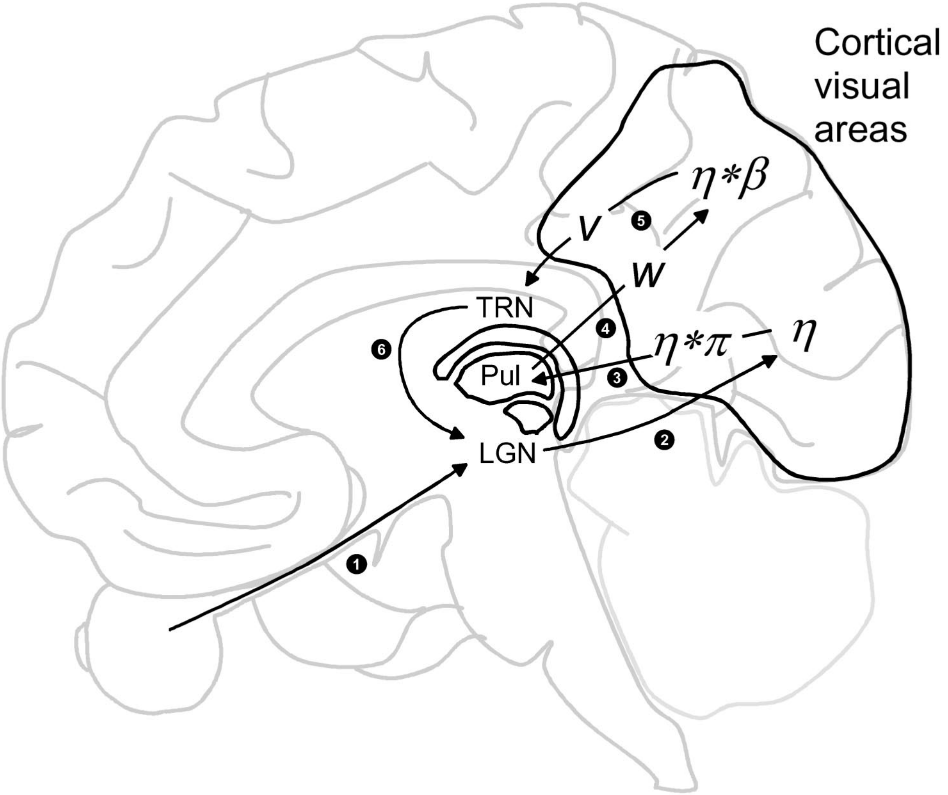

Figure 2. Possible distribution of visual processing across the human brain. Visual information from the eye enters the lateral geniculate nucleus (LGN) of the thalamus (1) and is transmitted to striate and extrastriate cortical areas, where η values (strengths of evidence that objects at particular scales and positions have particular features) are computed (2). The η values are multiplied by π (pertinence) values, and the products are transmitted from the cortex to a saliency map in the pulvinar (Pul) nucleus of the thalamus, where the products are summed up as attentional weights of the stimulus objects (3). After the first (unselective) wave of processing, cortical processing capacity is redistributed by means of attentional weight (w) signals from the pulvinar to the cortex, such that during the second (selective) wave of processing, objects with high attentional weights are processed by many neurons (4). The resulting η values are multiplied by β (bias) values, and the products are transmitted from the cortex to a visual short-term memory (VSTM) map of locations, which is tentatively localized in the thalamic reticular nucleus (TRN [5]). When the VSTM map is initialized, objects in the visual field effectively start a race to become encoded into VSTM. In this race, each object is represented by all possible categorizations of the object, and each possible categorization participates with a firing rate (v value) proportional to the corresponding η value multiplied by the corresponding β value. For the winners of the race, the TRN gates activation representing a categorization back to some of those cells in LGN whose activation supported the categorization (6). Thus, activity in neurons representing winners of the race is sustained by positive feedback.

the computations of these networks correspond to the original equations of TVA. We also detail where the networks may be localized in the primate brain.

In the first subsection, we present a basic building block of the model: a winner-take-all network that records the result of a competition between neural signals. This is the most fundamental implementation of the race for attentional selection. Next, we show how dynamic remapping of RFs of visual neurons in a hierarchical network can yield a distribution of processing resources conforming to the principles of TVA. The third subsection demonstrates how features extracted at different levels in the visual network may be linked to objects at particular locations by routing of information to topographic maps. The mechanism is used in modeling the perceptual cycle in NTVA (see the fourth subsection below and Appendix A). During the first wave of processing, attentional weight components are computed for objects in the visual field, and the components are routed to a topographic saliency map in which they are summed up in accordance with Equation 2 of TVA. A possible anatomical location of the saliency map is found in the pulvinar nucleus of the thalamus. When the saliency map has been configured, cortical processing resources are redistributed across the objects in the visual field by remapping of RFs such that the number of cells allocated to a particular object during the second wave of processing becomes proportional to the attentional weight of the object. We show that during the second wave of processing (i.e., when RFs have been remapped), visual categorizations are made in close agreement with Equation 1 of TVA.

The VSTM system (see the fifth subsection below) is assumed to depend on another topographic map of objects: a VSTM map, possibly located in the reticular nucleus of the thalamus. The VSTM map is constructed as a *K*-winners-take-all network that sustains perceptual categorizations of up to *K* selected (attended) objects by way of feedback loops to the hierarchical visual network. The feedback mechanism is assumed to be the neural basis of mental images. In the sixth subsection, we discuss how visual identification that is more complex than immediate perception may be based on the neural networks of NTVA (see also Appendix B). In the final subsection, we consider relations to other theories.

Winner-Take-All Networks: Recording the Winner of a Race

To make a neural network implementation of TVA, we need a device for recording the winner of a race. A winner-take-all (WTA) cluster of a general type proposed by Grossberg (1976, 1980) can be used. As shown in Figure 3, it consists of a set of units (populations of cells) such that each unit excites itself and inhibits all other units in the cluster. Suppose that when the cluster is initialized, a signal of a certain strength (e.g., *r* spikes) from the environment to one of the units is sufficient to trigger this unit. Suppose that once the unit has been triggered, it keeps on firing because of its self-excitation. And suppose that when the unit is firing, it inhibits the other units in the cluster so strongly that they cannot be triggered by signals from the environment. If this is the case, one can read from the state of the cluster which of the units received the first above-threshold signal from the environment after the cluster was initialized. Thus, the cluster can serve as a device for recording the winner of a race (for related applications of WTA networks, see, e.g., Cave, 1999; Koch & Ullman, 1985; Lee, Itti, Koch, & Braun, 1999; Tsotsos et al., 1995).

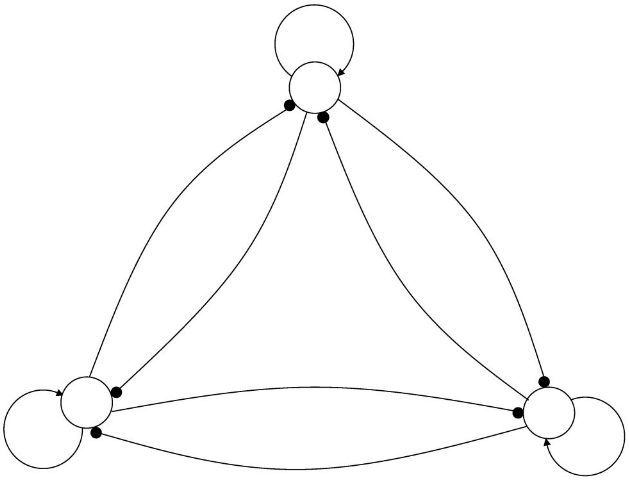

Figure 3. Winner-take-all network with three units (open circles). Excitatory connections are shown by arrows. Inhibitory connections are shown by lines ending in solid circles.

Dynamic Remapping of RFs: Distributing Processing Capacity

In this section, we sketch a simple neural network in which processing capacity (neurons) can be distributed and redistributed among stimuli by opening and closing of gates that control communication from lower to higher levels of the visual system. As gates are opened and closed, RFs are remapped, and reallocation of attention consists in dynamic remapping of RFs. The network can be controlled by attentional weight signals such that the expected number of cells that represent a particular object at the highest level of the visual system becomes proportional to the relative attentional weight of the object. The suggested network was inspired by findings of Moran and Desimone (1985).

In their classical work with monkeys, Moran and Desimone (1985) found that when a target and a distractor were both within the RF of a cell in visual area V4 or IT cortex, the response of the cell to the distractor was dramatically reduced. In the words of Desimone and Ungerleider (1989), the typical cell responded “as if its RF had contracted around . . . [the] attended stimulus” (p. 293). The effect depended on both the target and the distractor being located within the recorded neuron’s RF. “If one stimulus was located within the RF and one outside, the locus of the animal’s attention had no effect on the neuron’s response” (Desimone & Ungerleider, 1989, p. 292).

Desimone and Ungerleider’s (1989) description suggests the following hypothesis. Consider a cell in area V4 or the IT cortex. Suppose that several objects are present within the classical (anatomical) RF of the cell. By an attentional gating mechanism, the *effective* (functional) RF of the cell can be contracted around a single one of the objects. The probability that the RF contracts around a particular object increases with the attentional weight of the object.

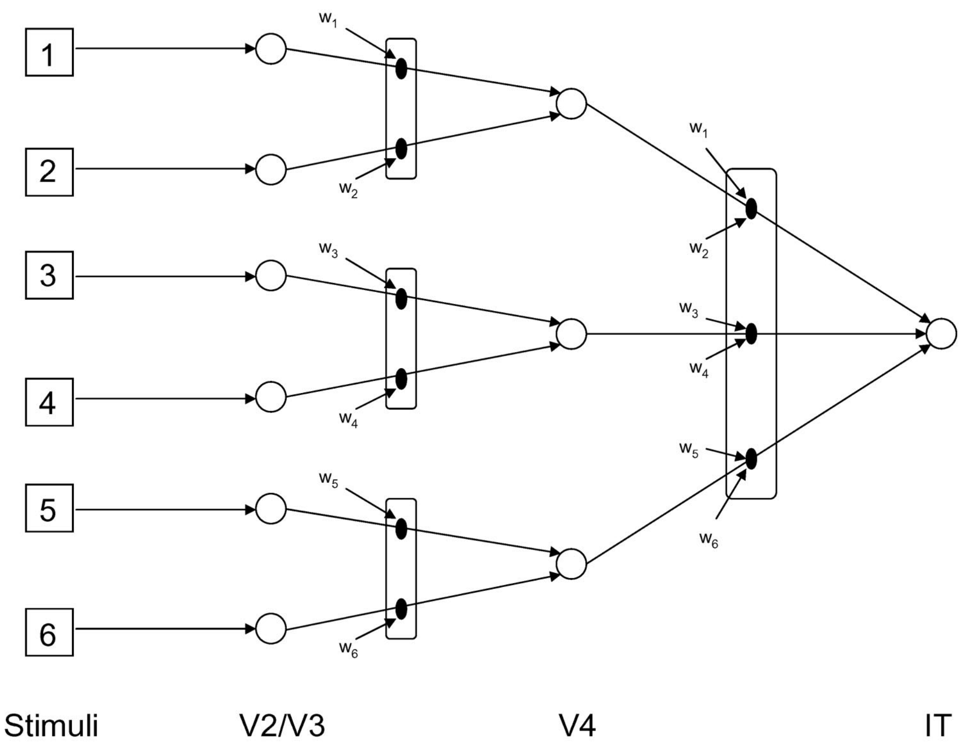

Figure 4. Hierarchical network with processing units at three levels: visual areas V2/V3 and V4 and inferotemporal (IT) cortex. Each V2/V3 unit has one stimulus in its classical receptive field (RF); each V4 unit has two stimuli in its classical RF; and the IT unit has all six of the stimuli in its classical RF. By attentional gating, the effective RFs of V4 and IT units are contracted such that each of the effective RFs contains only one object: The symbol ⁚ stands for a gate that is open on only the upper or the lower line, and ⁝ stands for a gate that is open on only one of the three lines. The gates are set by competing attentional weight signals. For $i = 1, \ldots, 6$, w_i is the attentional weight of stimulus i.

Consider how gating on the basis of attentional weights could be done in a simple hierarchical network. Figure 4 shows a network with processing units at three levels labeled *V2/V3*, *V4*, and *IT* and a number of gates. Figure 5 shows a gate in close-up. It consists of a WTA cluster of units (one unit for each line through the gate), logical AND units (one on each line through the gate), and a logical OR unit (which collects information from the AND units). The WTA cluster records the winner of a race between signals to the two units in the cluster. Thus, if the upper unit receives a signal before the lower unit does, the upper unit will be excited, the lower unit will be inhibited, and only the upper line through the gate will be open.

Signals from the environment to units in the WTA clusters represent attentional weights. As indicated in Figure 4, a WTA cluster that gates responses from cells at a lower level to a cell at a higher level receives one attentional weight signal for every object within the classical RF of the higher level cell. Attentional weight signals for objects within the RF of the lower level cell go to that unit in the WTA cluster that supports communication from the lower level cell to the higher level cell. Attentional weight signals for objects outside the RF of the lower level cell (but within the RF of the higher level cell) go to other, competing units in the WTA cluster.

Consider the probability that the receptive field of the IT unit contracts around Object 1 rather than around one of the other objects. This equals the probability that both the upper of the gates from V2/V3 to V4 and the gate from V4 to IT open on their upper lines. To estimate the probability, we make some simple assumptions. Let arrival times of attentional weight signals for an object x be exponentially distributed with a rate parameter equal to the attentional weight of x, w_x. Let attentional weight signals for different objects be stochastically independent. Finally, let attentional weight signals that represent the same object x but go to different gates be stochastically independent (say, they stem from different members of a population of cells, each of which fires independently with a rate of w_x). On these assumptions, the probability that the upper of the gates from V2/V3 to V4 opens on its upper line is $w_1/(w_1 + w_2)$.[1] The probability that the gate from V4 to IT opens on its upper line is

$$(w_1 + w_2)/ \sum_{i=1,6} w_i.$$

And the probability of both events equals the product of the two probabilities—that is,

[1] If $X_1, X_2, \ldots, X_n$ are mutually independent, exponentially distributed random variables with rate parameters $v_1, v_2, \ldots$, and v_n, respectively, then the probability that X_i ($i = 1, 2, \ldots, n$) is the smallest among the n random variables equals $v_i/(v_1 + v_2 + \ldots + v_n)$.

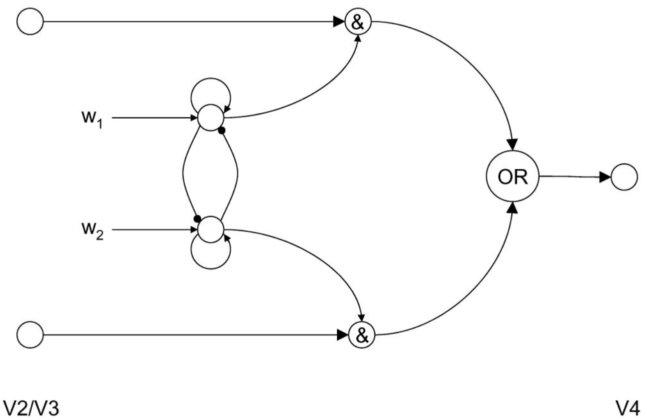

Figure 5. Close-up representation of the uppermost gate in Figure 4. The gate consists of a winner-take-all network with two units (open circles; each controlling one of the two lines through the gate), a logical AND (&) unit on each line, and a logical OR unit transmitting information from either AND unit. For $i = 1, 2$, w_i is the attentional weight of stimulus i. The signal representing the weight of a given stimulus comes from the representation of the stimulus in a topographic saliency map.

$$w_1/\sum_{i=1,6} w_i.$$

Consider a population of IT units, each of which is similar to the one we have considered so far. Each unit has Objects 1–6 and no other objects within its classical RF. If there are N units in the population, the expected number of IT units representing Object 1 equals

$$N\, w_1/\sum_{i=1,6} w_i,$$

the expected number of IT units representing Object 2 equals

$$N\, w_2/\sum_{i=1,6} w_i,$$

and so on. In general, the expected number of IT units representing a particular object should be proportional to the relative attentional weight of the object (the attentional weight of the object divided by the sum of attentional weights across all stimulus objects).

Topographic Maps of Objects: Keeping Things Separate

Dynamic remapping of RFs of V4 and IT neurons should dramatically increase their spatial resolution. In the model we have outlined, the effective RF of an IT neuron is coded by the way that the gates in the network are set. The code formed by the gate settings can be used for directing signals from an IT cell, whose effective RF is contracted around an object at a particular location, to a unit (a neuron or a group of neurons) representing the object at the given location. For example, the code formed by the gate settings can be used for directing the signals to the place in a topographic map corresponding to the location of the stimulus object. A simple network for doing this is shown in Figure 6.

Consider a cell in a WTA cluster in one of the gates on the left side of the figure. The cell has an excitatory connection to an AND unit on a particular input line on the route to the IT unit. Now, the cell also has an excitatory connection to an AND unit on a corresponding output line from the IT unit to a topographic map of objects. By this arrangement, the pattern of open and closed lines on the left side of the figure is duplicated on the right side of the figure. As a result, signals from the IT unit are directed to an appropriate place in the topographic map of objects. In the depicted state of the network, the effective RF of the IT unit equals the RF of the upper V2/V3 unit, and output from the IT unit is directed to a place in the topographic map of objects that corresponds to the RF of the upper V2/V3 unit.

The map of objects may be located at a high level of the visual system, such as the prefrontal (PF) cortex. In this particular case, the dashed lines in Figure 6 symbolize long-ranging connections from cells in gates controlling the flow of information from the retina to the IT (gates on the left side of the figure) to cells controlling the flow of information from IT to the PF cortex (AND units on the right side of the figure). Another possibility is that the map of objects is found at a low level of the visual system, such as the thalamus. By this hypothesis, the right side of Figure 6 depicts a flow of information from the IT unit back toward the retina: A gate and an AND unit connected by a dashed line are located at the same level of the visual system, and the connections symbolized by the dashed lines are quite short. This hypothesis is illustrated in Figure 7, which shows a gate in close-up. The figure is an extension of Figure 5, and the dashed lines correspond to dashed lines in Figure 6. The figure illustrates how the same gates routing information from a particular location on the retina to a certain cell at a higher level of the visual system can be used for routing information back from the high-level cell to a corresponding place in a map of objects found at a low level of the visual system.

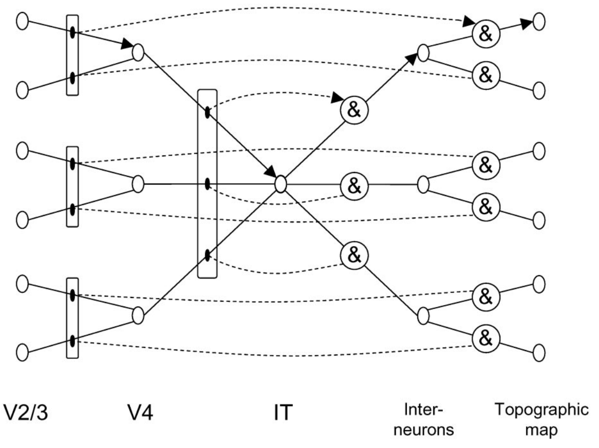

Figure 6. Network for directing signals from an inferotemporal (IT) cortex cell, whose effective receptive field is contracted around an object at a particular location, to a unit (open circle) representing the object at the given location in a topographic map. The network on the left side of the figure is similar to the network in Figure 4, but each cell in a winner-take-all cluster in one of the gates on the left side of the figure controls not only an input line on the route to the IT unit but also a corresponding output line from the IT unit to the topographic map. Dashed lines symbolize connections from cells in gates controlling the flow of information from the retina to the IT to cells controlling the flow of information from IT to the topographic map (AND [&] units on the right side of the figure).

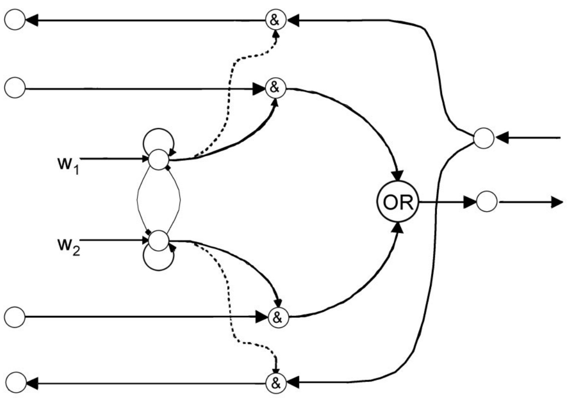

Figure 7. Close-up representation of the uppermost gate in Figure 6. The figure is an extension of Figure 5, and the dashed lines correspond to dashed lines in Figure 6. The figure illustrates how the same gates routing signals along a particular pathway from lower level to higher level cells in the visual system can be used for routing signals along a parallel pathway in the opposite direction. The gate is bistable. In one state, both upper lines through the gate are open and both lower lines are closed; in the other state, both lower lines are open and both upper lines are closed. When the upper lines are open, signals are directed from the retinal location corresponding to the pair of cells in the upper left corner through the ascending pathway originating in the lower member of the rightmost pair of cells. In this state, signals arriving through the descending pathway to the upper member of the rightmost pair of cells are directed back toward the retinal location corresponding to the pair of cells in the upper left corner rather than toward the retinal location corresponding to the pair of cells in the lower left corner. For $i = 1, 2$, w_i is the attentional weight of stimulus i. & = logical AND unit; OR = logical OR unit.

Visual pathways are generally found in pairs such that a pathway from Area A to Area B is accompanied by a pathway from Area B to Area A (see, e.g., Zeki, 1993). From an engineering point of view, it seems easy to design the visual system so that the two pathways in a pair open and close at the same time (bidirectional opening and closing). Thus, from an engineering point of view, it seems simple to design the system so that information about an object at a particular location in the visual field is processed at high levels of the visual system but is routed back to the correct place in a topographic map of objects at a low level of the system (cf. W. X. Schneider, 1995).

The Perceptual Cycle

In NTVA, the perceptual cycle consists of two waves: a wave of unselective processing followed by a wave of selective processing. During the first wave, cortical processing is unselective in that the processing capacity is distributed at random across the visual field. At the end of the first wave, an attentional weight has been computed for each object in the visual field. The weights are found as levels of activation in a saliency map, which may be located in the pulvinar. The weights are used for redistribution of cortical processing capacity across the objects in the visual field by dynamic remapping of RFs such that the expected number of cells allocated to a particular object becomes proportional to the attentional weight of the object. During the second wave, cortical processing is selective in that the processing capacity allocated to an object reflects the attentional weight of the object.

Wave of Unselective Processing

The wave of unselective processing comprises initial sensory processing, formation of perceptual units (object segmentation), and computation of attentional weights. Neither the initial sensory processing nor the process of unit formation is specified in TVA. However, Equation 2 of TVA describes the computation of attentional weights. Our neural interpretation of the equation is illustrated in Figure 8. In this interpretation, computation of attentional weights occurs at many different levels of the cortical visual system, perhaps all the way from primary visual cortex (V1) up to IT cortex. As shown in the figure, the attentional weight of an object x is assumed to be computed by summing up a number of inputs (activations measured in spikes per second) to a unit (a neuron or a population of similar neurons) in a saliency map. The saliency map is a topographic map of the attentional weights of objects in the visual field. Each input to the unit for object x in the saliency map is a product of a level of activation of a cortical neuron representing the strength of the sensory evidence that x has some feature, j, and a feedback factor, π'_j.

The cortical feature-j neurons that participate in the computation of the attentional weight of object x are those feature-j neurons that represent object x. The set of those feature-j neurons that represent object x is a subset of those feature-j neurons in whose classical RFs x is present. Little is known about the nature of this subset. For simplicity and specificity, we assume that at any point in time, the effective RF of a neuron is set for analyzing the visual field at a particular position and scale. Without proposing a computational account of position and size invariance in visual recognition (for different approaches, see Heinke & Humphreys, 2003; Riesenhuber & Poggio, 2000; see also Fukushima, 1980; Larsen &

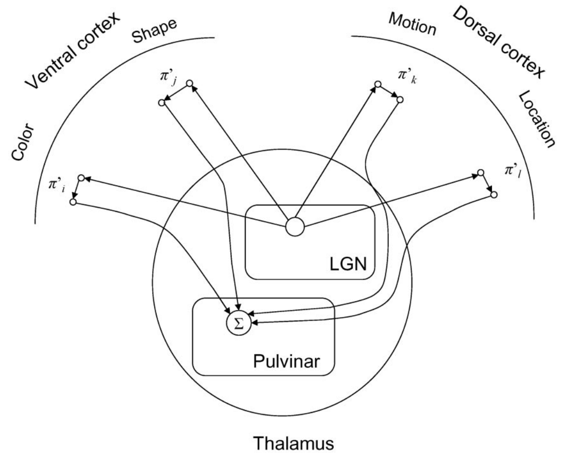

Figure 8. Network for computing attentional weights. Signals from the lateral geniculate nucleus (LGN) are transmitted to striate and extrastriate cortical areas, where η values (strengths of evidence that objects at particular scales and positions have particular features) are computed. The η values are multiplied by π' (pertinence) values, and the products are transmitted from the cortex to the pulvinar nucleus of the thalamus, where the products are summed up as attentional weights of the stimulus objects.

Bundesen, 1978, 1998), we assume (a) that the effective RF can vary in position (shifting) and size (scaling) within the boundaries of the classical RF and (b) that the neuron represents a given object if, and only if, the effective RF fits the position and scale of the object.[2]

For specificity, we use the following terminology: A spatial *position* is a point in space, and a point has no extension. A spatial *location* is centered at a certain position (point in space) and also has an extension (size or spatial scale). The location can be represented by a circle (in 2-D space) or a sphere (in 3-D space), but strictly speaking, it has no shape.

The saliency map is a multiscale map of locations in the visual field: an interconnected set of maps of locations at different scales. Each unit (population of cells) in the multiscale map of locations represents a particular location (i.e., a combination of a particular position and a particular size), and—within boundaries determined by the limited resolution—every location in the visual field is represented by a single unit. To a first approximation, each unit also represents a possible object. At a sufficiently high resolution, no two objects have exactly the same spatial location (position and size) at a given point in time, so the multiscale map of locations can also be regarded as a (more or less precise) map of objects: a multiscale topographic map of objects containing one unit for every possible object (specified by spatial position and size) in the visual field.

During the wave of unselective processing, the set of cortical feature-*j* neurons representing object *x* is a randomly selected subset of all the feature-*j* neurons in whose classical RFs *x* is present. The feedback factor, π'_j, equals that proportion of the activation that is fed back from feature-*j* neurons to the saliency map when attentional weights are computed. In Appendix A, we show that under fairly general assumptions, the network shown in Figure 8 computes attentional weights in accordance with Equation 2 of TVA,

$$w_x = \sum_{j \in R} \eta(x, j)\pi_j,$$

where the attentional weight, w_x, is the expected activation (spikes per second) of the unit representing object *x* in the saliency map; the strength of the sensory evidence that "object *x* has feature *j*," $\eta(x, j)$, equals the sum of the activations of all the feature-*j* neurons in whose classical RFs *x* is present when all of these represent object *x* rather than representing any other objects in their classical RFs; and the pertinence, π_j, is proportional to the feedback factor π'_j.

Anatomical Location of Saliency Map

A number of brain areas have been strongly linked to processing of visual saliency, including the pulvinar nucleus of the thalamus (e.g., Robinson & Cowie, 1997; Robinson & Petersen, 1992; see also Olshausen, Anderson, & Van Essen, 1993), the dorsum of the inferior parietal lobule (see, e.g., Bushnell, Goldberg, & Robinson, 1981), the lateral intraparietal area (see review by Colby & Goldberg, 1999), and the frontal eye fields (see review by Schall & Thompson, 1999). Below, we summarize the evidence that the saliency map is located in the pulvinar. The evidence is suggestive but not conclusive.

There are several types of evidence that the saliency map is located in the pulvinar. The lateral pulvinar nucleus is interconnected with all of the areas in the occipitotemporal visual pathway and large parts of the posterior parietal cortex, so anatomically, it seems well situated for collecting attentional weight components and influencing cortical responses (Desimone, Wessinger, Thomas, & Schneider, 1990; Robinson & Cowie, 1997). Studies of humans with thalamic lesions involving the pulvinar have shown attentional impairments (e.g., Danziger, Ward, Owen, & Rafal, 2001; Karnath, Himmelbach, & Rorden, 2002; Rafal & Posner, 1987; Sapir, Rafal, & Henik, 2002), and brain-imaging studies have shown activation in the pulvinar in attentional tasks (e.g., LaBerge & Buchsbaum, 1990). More direct evidence stems from electrophysiological and pharmacological studies in monkeys. Petersen, Robinson, and Keys (1985) recorded from single cells in various parts of the pulvinar. Nearly all cells increased their rate of firing in response to visual stimuli, but few cells discriminated for stimulus features like orientation, direction of movement, color, or disparity. In a dorsomedial portion of the lateral pulvinar (Pdm) with a crude retinotopic organization, about half of the tested cells showed visual responses whose strength reflected the behavioral relevance of the stimulus in their RF (enhanced responses to targets as compared with distractors). The enhancement (attentional modulation) seemed unrelated to any specific motor response, but it was demonstrable with and without saccadic eye movements to the attended stimulus, consistent with the hypothesis that the levels of activation of the cells represented attentional weights.

By use of Posner's (1980) cuing paradigm, Petersen, Robinson, and Morris (1987) showed attentional effects of microinjecting bicuculline (a gamma-aminobutyric acid antagonist) or muscimol (a gamma-aminobutyric acid agonist) into Pdm (see also Desimone et al., 1990). The major effects conformed to the hypothesis that injection of bicuculline elevated attentional weights of stimuli in the visual field contralateral to the injection, whereas injection of muscimol depressed attentional weights of contralateral stimuli. Thus, when an invalid cue was presented in the normal hemifield (ipsilateral to the injection), simple reaction times to the imperative stimulus in the affected hemifield were speeded up by bicuculline but slowed down by muscimol. When an invalid cue was presented in the affected hemifield (contralateral to the injection), simple reaction times to the imperative stimulus in the normal hemifield were slowed down by bicuculline but speeded up by muscimol (see Bundesen, 1990, p. 532, for a TVA-based model of performance in the cuing paradigm of Posner, 1980; see also Bundesen, 1998b, pp. 305–307). Thus, the attentional weight of an object seemed to depend on the level of activation in the contralateral Pdm. The study suggests that the saliency map is anatomically lateralized such that attentional weights of objects in the left and right visual hemifields are represented in the contralateral pulvinars. This hypothesis is consistent with lesion studies in humans (e.g., Karnath et al., 2002).

[2] The activation of a neuron can be affected by an object *x* in its classical RF even though the neuron does not represent object *x*. This is likely to happen when the neuron represents a smaller object that is a proper part of *x* or a larger object of which *x* is a proper part.

Pdm is directly connected with areas 7ip and PO of the posterior parietal cortex (Robinson & Cowie, 1997). The posterior parietal cortex—and, in particular, area 7—has traditionally been associated with spatial attention (e.g., Bushnell, Goldberg, & Robinson, 1981; Posner, Walker, Friedrich, & Rafal, 1984; see also Bundesen, 1998b), and it seems plausible that attentional weight components resulting from spatial selection criteria are computed in the posterior parietal cortex and transmitted to the pulvinar, where they are added to weight components resulting from nonspatial selection criteria (cf. Figure 8).

Wave of Selective Processing

The attentional weights computed during the wave of unselective processing are used for redistribution of cortical processing capacity across the objects in the visual field such that the expected number of cells allocated to a particular object during the wave of selective processing reflects the attentional weight of the object. Specifically, as previously described in the *Dynamic Remapping of RFs* section, cortical processing capacity is redistributed in such a way that the probability that a neuron represents a given object within its classical RF during the wave of selective processing equals the attentional weight of the object divided by the sum of the attentional weights of all objects within its classical RF.

In our neural interpretation, Equation 1 of TVA describes the effect of the wave of selective processing at a level at which the classical RFs of neurons are so large that each one covers the entire visual field. Below, we show that for a feature i represented at this level, $v(x, i)$ is the expected value of the total activation of the population of neurons representing the categorization that "object x has feature i," assuming that the activation of the neurons is modulated by the perceptual decision bias β_i.

Let the feature-i neurons at the level in question be numbered 1, 2, . . ., $c(i)$. At any point in time, the effective RF of each neuron is assumed to be contracted such that the neuron represents the properties of only one of the objects in the visual field. As discussed in the *Dynamic Remapping of RFs* section, each object x has an attentional weight w_x, such that the probability that the neuron represents object x (i.e., the probability that the RF of the neuron is contracted around x) equals

$$w_x / \sum_{z \in S} w_z .$$

Let $\eta_k(x, i)$ be the activation of the kth ($1 \leq k \leq c[i]$) feature-i neuron when the neuron represents object x (i.e., the RF is contracted around x) and the perceptual decision bias in favor of feature i is maximal.[3] In the same terminology, $\eta_l(y, i)$ is the activation of the lth ($1 \leq l \leq c[i]$) feature-i neuron when the neuron represents object y and the featural bias in favor of i is maximal. If the featural bias in favor of i is less than maximal, the activations of the two neurons are assumed to be reduced by multiplication with the same factor β_i ($0 \leq \beta_i \leq 1$). Thus, β_i is a measure of the strength of the featural bias in favor of i. Given that the featural bias in favor of i equals β_i, the activation of the kth feature-i neuron equals $\eta_k(x, i)\ \beta_i$ when the neuron represents object x.

The total activation of the population of neurons that represent the categorization that "object x has feature i" equals the sum of the activations of those feature-i neurons that represent object x. The contribution to this sum from the kth feature-i neuron equals $\eta_k(x, i)\ \beta_i$ if the neuron represents object x, but it equals 0 if the neuron represents any other object. Because the probability that the neuron represents object x is

$$w_x / \sum_{z \in S} w_z ,$$

the expected contribution to the sum from the kth feature-i neuron equals

$$\eta_k(x, i)\ \beta_i w_x / \sum_{z \in S} w_z .$$

Hence, the expected value of the sum, $v(x, i)$, is given by

$$v(x,i) = \sum_{k=1}^{c(i)} \eta_k(x, i) \beta_i \frac{w_x}{\sum_{z \in S} w_z} . \tag{3}$$

By defining

$$\eta(x, i) = \sum_{k=1}^{c(i)} \eta_k(x, i) \tag{4}$$

(by analogy with Equation A6 in Appendix A), Equation 3 reduces to Equation 1 of TVA. By Equation 4, $\eta(x, i)$ equals the total activation of the set of all feature-i neurons when every feature-i neuron represents object x (say, x is the only object in the visual field) and the featural bias in favor of i is maximal (i.e., $\beta_i = 1$).

In the neural interpretation we have outlined, the filtering mechanism of selection affects the number of neurons in which an object x is represented (the number of neurons allocated to the object). Filtering is done by dynamic remapping of RFs. At a level of processing where the classical RFs of neurons are so large that each one covers the entire visual field, the number of neurons in which object x is represented becomes proportional to the relative attentional weight of object x. Whereas the filtering mechanism affects the number of neurons in which an object x is represented, the pigeonholing mechanism of selection affects the way in which the object is represented in those neurons that are allocated to the object. For each feature-i neuron whose activation is modulated by featural bias, the activation becomes proportional to β_i. Thus, Equation 1 of TVA describes the combined effects of filtering and pigeonholing on the total activation of the population of neurons representing the categorization that "object x has feature i."

VSTM

Feedback Loops

As described in the *Dynamic Remapping of RFs* section, the effective RF of a cortical neuron in the visual system is coded by

[3] For a feature-i neuron whose activation is independent of the perceptual decision bias in favor of feature i, $\eta_k(x, i)$ is simply defined as the activation when the neuron represents object x. For a feature-i neuron whose activation depends on the bias in favor of feature i, $\eta_k(x, i)$ is defined as the activation when the neuron represents object x and the bias in favor of feature i is maximal (i.e., $\beta_i = 1$). The activations of the feature-i neurons considered in this section are assumed to be modulated by the bias in favor of feature i.

the way in which the gates that control the flow of information from the retina to the cortical neuron are set. As shown in the *Topographic Maps of Objects* section, the code formed by the gate settings can be used for directing signals from a cortical cell whose effective RF is contracted around an object at a certain location to a corresponding place in a topographic map of objects. Just as attentional weight components (of the form $\eta_k[x, j]\ \pi_j'$) computed during the wave of unselective processing are assumed to be routed to the correct places in a saliency map, we assume that components of v values (of the form $\eta_k[x, i]\ \beta_i$) computed during the wave of selective processing are routed to the correct (topographically corresponding) places in a VSTM map.

What does it mean that a visual categorization of a certain object enters VSTM? Following Hebb (1949), among many others, we assume it implies that the activation of the population of neurons representing the categorization is sustained by being incorporated into a feedback loop. Specifically, a visual categorization is encoded in VSTM if, and only if, the categorization is embedded in a positive feedback loop gated by a unit in the VSTM map of objects. Two feedback loops of this type are illustrated in Figure 9 (for a closely related proposal, see Usher & Cohen, 1999). As shown in the figure, impulses routed to a unit that represents an object at a certain location in the topographic VSTM map of objects are fed back to the feature units from which they originated, provided that the VSTM unit is activated. If the VSTM unit is and remains inactive, impulses to the unit are not fed back. Thus, for each feature-i neuron representing object x, activation of the feature-i neuron is sustained by feedback if, and only if, the unit representing object x in the topographic VSTM map of objects is activated.

Anatomical Location

The VSTM map of objects in the visual field may be located in the PF cortex, where regions implicated in VSTM have been tentatively identified by functional MRI (fMRI) in posterior parts of the superior frontal sulcus and the middle and inferior frontal gyri (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; for further evidence supporting this possibility, see Bundesen, Larsen, Kyllingsbæk, Paulson, & Law, 2002; see also Smith & Jonides, 1997). A different possibility is that the VSTM map of objects is located in the thalamic reticular nucleus (TRN), which forms a major component in the communication between the cortex and the thalamic nuclei, including the lateral geniculate nucleus (LGN; see Crick, 1984, for an early proposal concerning attentional functions of the TRN; see O'Connor, Fukui, Pinsk, & Kastner, 2002, for evidence of attentional effects in the LGN).

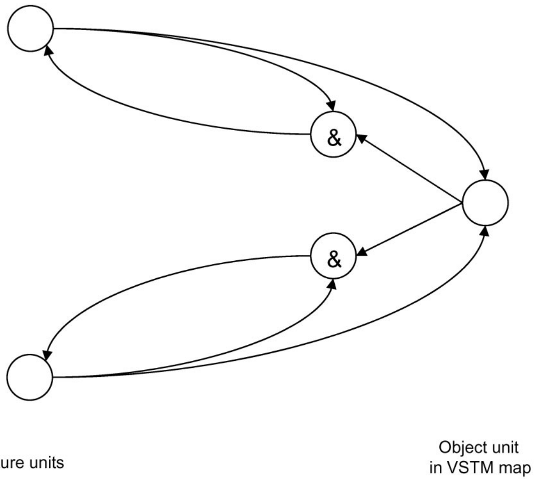

Figure 9. Two feedback loops gated by the same unit in the visual short-term memory (VSTM) map of objects. Activation of either feature unit is sustained by positive feedback if, and only if, the object unit is activated. & = logical AND unit.

The TRN lies like a thin shield between the thalamus and the cortex, so all fibers passing between the thalamus and cortex go through the TRN. Many of the fibers that go through the TRN have side branches (collaterals) with excitatory synapses to cells in the TRN. Cells in TRN are interconnected, and they also send axons to thalamocortical relay cells in the LGN. Like LGN and the visual cortex, the visual sector of TRN contains at least one topographically ordered representation of the visual field (see, e.g., Guillery, Feig, & Lozsádi, 1998), so it seems to be capable of focusing on limited parts of the visual field.[4]

Figure 10 shows how the circuits illustrated in Figure 9 might be implemented as thalamocortical feedback loops. In Figure 10, feature i is a shape feature of an object x, and feature j is a motion feature of the same object x. Feature-i neurons are found in a high-level cortical area in the ventral stream of processing (cf. Milner & Goodale, 1995; Ungerleider & Mishkin, 1982), where their rates of activation are modulated by multiplication with the perceptual decision bias β_i. Feature-j neurons are found in a high-level cortical area in the dorsal stream of processing, where their rates of activation are modulated by multiplication with the perceptual decision bias β_j. The rates of activation of both the feature-i and the feature-j neurons allocated to object x are projected back to the thalamus. The top-down impulses representing the categorization that "object x has feature i" are routed to some of those cells in LGN whose bottom-up activation supported this categorization, and the top-down impulses representing the categorization that "object x has feature j" are routed to some of those cells in LGN whose bottom-up activation supported that categorization. In either case, the feedback occurs through an AND unit, which tentatively is located in the LGN near the cells that are targeted by the feedback. For the feedback to be effective, the AND gates must be open, which means that the AND units must

[4] All cells in the TRN are inhibitory, but there are reasons to believe that TRN serves to support representations in LGN of selected objects or locations. In a model proposed by Sherman and Guillery (2001, Figures 8B and 8D, p. 75), a cortical neuron (Cell A) that provides excitatory input to a topographically corresponding thalamocortical relay cell in the LGN (Cell B) also sends excitatory collaterals to cells in TRN. The targeted cells in TRN are connected to Cell B in such a way that the net effect of the connection from Cell A to Cell B via TRN is excitatory. (The activation of the targeted cells in TRN inhibits neurons in LGN that excite inhibitory TRN neurons connected to Cell B, resulting in *feedforward disinhibition.*) In a related model, the same result (feedforward disinhibition) is obtained by allowing cells in TRN that are excited by Cell A to be connected to Cell B via a local interneuron in the TRN. In either model, the collaterals from the cortical Cell A into TRN should support activation of the topographically corresponding Cell B in LGN.

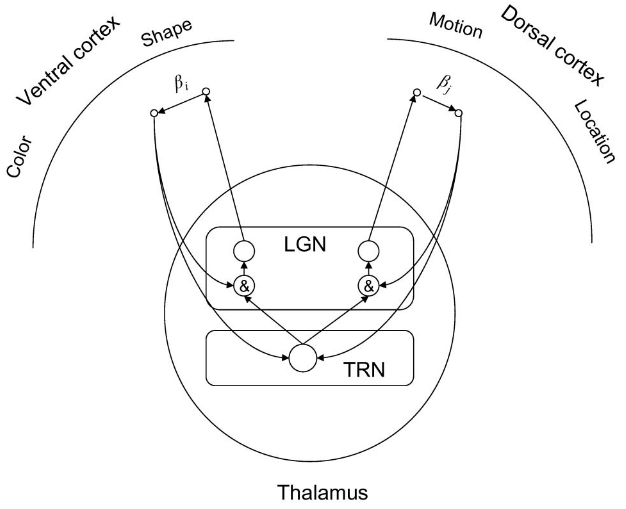

Figure 10. Possible implementation of the circuits illustrated in Figure 9 as thalamocortical feedback loops. The feature units are located in different cortical areas, where η values (strengths of evidence that objects at particular scales and positions have particular features) are computed. The η values are multiplied by β (bias) values, and the products are projected back to the thalamus as rates of activation. The visual short-term memory (VSTM) map of objects is located in the thalamic reticular nucleus (TRN). When the VSTM map is initialized, objects in the visual field effectively start a race to become encoded into VSTM. In this race, each object is represented by all possible categorizations of the object, and each possible categorization participates with a firing rate proportional to the product of the corresponding η and β values. For the winners of the race, the TRN gates activation representing a particular categorization back to some of those cells in lateral geniculate nucleus (LGN) whose activation supported the categorization. Thus, activity in neurons representing winners of the race is sustained by positive feedback. & = logical AND unit.

receive input from a unit representing object *x* in the topographic VSTM map of objects found in the TRN. Effectively, the feedback loops with information about features of object *x* are complete if the TRN unit representing object *x* in the VSTM map of objects is active. An obvious way of activating the TRN unit is illustrated in the figure: Both impulses representing the categorization that “object *x* has feature *i*” and impulses representing the categorization that “object *x* has feature *j*” are projected back, not only whence they came in the LGN but also to the unit representing object *x* in the TRN. If, for any reason, the TRN unit representing object *x* is and remains inactive, impulses from the cortex representing object *x* are not transmitted beyond the AND units in the LGN.

Storage Capacity

In TVA, VSTM can at most hold *K* objects, and space is available for a categorization of object *x* if object *x* is already represented in the store or if fewer than *K* objects are represented in the store. This strong limitation on storage capacity may be implemented through wide-ranging inhibitory connections in the VSTM map of objects. Below, we describe a neural network model of a short-term store with the required type of behavior.

Our neural network model of VSTM is a multiscale map of locations in the visual field. Each unit in the VSTM map has a number of input lines from the environment (different cortical areas). A signal on a particular input line to the unit represents a particular perceptual categorization of an object at the location (spatial position and size) corresponding to the unit; different types of perceptual categorizations are represented by different input lines to the unit. As in the WTA cluster illustrated in Figure 3, each unit excites itself and inhibits all other units in the network. Again, as in the WTA cluster, we suppose that when the network is initialized, a signal of a certain strength (e.g., *r* spikes) from the environment to one of the units is sufficient to trigger this unit, and once the unit has been triggered, it keeps on firing because of its self-excitation. Now we generalize the specification of the WTA cluster by claiming that when fewer than *K* units are firing at the same time, an above-threshold signal from the environment to one of the inactive units is sufficient to trigger this unit; but once *K* units are firing, they inhibit the other units in the network so strongly that these other units cannot be triggered by signals from the environment. Clearly, given these assumptions, *K* objects at most can be represented (by activated units) in the network. Each unit of the network (i.e., each object unit in the VSTM map) is used as illustrated in Figure 9, so impulses from the environment (feature units in different cortical areas) representing categorizations of object *x* are retained in VSTM (in reverberating circuits) if, and only if, a unit representing object *x* in the VSTM map is active (such that the feedback loops are complete). Hence, space is available for a categorization of object *x* if object *x* is already represented in the store, or if fewer than *K* objects are represented in the store, which was to be shown.

Race for Encoding

In the network model we have described, VSTM can be cleared (initialized) by deactivation of all units in the *K*-winners-take-all cluster that serves as the VSTM map of objects. Clearing of VSTM effectively starts a new race among objects to become encoded into VSTM (i.e., to [re]activate units representing the objects in the VSTM map). An object is encoded in VSTM if and when any categorization of the object is encoded in VSTM, so each object *x* is represented in the race by all possible categorizations of the object. Assuming the baseline rates to be negligibly small, each categorization of the form “*x* has feature *i*” participates in the race with a Poisson firing rate equal to $v(x, i)$, so the object effectively participates in the race with a Poisson firing rate, v_x, equal to the sum of the *v* values of all possible visual categorizations of the object,

$$\sum_{i \in R} v(x, i).$$[5]

[5] The Poisson rate given by a certain *v* value is the mean number of spikes per time unit. In a (homogeneous) Poisson process with rate parameter *v*, interspike times are independent, exponentially distributed random variables with the same rate parameter, *v*. A set of *n* parallel Poisson processes with rate parameters v_1, v_2, . . ., and v_n, respectively, is itself a Poisson process with rate parameter $v_1 + v_2 + \ldots + v_n$. Correspondingly, the minimum of *n* mutually independent, exponentially distributed random variables with rate parameters v_1, v_2, . . ., and v_n, respectively, is itself exponentially distributed with rate parameter $v_1 + v_2 + \ldots + v_n$.

Suppose that the first spike arriving at a given unit in the VSTM map of objects after the map has been cleared suffices to activate the unit (until K units have been activated). In this case, selection of objects for encoding can be described by a simple independent, exponential race model—a model assuming that the selection is determined by a parallel-processing race in which processing times for individual objects are mutually independent, exponentially distributed random variables. This model implies that the waiting time from the start of the race (the clearing of VSTM) to the activation of the first unit representing an object in the VSTM map is exponentially distributed with rate parameter

$$\sum_{x \in S} \sum_{i \in R} v(x, i).$$

Similarly, the waiting time from the activation of the first to the activation of the second unit is exponentially distributed with a rate parameter equal to the sum of the rate parameters for the remaining objects, and so on until K objects have been selected. Furthermore, throughout the race for encoding, the probability that any not-yet-encoded object x is the next one to be encoded equals v_x divided by the sum of the v values of all not-yet-encoded objects (choice rule of Luce, 1959).

If more than one spike is needed for activation of a unit in the VSTM map of objects, more complex race models will be needed to make exact predictions. For example, suppose that the units in the VSTM map of objects are accumulators (counters), each with a threshold set at r spikes ($r > 1$) such that r spikes must arrive at a unit to activate the unit after the VSTM map has been initialized. If this is so, then encoding of objects into VSTM can be described by an independent gamma race model—a model in which processing times for individual objects are mutually independent, gamma-distributed random variables with shape parameter r and rate parameters given by their v values (see Bundesen, 1987; Logan, 1996). For another example, suppose that the counting race goes on until one of the units has accumulated r more spikes than any of the other units. In this case, encoding of objects into VSTM can be described by a random-walk model with exponentially distributed interstep times (cf. Logan, 2002; Logan & Gordon, 2001).

Mental Images

Representations in VSTM are visual mental images. The representations can be more or less abstract. They can be bottom-up generated (from visual sense impressions). This occurs when concrete visual features of an object are encoded into VSTM by the processes described in the previous sections. Mental images can also be top-down generated (from long-term memory). This often occurs when a subject anticipates or prepares him-, her-, or itself for seeing a particular stimulus (a stimulus that matches the mental image). The mental image can be highly schematic. For example, if the set of possible display objects in an experiment is known (e.g., small quadrangular objects), but there is uncertainty about which particular object will appear next, only a highly schematic anticipatory mental image may be warranted.

Whether they are bottom-up or top-down generated, visual mental images are neural representations that are similar to visual sense impressions. Following Hume (1739/1896), among many others, we regard simple mental images as *faint copies* of sense impressions (and complex mental images as compounds of simple ones). A suggested implication is that mental images are positioned in the (subjective) visual field, although localization may be less precise than it is with real sense impressions. Neurophysiologically, we assume, a simple visual mental image consists in activity in a subset of all those neurons whose activity would be implicated in the corresponding visual sense impression. However, the responses of the activated neurons are weaker than they are when driven by the corresponding real stimulus.

With no external stimulus in the RF of a neuron, the firing rate, by definition, is at baseline. However, we assume that the baseline rate is a sum of two components: *inner-driven* activity and *undriven* activity. Inner-driven activity carries information about imagined objects. Undriven activity can be regarded as a response to internal random noise. Unlike stimulus-driven and inner-driven activity, undriven activity seems to be independent of a subject's state of attention (cf. McAdams & Maunsell, 1999a).

NTVA assumes that mental images can have various effects on the neural processing of upcoming stimuli. Obviously, maintaining a mental image of a target is a way of retaining the object so that it can be compared against other stimuli, encoded into long-term memory, and the like. Also, while the image of the target is being retained, the visual system may be tuned to search for the target by setting appropriate pertinence values (π values), and the system may be primed for perceiving the target again by the setting of appropriate categorization biases (β values). However, in NTVA, neural activity corresponding to a mental image does not directly bias competition between multiple stimuli in a visual cell's RF (cf. Desimone, 1999); in NTVA, input selection is controlled via the saliency map.

Perceptual Decision Making

In NTVA, attentional selection of an object x consists in encoding a categorization of the form "object x belongs to category i" or, equivalently, "object x has feature i" into VSTM. Feature i can be a raw sensory feature (e.g., a particular brightness or orientation) or a visual pattern stamped in by learning (e.g., letter type A). In either case, as described in the *VSTM* section, the encoding is done by establishing a feedback loop such that cortical neurons representing the selected categorization are kept active.

Visual identification of object x consists in making perceptual categorizations of x. Encoding a categorization into VSTM (i.e., attentional selection) is one way of making the categorization (viz., making the categorization by *immediate perception*). However, mutually contradictory categorizations can be made by immediate perception (i.e., encoded in VSTM), and decisional procedures for resolving contradictions (procedures for making categorizations by *mediate perception*) are needed. In Appendix B, we describe three types of procedures for mediate perception: an exponential race procedure (cf. Bundesen, 1990), a Poisson counter procedure (cf. Logan, 1996), and a Poisson random-walk procedure (cf. Logan, 2002; Logan & Gordon, 2001). We also analyze how decision making by these procedures is improved when the number of neurons used to represent the object to be identified is increased by attentional filtering.

Relations to Other Theories

The relationship between TVA and other formal theories of visual attention (e.g., Shih & Sperling, 2002; Sperling & Weichselgartner, 1995) has recently been analyzed by Logan (2004; see also Bundesen, 1996). Here, we consider the relationship between NTVA and other neural theories of visual attention.

Time and Locus of Selection

In *early-selection* theories of attention, attentional selection takes place before perceptual identification (e.g., Broadbent, 1958, 1982; Kahneman, 1973; LaBerge & Brown, 1989; Moray, 1969; Neisser, 1967; Rumelhart, 1970; Treisman, 1964a, 1964b; Treisman & Gelade, 1980). In *late-selection* theories, attentional selection takes place only after identification (e.g., Allport, 1977; Deutsch & Deutsch, 1963; Duncan, 1980; Duncan & Humphreys, 1989; Hoffman, 1978; Humphreys & Müller, 1993; Keele, 1973; Norman, 1968; Posner, 1978; Shiffrin & Schneider, 1977; van der Heijden, 1981; see Usher & Niebur, 1996, for an interesting neural interpretation of the late-selection theory of Duncan, 1980; see Bundesen & Habekost, 2005, for a general review). In NTVA, both immediate visual identification and attentional selection consist in encoding visual categorizations into short-term memory (VSTM), so immediate visual identification and attentional selection occur at the same time (*simultaneous selection*; Logan, 2002).

Typical early- and late-selection theories of attention differ in their assumptions concerning not only the time but also the anatomical locus of selection. In typical early-selection theories, selection occurs at a low (*early*) level of the visual system. In typical late-selection theories, selection occurs at a high (*late*) level of the system (see van der Heijden, 1981, for a notable exception). In NTVA, the wave of unselective processing occurs early in time at both low and relatively high anatomical levels of the visual system. The wave of selective processing occurs late in time at all levels of the visual system. Thus, to a wide extent, the waves of unselective (*preattentive*) and selective (*attentive*) processing occur at the same anatomical loci and involve the same cortical machinery.

Processing Capacity

In NTVA, visual processing capacity is limited because the number of cells in the visual system is limited (see van der Heijden, 1992, chap. 8, for pertinent references and a critical review; see also Palmer, Verghese, & Pavel, 2000). However, the limited processing capacity is distributed across stimulus objects such that more processing resources (neurons) are devoted to behaviorally important stimuli than to unimportant ones. The distribution is implemented by dynamic remapping of RFs of cortical neurons. These ideas were inspired by the findings of Moran and Desimone (1985). The idea of dynamic remapping of RFs is also found in the shifter circuit model of Anderson and Van Essen (1987); the dynamic routing circuit model of visual attention and recognition by Olshausen et al. (1993); and the selective attention for identification model of Heinke and Humphreys (2003), which uses a spatial window to select visual information for recognition, binding parts to objects and generating translation-invariant recognition (cf. Larsen & Bundesen, 1978, 1998).

Serial Versus Parallel Processing

In feature integration theory (Treisman & Gelade, 1980), the guided search model (Cave & Wolfe, 1990; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989), and other serial processing models (see Bundesen, 1996, for a review), only one object can be attended at any one time. By contrast, NTVA is a parallel-processing model. Like Reynolds, Chelazzi, and Desimone (1999), NTVA assumes parallel processing of simultaneously presented stimuli not only at low levels of the visual system but also at high levels. However, unlike Reynolds et al. (1999), NTVA assumes that at any given point in time, an individual cell with multiple stimulus objects within its classical RF is driven by only one of these objects, so separate objects are represented in separate sets of cells (cf. Rousselet, Thorpe, & Fabre-Thorpe, 2003).

Mechanisms of Selection

The theoretical account of Reynolds et al. (1999) builds on the biased-competition model of Desimone and Duncan (1995; see also Desimone, 1999). In the biased-competition model, a typical neural network for attentional selection works as follows: Connections between units in the network are arranged such that (a) units representing mutually compatible categorizations of the same object facilitate each other, but (b) units representing incompatible categorizations inhibit each other, and (c) units representing categorizations of different objects also inhibit each other. Search for a red target, for example, can then be done by preactivating units representing redness. If a red target is present, the preactivation will directly facilitate the correct categorization of the target with respect to color. Indirectly, the preactivation will facilitate categorizations of the target with respect to properties other than color, but it will inhibit categorizations of any objects other than the target (*integrated competition*; Duncan, 1996; see also Phaf, van der Heijden, & Hudson, 1990; cf. TVA's assumption of substantial independence between different visual categorizations of the same object). The theoretical emphasis on bias (preactivation) is reminiscent of TVA, but the biased-competition model has no distinction between pertinence parameters (controlling filtering) and bias parameters (controlling pigeonholing). Bundesen (1990), Logan (1996), and van der Heijden (1992, 2004) have emphasized the distinction between pertinence and bias, and NTVA elaborates the nature of the distinction by assuming that filtering (pertinence) changes the number of cells representing a stimulus, whereas pigeonholing (bias) changes the level of activation in cells coding for particular features.

It is interesting to compare the mechanisms of selection proposed in NTVA with the mechanisms proposed in a series of related connectionist models. Mozer (1991) and Mozer and Sitton (1998) proposed an influential connectionist model of object recognition and attentional selection (MORSEL) with an attentional mechanism (AM) for filtering (see also Mozer, 2002). The AM is a saliency map consisting of processing units in one-to-one correspondence with the locations in a topographic map of feature detectors. Each unit in the AM receives bottom-up input from feature detectors at the corresponding location. The activation of the AM unit indicates the saliency of the corresponding spatial location and serves to gate the flow of activation from feature detectors at the given location into a recognition network. The gating is done by multiplying (a) the activation flowing from the feature detectors into the recognition network by (b) the activation of the AM unit. Thus, filtering is done by spatial selection, and the spatial selection is performed by scaling the output from individual feature detectors rather than varying the number of units in which an object is represented. In a more abstract model proposed by Cohen, Romero, Farah, and Servan-Schreiber (1994), and in models developed by O'Reilly (see O'Reilly & Munakata, 2000, chap. 8), selection of objects (filtering) emerges from units representing spatial locations interacting with units in the object recognition network. In the complex attentional model presented by O'Reilly and Munakata (2000), activation of spatial representations in the dorsal cortical pathway enhances the levels of activation of representations of objects at the corresponding locations in lower level visual areas V1 and V2 rather than opening and closing gates controlling the communication from lower to higher levels of the visual system. Thus, in contrast to NTVA, filtering changes the activation of individual cells in which an object is represented rather than changing only the number of cells in which the object is represented.

Applications to Single-Cell Studies of Attentional Effects on Visual Representations

NTVA interprets attentional selection at the level of individual neurons. The interpretation was suggested by electrophysiological findings from single-cell studies in monkeys, and NTVA accounts for many findings from such studies. In this section, we review the major findings from single-cell studies of attentional effects on visual representations and interpret these findings in terms of NTVA. We focus on the *ventral stream* of visual processing (Ungerleider & Mishkin, 1982). The ventral stream includes areas V1, V2, V4, and IT and is central to categorization and object recognition in the primate visual system. The review also includes a few studies of the *dorsal stream* of visual processing (e.g., MT, MST), where attentional effects seem to be similar.

Two decades of single-cell studies have revealed several distinct types of attentional effects in the ventral stream. By far the strongest changes of a cell's firing rate occur when multiple objects are present in the classical RF. Under these conditions, a general finding is that attention to one of the objects modulates the firing rate either up or down, depending on the cell's preference for the attended object. As detailed in the *Attentional Effects With Multiple Stimuli in the RF* section, the finding is readily explained by NTVA's notion of filtering on the basis of attentional weights. A second typical effect of attention is a modest modulation of firing rates with a single stimulus in the RF. In the *Attentional Effects With a Single Stimulus in the RF* section, this finding is explained by pigeonholing or by the presence of stimuli other than the one defined by the experimenter. Even when an experiment is designed to include only one stimulus in the RF, the cell may respond to an individual part of the experimental stimulus instead of the whole stimulus, or the cell may respond to internally generated random noise. Effects of individual parts of an experimental stimulus should mainly be found in experiments with complex stimuli. Effects of internally generated random noise should mainly appear in experiments with faint stimuli. A third common effect of attention is an increase in a cell's baseline firing rate when a target is expected to appear in its RF. In the *Attentional Effects on Baseline Firing* section, this effect is explained as the neural correlate of a more or less schematic mental image of the anticipated stimulus (a representation in VSTM).

Attentional Effects With Multiple Stimuli in the RF

If several stimuli are presented simultaneously within the RF of a cell, NTVA predicts that the firing rate of the cell comes to be determined by just one of the stimuli. The probability that the cell's effective RF becomes adjusted to a particular stimulus (which may, of course, be a group of other stimuli) depends on the activation pattern of the saliency map. If the organism is precued about which location to attend (i.e., the location is ascribed pertinence before the experimental stimuli are presented), the saliency map may be configured before stimulus exposure. In this case, selective processing of targets may be seen shortly after the presentation of the stimuli (see the *Filtering by Location* section below). However, if the organism is precued about target-defining features but not the location of the target, a substantial period of time occurs after the presentation of the stimulus before the saliency map is configured. In this case, unselective processing typically proceeds until 150–200 ms after display onset; a clear demonstration of the first wave of processing in NTVA (see the *Filtering by Nonspatial Categories* section below). Finally, by Equation 2, the attentional weight of an object depends both on strengths of sensory evidence and on pertinence settings. The dependence on strengths of sensory evidence explains why the probability of selecting a particular stimulus in the RF depends on the luminance contrast of the stimulus (see the *Filtering of Stimuli With Different Contrast* section below).

Filtering by Location

In typical experiments that require filtering by location, a monkey is given a prestimulus cue that directs attention to one among several possible locations. To measure effects of directing attention to one object in the RF rather than another, experimenters choose stimuli that will elicit clearly different firing rates from the neuron when shown in isolation (*ineffective* and *effective* stimuli, respectively). In the studies reviewed in this section, the monkey's task was either (a) to *match to sample* (i.e., to encode, retain, and compare stimuli appearing at the cued location; Moran & Desimone, 1985) or (b) to monitor a sequence of displays for the appearance of a target at the cued location (Luck, Chelazzi, Hillyard, & Desimone, 1997; Reynolds et al., 1999). The findings from the two paradigms are similar.

Moran and Desimone (1985). NTVA's notion of attentional filtering was originally inspired by a study by Moran and Desimone (1985). These authors presented macaque monkeys with stimuli at two locations. The monkeys were trained to attend only to stimuli presented at one of the locations in the display and to ignore any stimuli shown at the other location. The monkeys performed a match-to-sample task, in which they encoded a sample stimulus that was shown at the attended location, retained it for a delay period, and then matched it to a test stimulus shown at the same location. During the presentation of both sample and test displays, Moran and Desimone recorded the responses of single neurons to the stimuli. They found that when the target and distractor stimuli were both within the RF of a cell in visual areas

V4 or IT, the rate of firing in the cell showed little effect of the distractor. For example, they recorded the response of a cell to a pair of stimuli consisting of (a) a stimulus that elicited a high rate of firing in the cell when the stimulus was presented alone (an effective sensory stimulus) and (b) a stimulus that had little or no effect on the rate of firing in the cell when the stimulus was presented alone (an ineffective sensory stimulus). On trials in which the effective sensory stimulus was the target and the ineffective sensory stimulus was the distractor, the cell showed a high rate of firing. However, on trials in which the ineffective sensory stimulus was the target and the effective sensory stimulus was a distractor, the cell showed a low rate of firing. Moran and Desimone (1985) remarked that the typical cell responded "as if the RF had contracted around the attended stimulus" (p. 783).

The effect depended on both the target and the distractor being located within the recorded neuron's RF. However, in IT, the RFs were very large, covering at least the central 12° of both the contralateral and ipsilateral fields. For these neurons, the distractors were always located inside the RF. In contrast, in V1, only one stimulus could be fitted into each neuron's RF. In this case, no significant effects of attention were found. Also, in V4, no effect of attention was found when only a single stimulus was placed in the RF. These negative findings fit with the predictions of NTVA. Overall, the results of Moran and Desimone (1985) clearly support NTVA's notion of filtering, and in fact, these results suggested the interpretation in the first place. The effect was replicated by Luck et al. (1997; see the *Attentional Effects on Baseline Firing* section below), who displayed two stimuli within the RF of V4 neurons. Similar to the findings of Moran and Desimone, Luck et al.'s results showed that the typical neuron tended toward responding to the stimulus in the attended location while being relatively unaffected by the other stimulus. Extending Moran and Desimone's findings, Luck et al. also found this effect in some V2 neurons that had RFs large enough to encompass two stimuli at the same time.

Whereas these two studies demonstrate the existence of a competitive mechanism with more than one stimulus present within the RF, the data do not allow for testing of a specific hypothesis about the effect: NTVA implies that the expected response to a pair of stimuli within the RF is given by a weighted average of the responses to each of the stimuli presented alone. A more recent study has supplied the information needed to test this hypothesis.

Reynolds, Chelazzi, and Desimone (1999). Reynolds et al. (1999) recorded from single neurons in areas V2 and V4. In their first experiment, they attempted to characterize the response to a pair of stimuli in the RF when attention was not involved (i.e., the monkey was passively fixating a point outside the RF). The stimuli were drawn randomly from a set of 16 bar stimuli, encompassing all possible combinations of four orientations and four colors. In this manner, the set covered a range of selectivity, from ineffective to effective stimuli. The stimuli could appear at two locations in the RF. In each recording session, a *reference* stimulus was selected to be shown at one of the locations. In one condition of the experiment, the reference stimulus was shown in isolation. In another condition, the reference stimulus was accompanied by 1 of the 16 bar stimuli (a *probe*) at the second location. In the third condition, the probe stimulus was shown alone. Reynolds et al. (1999) systematically varied the combinations of reference and probe stimuli. The results showed that the mean firing rate of a cell to a pair of stimuli in the RF approximated a weighted average of the firing rates to each of the stimuli in the pair when presented alone. For example, for one cell tested with the 16 different probes, the pair response equalled approximately 67% of the response to the probe plus 33% of the response to the reference stimulus. Other cells attached more weight to the reference stimulus than to the probe. The degree of influence exerted by a given stimulus on a certain cell (the weighting factor) showed little dependence on the magnitude of the response elicited when the stimulus was presented alone.

In their second experiment, Reynolds et al. (1999) studied how attention to one of the objects in a pair modulated this basic pattern of responses. The monkey's task was to monitor a target location in a sequence of displays, reacting when a diamond-shaped object was shown (see Figure 11). Stimuli could also appear at other locations, but this was irrelevant to the task. Before the target appeared, the monkey was presented with from one to six brief displays, each of which contained either one or two distractors in the RF. The distractors were drawn from the same set of oriented bars that was used in the first experiment and were also arranged in terms of reference and probe stimuli. The responses to these stimuli, not the response to the less frequent target displays, formed the basis of the analysis.

Stimuli could appear at four locations: Two within the RF and two outside. In the *attend-away* condition of the experiment, the monkey attended to one of the two locations outside the RF, but in the *attend-RF-stimulus* condition, attention was directed to one of the locations inside the RF. The attend-away condition was essentially a replication of the first experiment. The (pretarget) displays in this condition consisted of (a) a reference stimulus, (b) a probe stimulus, or (c) both a reference and a probe stimulus in the (unattended) RF. Reynolds et al. (1999) found the same basic pattern as in the first experiment, with the response to a pair being a weighted average of the responses to the individual stimuli. The weight of each stimulus (reference or probe) in the pair was about

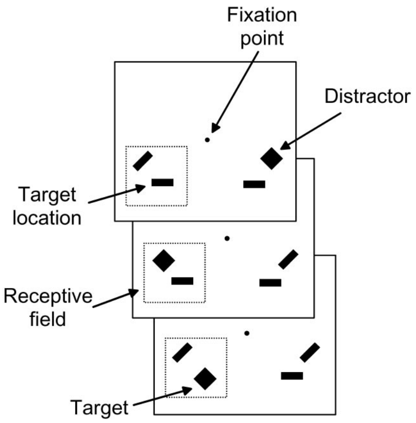

Figure 11. Structure of a trial in Experiment 2 of Reynolds, Chelazzi, and Desimone (1999). A monkey's task was to monitor a target location in a sequence of displays, responding when a diamond-shaped target object was shown. In the illustrated trial, the target location was the lower one of the two possible locations inside the receptive field (dotted squares) of the recorded neuron. From "Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4," by J. H. Reynolds, L. Chelazzi, and R. Desimone, 1999, *Journal of Neuroscience, 19,* Figure 1C, p. 1738. Copyright 1999 by the Society for Neuroscience. Adapted with permission.

equal on average, consistent with the fact that no selection was required in this part of the visual field.

In the attend-RF-stimulus condition, Reynolds et al. (1999) were able to study the effect of attention on the pair response. When the monkey's attention was directed to the location of one of the stimuli in the pair, this increased the weight on the stimulus in the target location such that the mean response of the neuron was driven (up or down) toward the response elicited when the stimulus was presented alone. This pattern was found in both V2 and V4, although the effect of attention was somewhat stronger in V4. In V2, when attention was drawn to the probe stimulus, the response was a weighted average approximating 69% of the response to the probe plus 31% of the response to the reference stimulus. In V4, the same weight estimates were 83% and 17%, respectively. With attention to the reference stimulus, the strength of the weight relation was reversed, such that the pair response was given by 76% (in V2) or 79% (in V4) of the response to the reference stimulus plus 24% (in V2) or 21% (in V4) of the response to the probe.

The results of Reynolds et al. (1999) agree fairly well with the dual conjectures that (a) at any time, a typical cell is driven by one and only one of the stimuli in its RF, and (b) the probability that the cell is driven by any given stimulus (the probability that the cell represents the stimulus) is proportional to the attentional weight of the stimulus. This is exactly what is assumed in NTVA. However, Reynolds et al. (1999) proposed a different model of the weighting phenomenon, one that also seems consistent with the data. They assumed that pair responses of the cell at any given moment reflected both stimuli. Thus, in their account, the RF was not alternately monopolized by one of the stimuli. Whereas the published data in themselves give no grounds for choosing between these two hypotheses, the merits of each can be discussed on theoretical grounds. The mechanism proposed by Reynolds et al. (1999) is capable of generating the observed firing rates. However, it is not at all clear how a recognition system could use this information with several objects entangled in the neuron's representation. In this respect, we think that NTVA provides a more plausible explanation of the observed results. Here, the properties of different objects are separated by the filtering mechanism.

Another important finding by Reynolds et al. (1999) was an increase in the baseline firing rate (before the stimulus was presented) of many neurons when attention was directed to a location within the RF. This finding is discussed in the *Attentional Effects on Baseline Firing* section.

Filtering by Nonspatial Categories

In typical experiments requiring filtering by a nonspatial category, a monkey is cued to attend to a target stimulus defined by one or more features, but it is not given prior information about the location of the target. Accordingly, the activation in the saliency map must be configured after the onset of the display, resulting in a long initial period of unselective processing. Subsequently, when attentional weights have been computed and applied, selective processing takes effect. This two-stage mechanism seems clearly to be demonstrated in studies by Chelazzi, Duncan, Miller, and Desimone (1998) and Chelazzi, Miller, Duncan, and Desimone (2001).

Chelazzi, Duncan, Miller, and Desimone (1998). Chelazzi et al. (1998; see also Chelazzi, Miller, Duncan, & Desimone, 1993) studied the response of single neurons in anterior IT cortex during visual search. In the basic task, a cue stimulus was first presented at fixation. The cue showed the target to be searched for. Following the cue, the computer screen went blank for 1,500 ms while the monkey maintained central fixation. After this delay, two extrafoveal stimuli were displayed. If one of them matched the cue, the monkey was rewarded for making a saccadic eye movement toward that stimulus. If the target was absent from the display, the monkey was to just maintain fixation until a third, matching stimulus appeared. The two stimuli in the search display were both located within the large RF of the IT neuron so that both could influence the response. All stimuli were drawn from a set consisting of a *good*, a *neutral*, and a *poor* stimulus. When presented alone, the good stimulus and the poor stimulus elicited strong and weak responses, respectively.

When a target was present in the search display, and the display was presented in the contralateral visual field, Chelazzi et al. (1998) found strong effects of attention. The effects had a characteristic time course. During the first 150–200 ms after stimulus onset, the IT neuron responded in the same manner regardless of whether the good or the poor stimulus was the target. In terms of NTVA, this corresponds to the wave of unselective processing before attentional weights have been computed and applied. After this initial activation, a systematic change occurred in the firing pattern of the neuron. The response was rapidly driven toward the firing rate seen when the target stimulus was presented in isolation—either up (in case of the good stimulus) or down (in case of the poor stimulus). This is compatible with the hypothesis that a large majority of the recorded IT neurons contracted their RFs around the attended stimulus following the application of attentional weights. The effect of attention only occurred with competing stimuli in the RF. In a variation of the experiment in which just a single stimulus was shown in the search display, the neuron's response was almost unaffected by whether the stimulus equalled the cued target or not. This effect was also replicated with a different response (manual lever release), suggesting that the mechanism was not specific to the motor response but reflected basic perceptual processing.

When the search display was presented with the two stimuli on opposite sides of the vertical meridian, the response was dominated by the contralateral stimulus regardless of whether the contralateral stimulus was the target, the ipsilateral stimulus was the target, or neither stimulus was a target (see Sato, 1988, 1989, for related findings). When the stimuli were presented one at a time, responses to contralateral stimuli were no stronger than responses to ipsilateral stimuli, so the effect appeared to be attentional rather than sensory. In terms of NTVA, the results suggest that if the saliency map is located in the pulvinar, then the IT cortex receives stronger attentional weight signals from the ipsilateral than from the contralateral pulvinar. Bilateral filtering without any lateral bias may only occur at levels higher than the IT cortex (see Everling, Tinsley, Gaffan, & Duncan, 2002, for findings in the PF cortex).

Chelazzi, Miller, Duncan, and Desimone (2001). Using a variation of the experimental design described above (and the same monkeys), Chelazzi et al. (2001) extended the investigation to V4 neurons. The results were similar, with a few notable exceptions.

First, with two search stimuli in the RF (*inside/inside* condition), the wave of selective processing appeared to begin after approximately 150 ms. This was a little earlier than in IT, but several factors may underlie the difference in timing: The monkeys had more training than in the previous study, and the stimuli were positioned closer to each other. Second, in contrast to Chelazzi et al.'s (1998) study of IT neurons, this study of V4 neurons showed no increase in baseline activity with attention. This contrast is discussed in the *Attentional Effects on Baseline Firing* section.

Filtering of Stimuli With Different Contrast

By Equation 2 of NTVA, the attentional weight of an object depends on both the sensory evidence that the object has certain features (η values) and the behavioral relevance (π values) of those features. A recent study by Reynolds and Desimone (2003) demonstrates the influence of both of these factors on the filtering mechanism. Strength of sensory evidence was manipulated by varying the luminance contrast of objects. When two distractor objects were shown at different levels of contrast in the RF of a V4 neuron, Reynolds and Desimone found that the neuron's firing rate was primarily determined by the more visible stimulus. However, the effect could be reversed if attention was directed to the fainter stimulus. Similar results were obtained by Martínez-Trujillo and Treue (2002) in the dorsal stream of visual processing (area MT).

Reynolds and Desimone (2003). Reynolds and Desimone (2003) studied responses of V4 neurons to stimuli at various levels of contrast. A monkey's task was to monitor a sequence of displays, reacting when a target stimulus (a diamond) appeared at a cued location (cf. Luck et al., 1997; Reynolds et al., 1999). The target and distractor stimuli (rectangular patches of sinusoidal luminance grating) could appear at up to four locations, two inside and two outside of the recorded neuron's RF. For each neuron, a pair of distractor stimuli was chosen, and this formed the basis of the response analysis. The first of these stimuli, the *reference*, remained at a fixed (high) contrast throughout the experiment, whereas the second, *probe* stimulus was shown at varying contrast.

In the *attend-outside-RF* condition, the response (mean rate of firing) to the reference stimulus was compared with the response to a pair consisting of reference plus probe. When the probe was less effective at driving the neuron than was the reference stimulus, the addition of the probe to the display typically caused a reduction in response (cf. Reynolds et al., 1999). However, the influence of the probe depended on its level of contrast. If the probe had very low contrast, the neuron's response to the pair was approximately equal to the response to the reference stimulus alone. As the contrast of the probe was increased, the neuron's response became more and more suppressed, reflecting stronger influence of the probe. In other words, the attentional weight of the probe seemed to increase with its visibility (i.e., η values).

In the *attend-inside-RF* condition, the monkey's attention was drawn to the probe stimulus in the RF by cuing of the location of the probe. Presumably, the effect of cuing the location of the probe was to increase the pertinence (π value) of this location and, accordingly, the attentional weight of the probe. The manipulation caused the pair response (mean rate of firing to the pair consisting of the probe and the reference) to move toward the response to the probe alone, even when the probe was shown at lower contrast. Thus, the response to the pair could be driven toward the response to the probe both by increasing the contrast of the probe and by increasing the pertinence of the probe's location. In accordance with Equation 2, both manipulations (of η and π values, respectively) should increase the attentional weight of the probe, and both manipulations drove the mean rate of firing to the pair toward the response to the probe when it was presented alone.

Martínez-Trujillo and Treue (2002). Martínez-Trujillo and Treue (2002) recorded from area MT, which contains cells that are selective to direction of motion. They presented monkeys with patterns of coherently moving random dots. Four random-dot patterns (RDPs) were shown: one pair inside and one pair outside of the RF (see Figure 12). The two pairs were identical, either one consisting of (a) a pattern moving in the preferred direction of the neuron and (b) a pattern moving in the opposite, null direction (i.e., eliciting a response close to baseline). Before each trial, the monkey was cued to attend the null pattern either in the RF or at the other location. The task was to detect small changes in the movement of this pattern (cf. Treue & Martínez-Trujillo, 1999) while ignoring changes in the other patterns. Martínez-Trujillo and Treue varied the contrast of the unattended pattern in the RF to see how this changed the neuron's response, depending on whether the monkey was attending inside or outside the RF. The contrast was defined as the standard deviation of the pixel-by-pixel variation of the luminance of the stimulus (the dots and the background). The attended (null) pattern was always displayed at high contrast.

Given this design, NTVA predicts the following response pattern. With a probe pattern of very low contrast, the neuron's response should be almost entirely determined by the (high contrast) null stimulus, regardless of whether the monkey is attending inside or outside the RF. This should result in a response close to baseline in both conditions. As the contrast of the preferred stimulus is increased, its probability of being selected should also increase (cf. Equation 2), driving the response up. However, the increase should be weaker in the attend-inside-RF condition, because in this case, the behavioral relevance of the null pattern should reduce the probability that the preferred pattern was selected (represented) by the cell (cf. Reynolds & Desimone, 2003). At maximum contrast (i.e., when the preferred and null patterns are equally visible), the probability of selecting the preferred pattern should reach a ceiling of about 50% in the attend-outside-RF condition. However, in the attend-inside-RF condition, the probability should remain below 50% because of attention to the null pattern. Thus, NTVA predicts a difference between the responses in the attend-inside-RF and attend-outside-RF conditions even at maximum contrast.

Contrary to the prediction by NTVA, Martínez-Trujillo and Treue (2002) suggested that at maximum contrast, responses were essentially the same in the two attention conditions (attend-inside-RF vs. attend-outside-RF). This suggestion was based on the fact that although hyperbolic ratio contrast response functions (CRFs) fitted to their data showed significant differences between the two attention conditions, the estimated asymptotes of the CRFs did not differ significantly between the two conditions. However, considering previous demonstrations of robust effects of attention with multiple stimuli in the RF (also shown in MT: Treue & Martínez-Trujillo, 1999; Treue & Maunsell, 1999), it seems implausible that the effect of attention should disappear at high contrasts. Also, the empirical data of Martínez-Trujillo and Treue (2002; summarized in their Figure 5B [p. 367] and in the present

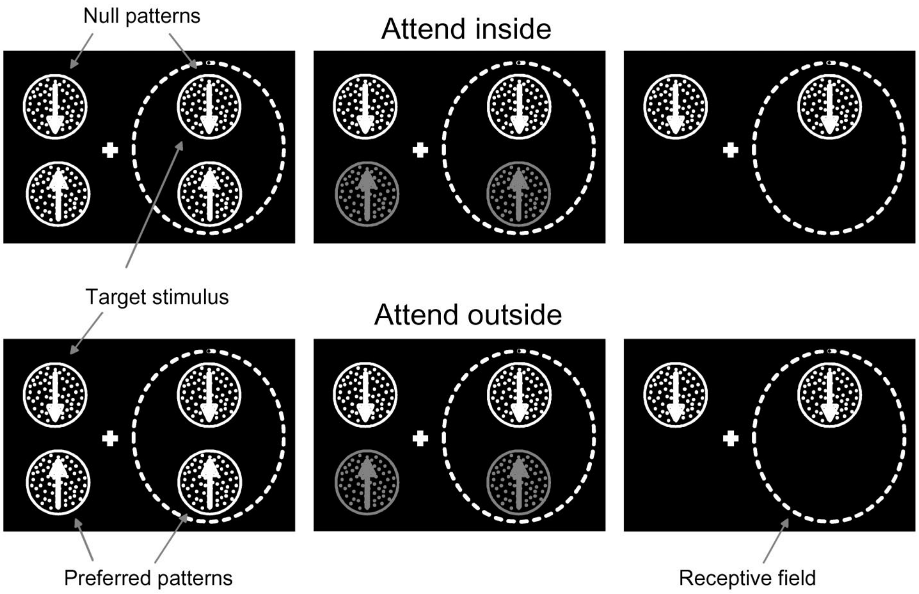

Figure 12. Experimental design for a cell preferring upward motion in the experiment of Martínez-Trujillo and Treue (2002). Each stimulus display showed two pairs of random-dot patterns, one pair inside and one pair outside of the receptive field (RF; dashed oval) of the recorded cell. Each pair consisted of one preferred and one null pattern. In the *attend-inside* condition, monkeys attended to the null pattern inside the RF. In the *attend-outside* condition, monkeys attended to the null pattern outside of the RF. From left to right, the panels show decreasing contrast of the preferred patterns. Adapted from *Neuron, 35,* J. C. Martínez-Trujillo and S. Treue, "Attentional Modulation Strength in Cortical Area MT Depends on Stimulus Contrast," pp. 365–370, Copyright 2002, with permission from Elsevier.

Figure 13) indicate substantial differences between responses in the two attention conditions, even at high contrast.

Figure 13 displays a quantitative fit to the data of Martínez-Trujillo and Treue (2002, Figure 5B) based on NTVA. The abscissa shows the contrast index, (contrast − C50)/(contrast + C50), where C50 is the contrast generating a response in the attend-outside-RF condition that is half as strong as the response obtained with the maximum contrast. The maximum firing rate varied between MT neurons, but for each of the tested neurons in each of the two attention conditions, the activations of the neuron (i.e., the firing rates of the neuron minus the neuron's baseline rate) were normalized by being expressed as proportions of the neuron's maximum activation in the attend-outside-RF condition. The normalized activations for each condition were binned by the contrast index, averaged across neurons within the bins, and plotted along the ordinate in Figure 13.

The smooth curves in the figure show a fit based on the following assumptions. First, for the investigated neurons, the mean firing rate of a cell is a sum of a baseline rate (undriven activity) and a mean activation (stimulus-driven activity). In accordance with Equation 3, the mean activation of the cell (say, the kth feature-i neuron) can be written as

$$v_{ki} = \eta_k(x, i)\beta_i \frac{w_x}{w_x + w_z}, \tag{5}$$

where x is the preferred pattern inside the cell's RF, z is the null pattern inside the cell's RF, and w_x and w_z are the attentional weights of x and z, respectively; β_i is the featural bias in favor of i; $\eta_k(x, i)\ \beta_i$ is the activation of the (kth feature-i) neuron when it responds to the preferred pattern (x) with the given featural bias (β_i); and $w_x/(w_x + w_z)$ equals the probability that the cell responds to the preferred pattern rather than to the null pattern.

Second, all η values for features of pattern x increase with the contrast of pattern x. Let c be the natural logarithm of the stimulus contrast. For simplicity, we assume that all η values for task-relevant features of pattern x increase by the same sigmoid CRF,

$$f_{lsa}(c) = \{1 + \exp[-2(c - l)/s]\}^{-a}, \tag{6}$$

which implies that for all k and i,

$$\eta_k(x, i) = \eta_k^*(x, i) f_{lsa}(c), \tag{7}$$

where $\eta_k^*(x, i)$ is the value of $\eta_k(x, i)$ at maximum contrast (theoretically, the asymptotic value of $\eta_k[x, i]$ as c approaches infinity). Similarly, for all $\eta(x, j)$ values contributing to the attentional weight of pattern x (see Equation 2),

$$\eta(x, j) = \eta^*(x, j) f_{lsa}(c), \tag{8}$$

where $\eta^*(x, j)$ is the value of $\eta(x, j)$ at maximum contrast.

For $l = 0$, $s = 1$, and $a = 1$, $f_{lsa}(c)$ is the standard logistic function of the logarithm of the contrast, c. The standard logistic function is rotationally symmetric about the point (0, ½), approaches 1 as c approaches infinity, and approaches 0 as c approaches minus infinity. Parameter l determines the location of the CRF $f_{lsa}(c)$ along the c axis, scale parameter s determines the steepness of the function, and parameter a determines the degree of

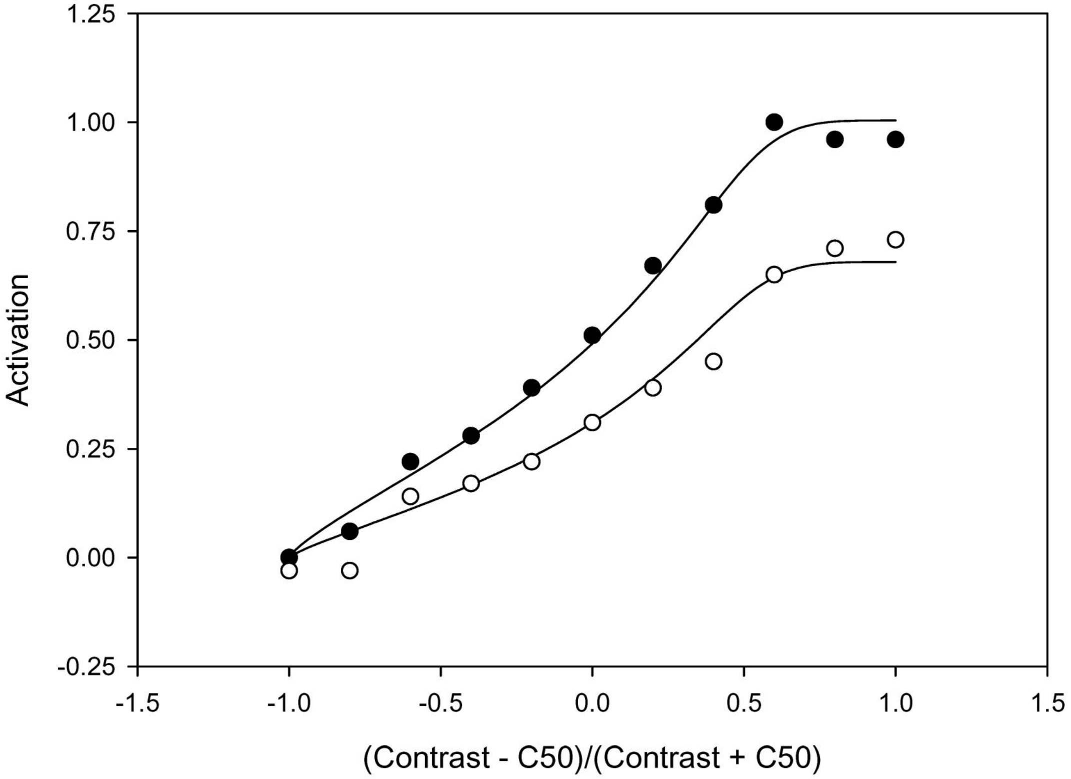

Figure 13. Effect of attention in the experiment of Martínez-Trujillo and Treue (2002). The normalized activation is shown as a function of an index of the stimulus contrast when a monkey was attending to a stimulus inside the receptive field (RF; open circles) and when the monkey was attending to a stimulus outside the RF (solid circles). A theoretical fit based on NTVA (neural theory of visual attention) is indicated by smooth curves. C50 = the contrast generating a response in the attend-outside-RF condition that is half as strong as the response obtained with the maximum contrast. The data are adapted from *Neuron, 35,* J. C. Martínez-Trujillo and S. Treue, "Attentional Modulation Strength in Cortical Area MT Depends on Stimulus Contrast," pp. 365–370, Copyright 2002, with permission from Elsevier.

rotational asymmetry of the function. Examples of theoretical CRFs $f_{lsa}(c)$ with different values of parameters l, s, and a are shown in Figure 14.[6]

By Equations 2 and 8, the attentional weight of pattern x also increases by the CRF $f_{lsa}(c)$—that is,

$$w_x = w_x^* f_{lsa}(c), \tag{9}$$

where

$$w_x^* = \sum_{j \in R} \eta^*(x, j)\pi_j$$

is the value of w_x at maximum contrast. By Equations 5, 7, and 9, the mean activation of the kth feature-i neuron is

$$v_{ki} = m\, f_{lsa}(c)\, \frac{f_{lsa}(c)}{f_{lsa}(c) + w_{\text{ratio}}}, \tag{10}$$

where $m = \eta_k^*(x, i)\ \beta_i$ and $w_{\text{ratio}} = w_z/w_x^*$. Note that at maximum contrast, Equation 10 reduces to

$$\max(v_{ki}) = \frac{m}{1 + w_{\text{ratio}}} = m\, \frac{w_x^*}{w_x^* + w_z}.$$

In the attend-outside-RF condition, both the null pattern and the preferred pattern inside the neuron's RF were distractors, so w_{ratio} should be about 1. In the attend-inside-RF condition, the null pattern was the target, so w_{ratio} should be greater than 1. The fit by Equation 10, shown in Figure 13, was obtained with w_{ratio} kept constant at a value of 1 in the attend-outside-RF condition. The fit was found with w_{ratio} at 1.96 in the attend-inside-RF condition, normalized maximum activation m at 2.01, and a CRF f_{lsa} with location parameter l at 1.11 log units of contrast, scale parameter s at 0.52, and asymmetry parameter a at 0.11.

The fit shown in Figure 13 is close. Note, in particular, that in both the observed data and the fitted functions, the relative effect of attention (measured, e.g., by [response attending outside RF − response attending inside RF]/[response attending outside RF +

[6] The theoretical CRFs that Martínez-Trujillo and Treue (2002) fitted to their data were functions of the form f_{lsa}[log(contrast)] with $a = 1$ (i.e., rotationally symmetric functions of the logarithm of the contrast). Sigmoid cortical CRFs have also been reported by Albrecht and Hamilton (1982) and Tolhurst, Movshon, and Thompson (1981), among others.

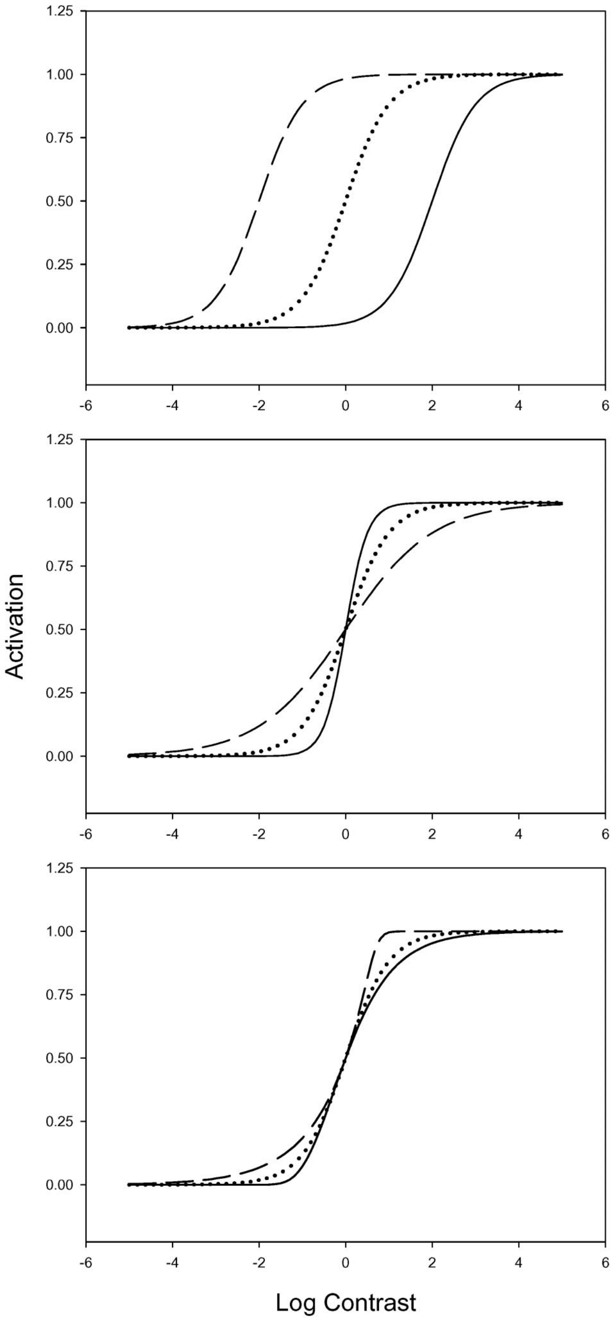

Figure 14. Theoretical contrast response functions $f_{lsa}(c)$ with different values of location parameter l, scale parameter s, and asymmetry parameter a. Top: $l = -2$ (dashed), 0 (dotted), or 2 (solid); $s = 1$; $a = 1$. Middle: $l = 0$; $s = 0.5$ (solid), 1.0 (dotted), or 2.0 (dashed); $a = 1$. Bottom: $l = 0.693$, $s = 0.2$, and $a = 0.1$ (dashed); $l = 0$, $s = 1$, and $a = 1$ (dotted); or $l = -4.55$, $s = 1.5$, and $a = 300$ (solid).

response attending inside RF]) was greatest in the midcontrast range. Thus, the main findings of Martínez-Trujillo and Treue (2002) can be explained by Equation 5 on the assumption that η values and, therefore, attentional weights are sigmoid functions of stimulus contrast.

Attentional Effects With a Single Stimulus in the RF

Experiments with multiple stimuli in the RF of the recorded neuron have consistently shown strong effects of attention. In studies with only one experimental stimulus in the RF, effects of attention have generally been much smaller and less consistent. Moran and Desimone (1985) and Luck et al. (1997) found no effect of attention when only a single stimulus was present in the RF of the recorded neuron. Other investigators (Connor, Preddie, Gallant, & Van Essen, 1997; McAdams & Maunsell, 1999a, 1999b, 2000; Motter, 1994a, 1994b; Reynolds et al., 1999; Reynolds, Pasternak, & Desimone, 2000; Treue & Martínez-Trujillo, 1999) have reported some enhancement of firing rates with attention, whereas Motter (1993) has reported both positive and negative modulations of firing rates. The present section contains an analysis of the findings.

When only one object x appears in the RF of a recorded feature-i neuron, NTVA implies that the neuron responds to object x. The resultant activation of the kth feature-i neuron should equal $\eta_k(x, i)\ \beta_i$, where $\eta_k(x, i)$ is independent of attention, whereas β_i is the perceptual bias in favor of feature i. Hence, an effect of attention should be found if, and only if, the perceptual bias in favor of feature i is varied (pigeonholing), and the effect of varying the bias (β_i) should be a multiplicative scaling of the activations of all feature-i neurons.

When a single experimental stimulus appears in the RF of a recorded neuron, the stimulus may be the only noteworthy object in the RF, but this need not be the case. Even when an experiment is designed to include only one stimulus in the RF, the neuron may respond to an individual part of the experimental stimulus as a separate object instead of responding to the whole stimulus, or the neuron may respond to a ghost object formed by internal random noise. Effects of individual parts of an experimental stimulus should prevail in experiments with a complex stimulus in the RF. For example, if the experimental stimulus is a cloud of dots moving within the classical RF of a recorded neuron (cf. Treue & Martínez-Trujillo, 1999), the neuron may sometimes respond to the whole pattern but at other times respond to a subpattern or a single dot. Effects of internally generated random noise should prevail in experiments with a faint stimulus in the RF. In these cases, the neuron is faced with a classical problem of signal detection (cf. Green & Swets, 1966): discrimination of a weak signal (the experimental stimulus) from pure noise (the ghost object formed by internal random noise).

In general, a neuron presented with a single stimulus in its RF may perform a filtering operation and respond to either the experimental stimulus (as a whole) or one out of a larger set of noise objects. The set of possible noise objects includes individual parts of the experimental stimulus as well as ghost objects formed by internal random noise. The probability that the neuron responds to a particular object equals the attentional weight of the object divided by the sum of the attentional weights of all objects in the RF. Thus, the neuron's response can be regarded as a probability mixture of its response to the experimental stimulus and its responses to the noise objects. Accordingly, the neuron's mean response is a weighted average of its response to the experimental stimulus object and its mean responses to each of the noise objects (with a weight on the mean response to a particular object equal to the probability that the object in question is selected by the cell).

Hence, the neuron's mean response can also be described as a weighted average of the mean response to the experimental stimulus object (with a weight, p, equal to the probability that the neuron responds to the experimental stimulus) and the mean response across all noise objects (with a weight equal to $1 - p$).

An extensive study of attentional effects with a single, faint stimulus in the RF of the recorded neuron is treated in the first subsection below. Studies of attentional effects with a single, complex stimulus in the RF are treated in the second subsection. Attentional studies with a single, relatively simple and strong stimulus in the RF are treated in the third subsection.

Attentional Effects With a Faint Stimulus in the RF

With a single, faint experimental stimulus in its RF, a neuron is faced with a classical problem of signal detection (cf. Green & Swets, 1966): discrimination of a weak signal (the stimulus) from pure noise (ghost objects formed by internal random noise). In this situation, NTVA assumes a substantial probability that the neuron responds to noise instead of responding to the faint stimulus. A recent study by Reynolds et al. (2000) has provided an intricate pattern of data on the way that attentional effects depend on stimulus contrast. Below, we show how the data can be accounted for in terms of NTVA.

Reynolds, Pasternak, and Desimone (2000). Reynolds et al. (2000) studied responses of V4 neurons to single stimuli presented in the RF, using a method adapted from Luck et al. (1997) and Reynolds et al. (1999). A monkey fixated a small dot at the center of a computer screen. Sequences of oriented, bar-shaped patches of grating were simultaneously presented at two locations, one inside the RF of the recorded V4 neuron and the other at an equally eccentric position in the opposite hemifield. At the beginning of a block of trials, a cue indicated which sequence should be attended. On each trial, stimulus sequences of variable length appeared simultaneously at the two locations. The monkey's task was to release a bar when it detected a target stimulus (a rotated square patch of grating) at the cued location.

The contrast of a stimulus was defined as (maximum luminance − minimum luminance)/(maximum luminance + minimum luminance). For each neuron Reynolds et al. (2000) selected five contrasts that were spaced at equal log intervals of contrast (typically by doubling the next lower contrast) so that these contrasts spanned the dynamic range of the cell. The contrast of the stimuli (including targets) varied at random among the five values from presentation to presentation. The orientation and spatial frequency of the grating stimuli were selected to be nonoptimal for the cell, so not even the one with highest contrast elicited the strongest possible response from the neuron. For this reason, lack of attentional enhancement of responses to the strongest stimuli could not be due to a physiological ceiling effect on the firing rate.

All analyses were based on the neurons' responses to distractors. Figure 15 shows the mean firing rate of 84 tested neurons as a function of the stimulus contrast with attention condition (attend-inside-RF vs. attend-outside-RF) as the parameter. As can be seen, the mean firing rate increased as the contrast of the stimulus grating was increased, and at all levels of stimulus contrast, the firing rate was higher to attended than to ignored stimuli (i.e.,

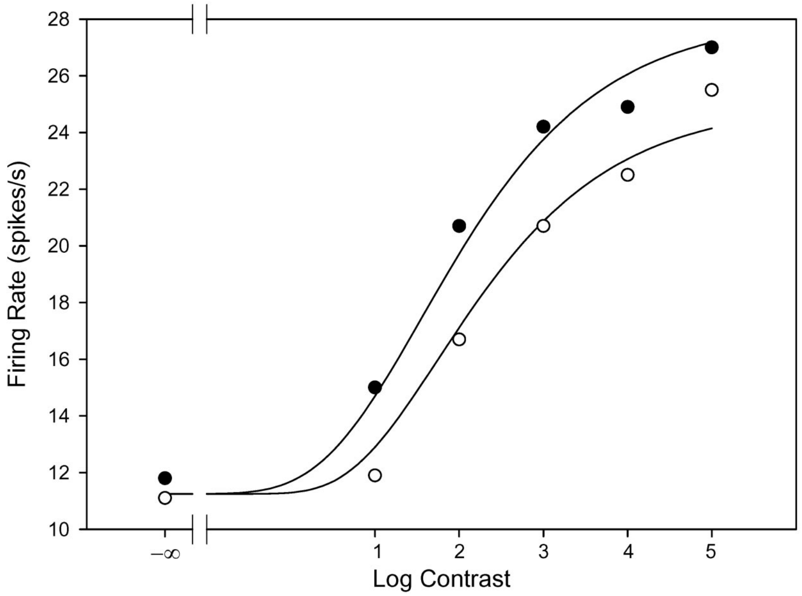

Figure 15. Effect of attention in the experiment of Reynolds, Pasternak, and Desimone (2000). The mean firing rate is shown as a function of the logarithm of the stimulus contrast when a monkey attended to a stimulus inside the receptive field (RF; solid circles) and when the monkey attended to a stimulus outside the RF (open circles). Values of the contrast along the abscissa have been normalized so that for each neuron, the five selected contrasts are found at 1, 2, 3, 4, and 5 log units, respectively. A theoretical fit based on NTVA (neural theory of visual attention) with the same contrast response function for all features is indicated by smooth curves. Data are adapted from *Neuron, 26,* J. H. Reynolds, T. Pasternak, and R. Desimone, "Attention Increases Sensitivity of V4 Neurons," pp. 703–714, Copyright 2000, with permission of Elsevier.

higher in the attend-inside-RF condition than in the attend-outside-RF condition).

The smooth curves in Figure 15 show a fit based on the hypothesis that when only a faint stimulus is present inside the classical RF of a V4 neuron, there is a substantial probability that the neuron responds to internal random noise rather than responding to the faint stimulus. We further assumed that when the neuron responds to internal random noise, the firing rate of the neuron equals the baseline rate (undriven activity) of the neuron, *b*. When the neuron responds to the stimulus, *x*, the firing rate of the neuron (say, the *k*th feature-*i* neuron) equals $b + \eta_k(x, i)\ \beta_i$, where β_i is the featural bias in favor of *i*. The probability that the neuron responds to the stimulus (*x*) rather than to the noise (*z*) equals $w_x/(w_x + w_z)$, where w_x is the attentional weight of the stimulus and w_z is the attentional weight of the noise. Hence, the mean rate of firing when stimulus *x* is the only real stimulus in the RF can be written as $b + v_{ki}$, where v_{ki} is given by Equation 5,

$$v_{ki} = \eta_k(x, i)\beta_i \frac{w_x}{w_x + w_z}.$$

As we did when fitting the data of Martínez-Trujillo and Treue (2002), we assumed that all η values for features of the stimulus grating increase with the contrast of the grating in accordance with the same CRF $f_{lsa}(c)$, given by Equation 6, so that (by Equations 7–10)

$$v_{ki} = m\, f_{lsa}(c) \frac{f_{lsa}(c)}{f_{lsa}(c) + w_{\text{ratio}}},$$

where $m = \eta^*_k(x, i)\ \beta_i$ (the activation of an average feature-*i* neuron when it responds to stimulus *x* rather than responding to noise and *x* is presented at maximum contrast), and $w_{\text{ratio}} = w_z/w^*_x$ (the ratio between the attentional weight of the noise and the attentional weight of the stimulus grating at maximum contrast).

In the attend-inside-RF condition, stimuli inside the RF should be attended, but noise should be ignored, so w_{ratio} should be as small as possible. The fit shown in Figure 15 was obtained with $w_{\text{ratio}} = 0$ in the attend-inside-RF condition but $w_{\text{ratio}} = 0.22$ in the attend-outside-RF condition. In both conditions, baseline rate *b* was 11.2 spikes/s, parameter *m* was 16.9 spikes/s, and the CRF had a location parameter *l* at −5.25 log units of contrast, scale parameter *s* at 2.39, and asymmetry parameter *a* at 296.

The fit shown in Figure 15 is based on very simple assumptions, and it is close. In this fit, the relative effect of attention (measured, e.g., by [response attending inside RF − response attending outside RF]/[response attending inside RF + response attending outside RF]) is greatest in the midcontrast range, but the absolute effect of attention (i.e., response attending inside RF − response attending outside RF) increases monotonically as the contrast is increased. There is a trend in the observed data that not only the relative but also the absolute effect of attention is greatest in the midcontrast range. This trend may be captured in the fit by relaxing the assumption that all η values for features of the stimulus grating increase with the contrast of the grating in accordance with the same CRF $f_{lsa}(c)$. Figure 16 shows a very close fit to the data, which was obtained with two different CRFs of the form $f_{lsa}(c)$. For both CRFs, parameter *a* was fixed at a value of 1. The CRF for $\eta_k(x, i)$ had location parameter *l* and scale parameter *s*, so

$$\eta_k(x, i) = \eta^*_k(x, i)\ f_{ls1}(c),$$

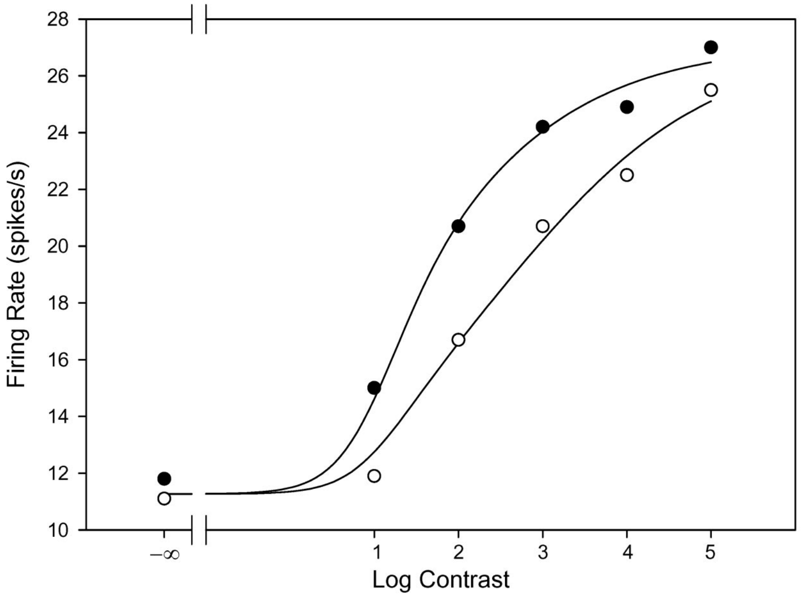

Figure 16. Alternative fit to the effect of attention in the experiment of Reynolds, Pasternak, and Desimone (2000). The empirical data are the same as those shown in Figure 15. A theoretical fit based on NTVA (neural theory of visual attention) with different contrast response functions for different features is indicated by smooth curves. Data are adapted from *Neuron, 26,* J. H. Reynolds, T. Pasternak, and R. Desimone, "Attention Increases Sensitivity of V4 Neurons," pp. 703–714, Copyright 2000, with permission from Elsevier.

where $\eta_k^*(x, i)$ is the value of $\eta_k(x, i)$ at maximum contrast. The CRF for η values contributing to the attentional weight of x had location parameter l' and scale parameter s', so

$$w_x = w_x^* f_{l's'1}(c),$$

where w_x^* is the value of w_x at maximum contrast. The fit was obtained with $w_{\text{ratio}} = 0.004$ in the attend-inside-RF condition and $w_{\text{ratio}} = 0.014$ in the attend-outside-RF condition. Baseline rate b was 11.3 spikes/s, and parameter m was 15.9 spikes/s. The CRF for $\eta_k(x, i)$ had a location parameter l at 1.02 log units of contrast and a scale parameter s at 0.72, whereas the CRF for w_x had a location parameter l' at 7.68 log units of contrast and a scale parameter s' at 2.33.

Attentional Effects With a Complex Stimulus in the RF

With a single, complex experimental stimulus in the RF of a recorded neuron, NTVA assumes a substantial probability that the neuron responds to an individual part of the experimental stimulus rather than responding to the stimulus as a whole. The probability should depend on a monkey's state of attention. If the monkey is instructed to treat the complex stimulus as a target, this should increase the attentional weight on the complex stimulus and, therefore, increase the probability that the recorded neuron responds to the complex stimulus rather than responding to a noise object, such as a smaller part of the stimulus. Hence, when the experimental stimulus object is a more effective stimulus for the cell than are any of the noise objects, attention should increase the firing rate of the cell. As detailed below, this filtering mechanism explains cases in which spatial attention has enhanced the response to a cloud of moving dots (Treue & Martínez-Trujillo, 1999) or a Gabor pattern (McAdams & Maunsell, 1999a, 1999b, 2000).

In addition to providing evidence of filtering with a complex stimulus in the RF, the studies by Treue and Martínez-Trujillo (1999) and McAdams and Maunsell (2000) have also provided evidence of pigeonholing: A feature-based mechanism of attention that selects a group of neurons with a given stimulus preference for a multiplicative enhancement in activation. The evidence consists, in part, in data showing that attention to a preferred feature of a neuron (e.g., a certain direction of motion) enhanced the response of the neuron even though the stimulus to be attended was outside the RF of the neuron.

Finally, McAdams and Maunsell (1999a) and Treue and Martínez-Trujillo (1999) have provided evidence of multiplicative scaling of neural tuning curves with attention. More precisely, attention was found to increase a neuron's activation (firing rate above baseline) by the same factor for all objects across a stimulus dimension (e.g., orientation). NTVA directly predicts such an effect resulting from pigeonholing: the multiplicative effect of the β factor. NTVA also predicts multiplicative scaling of the tuning curve when the attentional weight of the experimental stimulus is varied, provided that activations caused by noise objects are either zero or proportional to the activation caused by the experimental stimulus. To see this, consider the orientation tuning of a neuron whose mean activation, $v(\theta)$, is a function of the stimulus orientation, θ. Let w_x be the attentional weight of the stimulus, and let w_z be the attentional weight of the noise (i.e., the sum of the attentional weights of all noise objects). The probability that the neuron responds to the experimental stimulus is given by

$$p = \frac{w_x}{w_x + w_z}.$$

The mean activation of the neuron is a weighted average of the mean activation, $v_x(\theta)$, when the neuron responds to the experimental stimulus (which happens with probability p) and the mean activation, $v_z(\theta)$, when the neuron responds to a noise object (which happens with probability $1 - p$):

$$v(\theta) = p\, v_x(\theta) + (1 - p)\, v_z(\theta).$$

If activations caused by noise objects are zero (i.e., $v_z(\theta) = 0$), we get

$$v(\theta) = p\, v_x(\theta),$$

which implies multiplicative scaling of the tuning curve: As the relative attentional weight (p) of the experimental stimulus is varied, the mean activation of the neuron is scaled by the same factor (p) for all stimulus orientations (θ).

A similar argument applies if the mean activation when the neuron responds to a noise object is proportional to the mean activation when the neuron responds to the experimental stimulus (i.e., $v_z[\theta] = k\, v_x[\theta]$, where k is a constant independent of θ). In this case, we get

$$\begin{aligned} v(\theta) &= p\, v_x(\theta) + (1 - p)k\, v_x(\theta) \\ &= [p + (1 - p)k]v_x(\theta) \\ &= q v_x(\theta), \end{aligned}$$

where

$$q = p + (1 - p)k,$$

which also implies multiplicative scaling of the tuning curve: As the relative attentional weight (p) of the experimental stimulus is varied, the mean activation of the neuron is scaled by the same factor (q) for all stimulus orientations (θ).

Activations at zero (i.e., responses at a baseline corresponding to undriven activity) should result from ghost objects formed by internal random noise. Activations proportional to the activation caused by the experimental stimulus would be expected from noise objects that are parts of the experimental stimulus if the parts have the same value as the whole stimulus on the dimension along which the tuning curve is defined (e.g., the same orientation if this is the dimension along which tuning is being measured). Below, we argue that this condition appears to have been satisfied by the stimuli used by Treue and Martínez-Trujillo (1999) and McAdams and Maunsell (1999a, 1999b).

Treue and Martínez-Trujillo (1999). Treue and Martínez-Trujillo (1999) recorded from area MT, which contains cells that are selective to direction of motion. Although the area is located in the dorsal visual stream, the effects of attention were similar to those found in studies of ventral areas (McAdams & Maunsell, 1999a, 1999b, 2000; see below). In the basic task, Treue and Martínez-Trujillo presented monkeys with two coherently moving RDPs, one placed inside the RF of the neuron being recorded and the other placed in the opposite visual hemifield. At the start of each trial, a monkey was shown a cue at one of the locations. Following this, the RDPs appeared and the monkey was required to detect small changes in the speed or direction of the pattern at the cued location. These changes occurred after a random delay ranging between 270 and 4,000 ms.

Experiment 1 of Treue and Martínez-Trujillo (1999) demonstrated an effect of filtering based on spatial location (*spatial attention*). In this experiment, the two RDPs were moving in the same direction on any given trial (12 different directional pairs were used to obtain a tuning curve for the neuron). When attention was directed at the RDP in the RF of the recorded cell, the cell's mean activation (i.e., response above baseline) was about 10% higher than it was when attention was directed at the stimulus outside the RF. The increase in activation occurred without any sharpening of the tuning curve around the preferred direction (cf. McAdams & Maunsell, 1999a). In contrast, the relative increase with attention was approximately the same across all orientations (*multiplicative modulation*).

In NTVA, these results can be explained by assuming that the probability of selecting the whole RDP differed between the two experimental conditions. The stimulus used by Treue and Martínez-Trujillo (1999), a cloud of moving random dots, was so complex that there should be a substantial probability that the recorded cell responded to only a part of it. In the attended condition, the monkey presumably treated the complex stimulus in the RF of the recorded neuron as the target, which should increase the probability that the neuron responded to the complex stimulus rather than responding to a smaller part of the stimulus. Furthermore, the whole RDP should elicit a stronger response than would parts consisting of only one or a few moving dots (we assume that Treue and Martínez-Trujillo, 1999, chose the cloud stimulus because it was more effective than individual dots at driving the cell). This explains why activations were stronger in the attended condition. Furthermore, all dots were moving coherently, so selection of smaller parts of the RDP should elicit activations that were approximately proportional to, but lower than, the activation elicited by the whole RDP. Given these conditions, attentional modulation should be multiplicative across the tuning curve. The effect is illustrated in Figure 17.

Experiment 2 of Treue and Martínez-Trujillo (1999) demonstrated an effect of pigeonholing with respect to a given direction of movement (*feature-based attention*)—probably the first clear demonstration of pigeonholing in the single-cell literature. In this experiment, the recorded neuron's RF was stimulated by an RDP moving in the direction preferred by the neuron. Outside the RF, an RDP was presented that moved either in the same direction or in the opposite direction. When the monkey attended to the RDP outside the RF, the response of the recorded neuron varied with the direction of movement being attended. Depending on whether the direction in the attended pattern was the same as or the opposite of

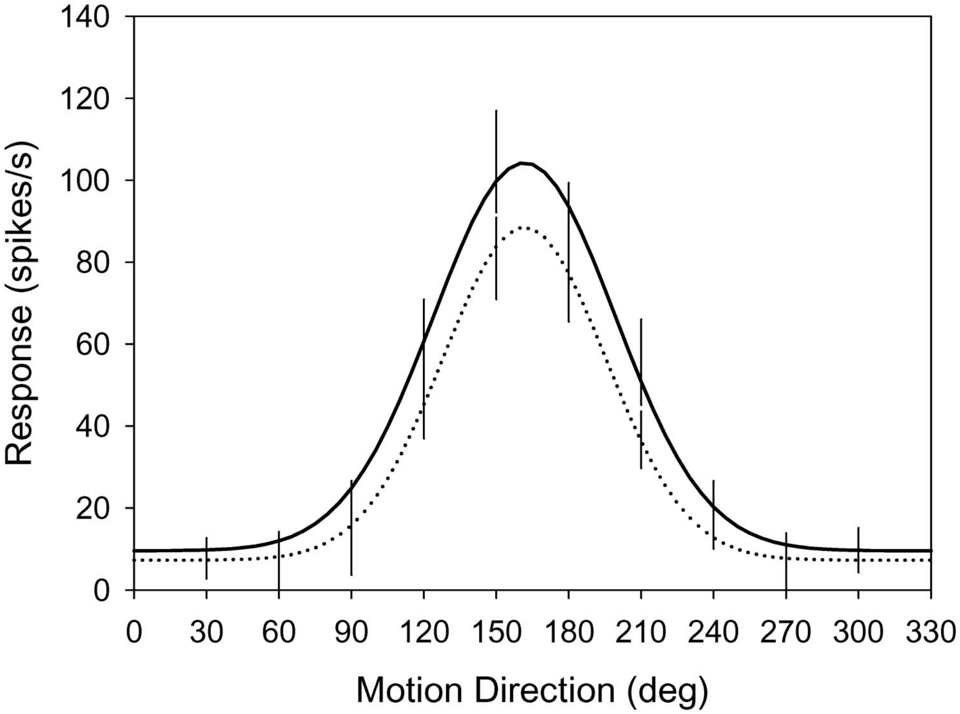

Figure 17. Tuning curves for a motion-sensitive cell in the middle temporal visual area in Experiment 1 of Treue and Martínez-Trujillo (1999). The stimulus in the receptive field (RF) of the recorded cell moved in the same direction as the stimulus outside the RF. The solid curve shows the mean firing rate of the cell as a function of the stimulus direction when a monkey was attending to the stimulus inside the RF, and the dotted curve shows the mean firing rate when the monkey was attending to the stimulus outside the RF. From "Feature-Based Attention Influences Motion Processing Gain in Macaque Visual Cortex," by S. Treue and J. C. Martínez-Trujillo, 1999, *Nature, 399,* Figure 1b, p. 576. Copyright 1999 by Nature Publishing Group (http://www.nature.com). Adapted with permission.

that preferred by the cell being recorded, the firing rate went, respectively, up or down. The attentional modulation of the neuron's response occurred even though the spatial location of the attended pattern was unchanged between the two conditions. Thus, a nonspatial, feature-based mechanism of attention seemed to be at work: pigeonholing (see Saenz, Buracas, & Boynton, 2002, for highly similar findings obtained by fMRI).

The result is readily explained by the hypothesis that β values for particular directions of movement differed between the two conditions of the experiment. Because the monkey was monitoring the pattern for hundreds to thousands of milliseconds, there should be plenty of time to adjust the β values in accordance with the display within each trial. Consider the situation in which the preferred movement of the recorded neuron was downward but the monkey attended to movement in the opposite direction (upward), trying to detect a small change in the speed or direction of movement of the RDP to be attended. Because the monkey had learned that only small changes in the direction of movement of the target would occur, β_{downward} should be low (perceptual bias should generally reflect expectations). Because the activation of the recorded cell should be proportional to β_{downward}, the recorded activation also should be low. By contrast, when the monkey attended to movement in the preferred direction of the recorded cell, β values should be high for downward and nearby directions. In this case (β_{downward} being high), the activation of the recorded neuron also should be high.

Treue and Martínez-Trujillo (1999) estimated the combined effects of filtering (spatial attention) and pigeonholing (feature-based attention) by comparing (a) trials on which the monkey was attending the antipreferred direction outside the RF with (b) trials on which attention was directed inside the RF to a stimulus moving in the preferred direction. The total increase in activation due to filtering and pigeonholing was about 25%, corresponding to a 10% increase in activation due to change in the relative attentional weight of the complex stimulus within the RF (i.e., increase in $w_x/(w_x + w_z)$, where w_x is the weight of the complex stimulus, and w_z is the sum of the weights of the noise objects in the RF) combined with a 13% increase in the perceptual bias (β value) for the preferred direction of the recorded neuron ($110\% \times 113\% \approx 125\%$, consistent with Equation 1 of TVA).

McAdams and Maunsell (1999a, 1999b). McAdams and Maunsell (1999a) studied the effect of attention on the orientation tuning of V4 neurons. They wanted to test whether attention alters the stimulus selectivities of the neurons or, more specifically, whether attention sharpens the tuning curves around a preferred orientation (as suggested by Haenny & Schiller, 1988, and Spitzer, Desimone, & Moran, 1988). Monkeys were tested in a delayed match-to-sample task. They were shown two sample stimuli for 500 ms, one of them at a cued location. The monkeys had to retain the cued stimulus during a delay period of 500 ms and then match it to a test stimulus presented at the same location (ignoring the test stimulus at the other location). Throughout each testing block, the cued location was kept constant. The stimuli used in the experimental task were Gabor patterns (constructed by multiplying a sinusoidal grating and a 2-D Gaussian) and colored Gaussians (isoluminant colored patches whose saturation varied with a 2-D Gaussian profile). The stimuli were always located so that the Gabor stimulus was in the RF of the neuron being recorded, whereas the colored Gaussian was presented outside. If the Gabor stimulus was cued, the monkey was to indicate whether the orientation of the (Gabor) test stimulus matched the sample orientation. If the Gaussian was cued, the monkey was to indicate whether the color of the (Gaussian) test stimulus matched the sample color. For each neuron recorded, McAdams and Maunsell (1999a) systematically varied the orientation of the Gabor stimulus in the RF to see whether attention to the object would change the shape of the orientation tuning curve.

In the experiment of McAdams and Maunsell (1999a), the attended and unattended stimuli differed in both spatial location and relevant feature dimension (orientation or color). McAdams and Maunsell (1999a) chose this design to increase the chances of encountering attentional modulations in area V4. In terms of NTVA, both filtering and pigeonholing should be expected. Consider, first, the effect of filtering. The Gabor stimulus for a cell was adjusted in spatial frequency, color, and size to generate the strongest possible response using the match-to-sample task. We assume that (a) the optimized Gabor stimulus for a given cell was a more effective stimulus object for the cell than were possible noise objects including individual parts (such as individual bars) of the Gabor stimulus. We also assume that (b) activations in response to noise objects were either zero (responses at baseline caused by ghost objects formed by internal random noise) or approximately proportional to the activation caused by the experimental stimulus as a whole (responses caused by individual parts of the Gabor pattern with the same orientation as the whole pattern). Finally, we assume that (c) the instruction to respond to the Gabor stimulus increased the attentional weight of the experimental stimulus (the Gabor pattern as a whole) relative to the attentional weights of noise objects. By the first and third assumptions above, the instruction to respond to the Gabor stimulus should enhance the activation of the cell; by the second assumption, the enhancement should be a multiplicative scaling of the activation.

Next, consider the effect of pigeonholing. The requirement to report orientation instead of color should increase the perceptual bias in favor of categorizing objects with respect to orientation instead of categorizing objects with respect to color. Thus, β values should be high for orientations and low for colors. Following McAdams and Maunsell (1999a), we assume that the recorded neurons were predominantly selective for orientation such that the activation of a neuron scaled with the attentional emphasis (β value) on orientation rather than the emphasis on color. Accordingly, the pigeonholing due to the requirement to report orientation rather than color should also cause a multiplicative enhancement of the activation (multiplication by the β value for the orientation preferred by the cell regardless of the orientation of the stimulus).

Finally, consider the combined effects of filtering and pigeonholing. Given that both filtering and pigeonholing scaled the activation of a recorded neuron multiplicatively (with the same factor for all stimulus orientations), the combined effect of the two mechanisms of attention should also be a multiplicative scaling of the activation of the neuron (the firing rate minus the baseline rate) to stimuli in all orientations.

For each recorded neuron, McAdams and Maunsell (1999a) fitted the neuron's mean rates of firing at the tested orientations by a theoretical tuning curve that was a sum of a 1-D Gaussian and two constants, one at the level of the baseline firing of the neuron (the undriven activity found when the RF was empty) and one representing the activation caused by a Gabor stimulus in the least preferred orientation. For those neurons that yielded satisfactory fits, the standard deviation of the Gaussian function was interpreted as the tuning curve's width, and the mean of the function was interpreted as the neuron's preferred orientation. The normalized population tuning curves for all V4 neurons are shown in Figure 18A. As can be seen, the responses to the Gabor patterns were substantially enhanced by attention at all possible orientations. Specifically, the amplitude of the Gaussian component of the tuning curve (i.e., the effect of varying stimulus orientation) and the activation caused by stimuli in the least preferred orientation were both enhanced by attention; however, the width of the Gaussian component of the tuning function and the (undriven) baseline firing were unaffected by attention. On the basis of the same data, Figure 18B shows a plot of the attended response against the unattended response at each of the 12 orientations. A strikingly good fit is provided by a straight line with a slope of 1.32 through the point representing the baseline firing rate (undriven activity) in both conditions. The goodness of fit strongly suggests that the effect of attention on the mean rate of firing of a recorded V4 neuron was a proportional scaling (multiplicative enhancement by a factor of about 1.32) of the activation of the neuron (the firing rate minus the baseline rate) to stimuli in all orientations.

The finding that the (undriven) baseline firing was unaffected by attention contrasts with findings by Luck et al. (1997) and Reynolds et al. (1999). This contrast is discussed in the *Attentional Effects on Baseline Firing* section. The finding that the width of the tuning curve was unaffected by attention contrasts with the results of Haenny and Schiller (1988) and Spitzer et al. (1988). McAdams and Maunsell (1999a) explained this discrepancy by different definitions of width. Both of the earlier studies measured tuning-curve width at a given fraction of the peak of the curve, such as the width at half height, where height was measured relative to undriven activity or zero activity. In measuring the width as the standard deviation of just the Gaussian component of the tuning function, McAdams and Maunsell's (1999a) procedure corresponds to measuring height relative to the response to the least preferred orientation rather than to undriven activity or zero activity. If McAdams and Maunsell (1999a) had measured height relative to undriven activity or zero activity, the width of the tuning curve would have seemed to change with attention.

McAdams and Maunsell (1999b) presented further analyses of the data considered above. They tested the hypothesis that attention changes the signal-to-noise ratio in the neuron's firing by decreasing the noise component. Specifically, when the firing rate is scaled up by attention, the variability in the neuron's response (i.e., the noise) might be reduced relative to the mean response (i.e., the signal). Such an increase in the reliability of the response would improve stimulus discrimination. However, McAdams and Maunsell (1999b) found no systematic change in the relation between response magnitude and response variance with attention to the stimulus. Instead, they pointed out that higher firing rates in themselves produce a better signal-to-noise ratio, even with a constant mean–variance ratio. When the firing rate (signal) is increased, the variance tends to increase proportionately. The standard deviation, being only the square root of the variance, increases less rapidly. Therefore, when the noise is measured by the standard deviation, the signal-to-noise ratio generally increases with the firing rate.

McAdams and Maunsell (1999b) also tested the notion that attention affects the temporal pattern of firing. In particular, attention might make neurons more likely to fire in bursts, which is more effective at driving other neurons and might increase the information transmitted by the neuron. However, attention did not change the rate of bursting, the number of spikes within each burst, or the length of each burst (when corrected for the increase in firing rate, which decreases the interspike interval). Overall, McAdams and Maunsell (1999b) concluded that the only systematic effect of attention in their study was a general upscaling of the activation (firing rate minus baseline). The qualitative pattern of firing did not change systematically, either by narrowing of the tuning curve or by a decrease in the variability of responses.

McAdams and Maunsell (2000). McAdams and Maunsell (2000) modified their experimental design to separate effects of spatial attention (filtering) from effects of feature-based attention (pigeonholing). One experimental condition was similar to the experiment of McAdams and Maunsell (1999a): The stimuli outside the RF of the recorded neuron were Gaussians, the stimuli inside the RF were Gabors, and either the Gaussians outside the RF or the Gabors inside the RF were to be attended. In this combined space-and-feature attention task (filtering by spatial locations inside or outside the RF combined with pigeonholing by orientation or color, respectively), the raw firing rates were 54% higher (median across neurons) when a monkey attended to the orientation of the Gabor in the RF than when the monkey attended to the color of the Gaussian outside the RF.

The other experimental condition was similar, but both the stimuli inside the RF and the stimuli outside the RF were Gabors (cf. Experiment 1 of Treue & Martínez-Trujillo, 1999). In this spatial attention task (filtering by spatial locations inside vs. outside the RF), firing rates were 31% higher when the monkey attended to the orientation of the Gabor in the RF than when the monkey attended to the orientation of the Gabor outside the RF.

A direct measure of the effect of pigeonholing (by orientation vs. color) was obtained by comparing firing rates to the Gabor in the RF when the monkey attended the orientation (of a Gabor) outside the RF against firing rates when the monkey attended the color (of a Gaussian) outside the RF (cf. Experiment 2 of Treue & Martínez-Trujillo, 1999). By this comparison, pigeonholing by orientation rather than color increased the firing rates by 11%.

Attentional Effects With a Relatively Simple and Strong Stimulus in the RF

Single-cell studies with only one stimulus in the RF of the recorded neuron have shown relatively small but fairly consistent effects of attention for stimuli that are faint (see the *Attentional Effects With a Faint Stimulus in the RF* section above) or complex (see the *Attentional Effects With a Complex Stimulus in the RF* section above). Results from studies with a single, relatively simple and strong stimulus in the RF have been less consistent. Some studies have shown no effects of attention. Using 200-ms exposures of bars of various colors, orientations, and sizes, Moran and Desimone (1985) found no effect of attention when only one stimulus was present in the RF of a recorded neuron in area V4 or the IT cortex. Similarly, using 50-ms presentations of bars (rectangles) followed by blank intervals of at least 300 ms, Luck et al. (1997) found essentially no effect of attention when only one

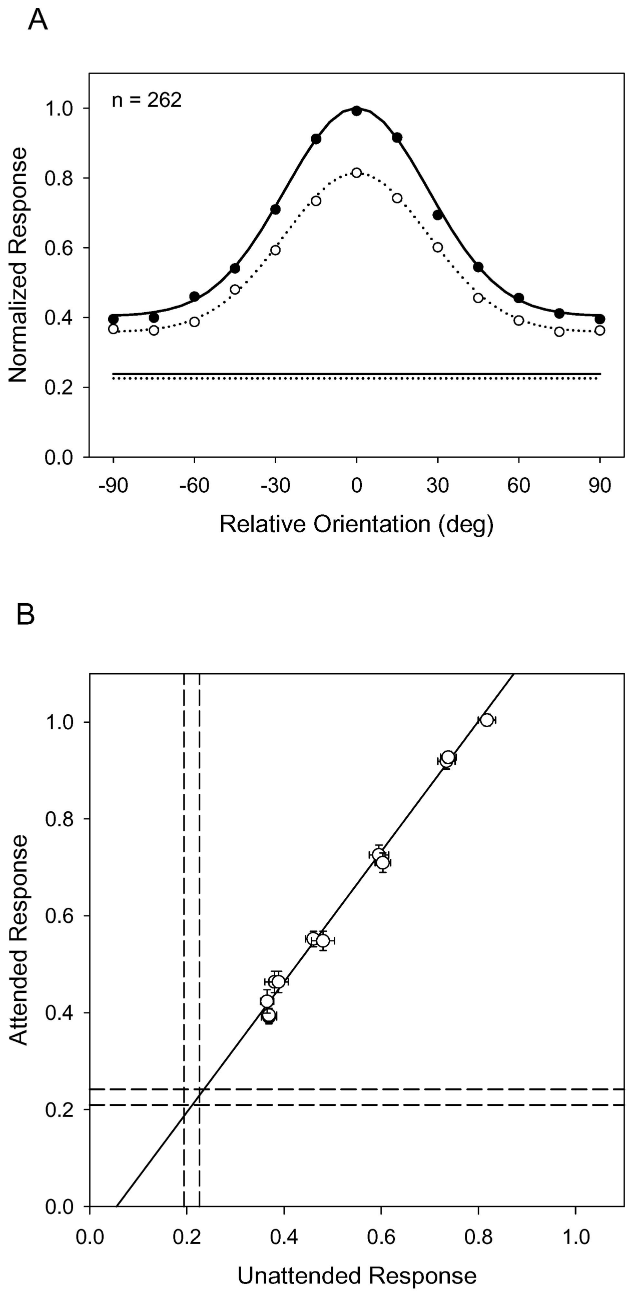

Figure 18. Effects of attention on mean rates of firing in the experiment of McAdams and Maunsell (1999a). A: Normalized population tuning curves for all V4 neurons. Solid circles fitted by a solid Gaussian curve show the normalized response as a function of the angular deviation between the stimulus and the preferred orientation when a monkey was attending to the stimulus inside the receptive field (RF). The solid horizontal line represents the undriven activity, measured as the mean firing rate during the fixation period before stimulus presentation in the same attention condition. Corresponding data for the condition in which the monkey was attending to the stimulus outside the RF are shown by open circles, the dotted Gaussian curve, and the dotted horizontal line. B: The attended response versus the unattended response for each of the tested orientations. The results are fitted by a least-squares line with a slope of 1.32. The pairs of dashed lines show undriven activity plus or minus 1 standard error. The strikingly close fit and the finding that the line very nearly passes through the point where attended response = unattended response = undriven activity show that attention very nearly effected a multiplicative scaling of the activation (the total firing rate minus the level of undriven activity). From "Effects of Attention on Orientation-Tuning Functions of Single Neurons in Macaque Cortical Area V4," by C. J. McAdams and J. H. R. Maunsell, 1999, *Journal of Neuroscience, 19,* Figure 7, p. 437. Copyright 1999 by the Society for Neuroscience. Adapted with permission.

stimulus was present in the RF of a recorded neuron in areas V1, V2, or V4. Such null effects of attention are readily explained by NTVA by assuming that the attentional weights of noise objects were negligibly small relative to the attentional weight of the experimental stimulus: The weight of the experimental stimulus should be much higher than the weights of ghost objects formed by internal random noise, because the experimental stimulus was high in contrast; and the weight of the experimental stimulus should be much higher than the weights of any individual parts of the stimulus, because the experimental stimulus was too simple to have any noteworthy parts.

However, other studies with a single, relatively simple and strong stimulus in the RF of each recorded neuron have shown significant effects of attention. With bars as stimuli, Motter (1993) found both positive and negative effects of spatial attention on firing rates of neurons in V1, V2, and V4; Motter (1994a, 1994b) found attentional enhancement of firing rates of V4 neurons in a task that required filtering by a nonspatial category (cf. Chelazzi et al., 1998, 2001); and Connor et al. (1997) found evidence of dynamic remapping of RFs of V4 neurons with changes in spatial attention. These findings are analyzed in the three subsections below. The analyses show that the main findings can be accounted for in terms of NTVA if it is assumed that although the experimental stimuli were relatively simple and strong, attentional weights of noise objects were noticeable relative to the weight of the experimental stimuli.

Motter (1993). Motter (1993) showed that spatial attention can modulate responses of neurons as early in the (macaque) visual system as V1, as well as in V2 and V4. Most important in the present context, Motter's results also suggest that such modulation can be both positive and negative with only one stimulus in the RF. In Motter's experiment, a monkey's attention was cued to a particular location on a display screen. The cuing procedure was as follows. The monkey was shown a circular array of small dots centered on fixation. After 400–1,000 ms, all but one of these dots disappeared. The remaining (cue) dot remained onscreen for 200–400 ms at the location where the target stimulus would appear. The cued location was either inside or outside the RF of the neuron being recorded. Immediately after the cue stimulus disappeared, from three to eight bars with different orientations were displayed, only one of them located in the RF. Thus, the stimulus in the RF was sometimes attended and sometimes unattended. The monkey was required to make a discrimination of the orientation of the bar at the cued location. More than one third of the recorded neurons in each area (V1, V2, and V4) showed significant differences between firing rates in the attended and unattended conditions. Remarkably, a substantial portion of the neurons had lower firing rates when the object in the RF was cued compared with when it was unattended. This was the case for 30% of the neurons in V1 and V2 and about half of the neurons in V4.

Motter's (1993) findings of both increases and decreases in mean firing rates when attention was directed to RF stimuli are reminiscent of the findings obtained in studies with multiple stimuli in the RF. When both an effective and an ineffective sensory stimulus are present in the RF of a recorded neuron, the firing rate increases when attention is directed to the effective sensory stimulus, whereas the firing rate decreases when attention is directed to the ineffective sensory stimulus (e.g., Moran & Desimone, 1985). Motter's (1993) findings can be explained by assuming that although his experimental stimuli were relatively simple and strong, there were noise objects with appreciable attentional weights that stimulated some of the recorded neurons more effectively than did the experimental stimuli. Although a bar is a relatively simple object, a bar has individual parts, such as edges, and it is possible to attend to a particular edge rather than attending to the bar as a whole. Thus, it seems plausible that individual edges could have appreciable attentional weights. It also seems plausible that for some neurons (e.g., *bar detectors* in V1), the bar was a more effective sensory stimulus than was an individual edge of the bar, but for other neurons (e.g., *edge detectors* in V1), a particular edge of the bar was a more effective stimulus than was the bar as a whole. Hence, when the attentional weight of the bar in the RF of the recorded neuron was increased in relation to the weights of individual edges of the bar (the *attended* condition), the expected firing rate increased in some neurons (e.g., bar detectors in V1) but decreased in others (e.g., edge detectors in V1, responding strongly to a particular edge but weakly to any other objects).

Motter (1994a, 1994b). Motter (1994a, 1994b) modified his experimental design to investigate attentional selection by color. He recorded from neurons in area V4, most of which were selective to both orientation and color. A monkey was required to select a bar stimulus on the basis of color (or luminance) and to report its orientation. First, the monkey fixated a cue stimulus that showed the target color of the trial. Then, the monkey was presented with an array of from four to six colored bars, only one of them located in the RF of the neuron being recorded. Initially, there were several possible targets (i.e., objects matching the color cue) in the display. At this stage, the monkey's attention was presumably uniformly distributed across all objects with colors matching the cue once attentional weights had been computed and applied (cf. the *Filtering by Nonspatial Categories* section above). Finally, after a period of 1,500–2,700 ms, all possible targets but one were deleted, and the monkey reported the orientation of the remaining target stimulus.

For a large majority of the recorded V4 neurons, the response to the object in the RF was significantly stronger when the cued color matched the color of the object. A relative increase in cases of nonmatch was not seen in any neuron. Thus, whereas Motter (1993) found both increases and decreases in firing rates of V4 neurons with attention to the RF stimulus, Motter (1994a, 1994b) found only increases. The cause of this discrepancy is not clear. However, the fact that attentional modulations occurred can be explained by assuming that although the experimental stimuli were relatively simple and strong, weights of noise objects were not negligible, so the probability that a V4 neuron responded to the bar in its RF instead of responding to noise objects increased with the attentional weight of the bar in the RF.

The difference in response to matching and nonmatching stimuli began 150–200 ms after stimulus onset, continued to rise until 500 ms, and remained stable for the remainder of a trial. Thus, as in the study by Chelazzi et al. (1998), the wave of unselective processing (before attentional weights had been computed and applied) seemed to take ~150–200 ms, which is two to three times the standard visual latency observed in area V4 (Motter, 1994b, p. 2195). If the cue that showed the target color was changed during the trial, the effect could be reversed over the course of 150–300 ms (Motter, 1994b).

Connor, Preddie, Gallant, and Van Essen (1997). In a study of V4 neurons, Connor et al. (1997; see also Connor, Gallant, Preddie, & Van Essen, 1996) found an interesting effect of spatial attention. In their interpretation, the results suggested that attentional enhancement of neural responses spreads from an attended object to behaviorally irrelevant objects at nearby locations, as though the attended object was illuminated by a diffuse *spotlight of attention* (cf. Eriksen & Yeh, 1985; Posner, Snyder, & Davidson, 1980). The basic task was as follows. A monkey was shown a central fixation point and an array of ring stimuli. After fixation was achieved, the monkey depressed a lever, and 500 ms later a target ring appeared. The delayed onset indicated that this was the target object. The monkey was to monitor the target ring continuously for up to 4,500 ms and respond to the deletion of a 90° section anywhere along the ring's circumference. Changes in the distractor rings were to be ignored. The target ring was always placed slightly outside the RF of the neuron being recorded. At the same time, behaviorally irrelevant bar stimuli (with optimal values of color, orientation, and width for the cell's response) were flashed inside the RF. The bars were displayed one at a time, with the first appearing 1,000 ms after the target ring (by which time spatial attention should long have taken effect). The bar stimuli were displayed for 150 ms, and a new stimulus appeared every 1,000 ms until completion of the trial. The bars were shown at varying distances from the attended ring. For example, in the *4 ring/5 bar* experiment, five locations were probed.

The mean rate of firing in response to a bar at a given location increased as the attended ring was moved closer to the bar. No cells showed the opposite effect. For example, in the 4 ring/5 bar experiment, the average cell shifted 16% of its total response profile from one half of the RF to the other half as attention was directed from one side to the other. Analyzed in another way, the RF position in which the mean response was strongest shifted 0.1 of RF diameter on average, depending on the position of the attended ring. In a variation of the experiment (*2 ring/7 bar*), even larger response shifts were found. In general, it seemed that the RF was remapped such that the most responsive part (the *hot spot*) moved in the direction of the attended object.

This finding can be explained in NTVA by assuming that although any bar flashed in the RF of the recorded neuron was a relatively strong stimulus, the weight of ghost objects (pure noise) in the RF was noticeable compared with the weight of the bar. Because the monkey's task was to monitor the target ring for deletion of a section anywhere along the ring's circumference, the monkey ascribed attentional weights on the basis of location, so objects in the immediate vicinity of the target ring got high attentional weights. (For optimal performance, the location being monitored should probably be a ring-shaped area extending somewhat beyond the edges of the target ring.) Because the target ring was close to the border of the RF of the recorded neuron, bars flashed inside the RF got higher attentional weights the closer they were to the target ring. Hence, when the target ring was moved closer to the bar flashed in the RF, the probability that the recorded neuron responded to the bar (with the color, orientation, and width preferred by the neuron) instead of responding to a noise object increased, so the mean rate of firing also increased.

Like our analyses of the studies by Motter (1993, 1994a, 1994b), our analysis of the study by Connor et al. (1997) shows that the main findings can be accounted for in terms of NTVA if it is assumed that although the experimental stimuli were relatively simple and strong, attentional weights of noise objects were noticeable relative to the weight of the experimental stimuli. It is not clear why effects of noise stimuli were negligibly small in the studies of Moran and Desimone (1985) and Luck et al. (1997) but noticeable in the studies of Motter (1993, 1994a, 1994b) and Connor et al. (1997). However, it seems clear that when only one stimulus is present in the RF of a recorded neuron, attentional effects are much smaller than they are when multiple stimuli are present. It also seems clear that the effects depend on the complexity and the strength of the RF stimulus. Studies with a single stimulus in the RF of the recorded neuron have shown consistent effects of attention for stimuli that are faint (see the *Attentional Effects With a Faint Stimulus in the RF* section above) or complex (see the *Attentional Effects With a Complex Stimulus in the RF* section above). With a single, relatively simple and strong stimulus in the RF, effects of attention have been less consistent (sometimes noticeable and significant, sometimes not). When effects of attention have been found, they have generally conformed to expectations from NTVA.

Attentional Effects on Baseline Firing

With no stimulus in its RF, a cell fires at baseline level. The baseline firing rate has been found to depend on the attentional state of the organism. Many investigators have reported that when a target is expected to appear inside a recorded cell's RF, the baseline firing rate is increased (Chelazzi et al., 1998; Fuster, 1990; Luck et al., 1997; Miller, Erickson, & Desimone, 1996; Miller, Li, & Desimone, 1993; Miyashita & Chang, 1988; Rainer, Rao, & Miller, 1999; Reynolds et al., 1999, 2000). The baseline shift is usually interpreted as being a result of top-down signals that prepare the organism for processing of an upcoming stimulus (e.g., Desimone, 1999).

In NTVA, the baseline shift reflects the fact that a mental image is held in VSTM such that inner-driven activity is added to the undriven activity. Use of a more or less schematic mental image of the target seems plausible in tasks like delayed match-to-sample (McAdams & Maunsell, 1999a; Miller et al., 1993, 1996), detection of a target in a sequence of displays (Luck et al., 1997; Reynolds et al., 1999, 2000), or standard visual search (Chelazzi et al., 1998, 2001). The mental image may be bottom-up generated (from a stimulus presentation) or top-down generated (from long-term memory; cf. the *Mental Images* section above).

Miller, Li, and Desimone (1993). Miller et al. (1993) recorded from neurons in anterior IT cortex while monkeys performed a delayed match-to-sample task. Consistent with previous studies of IT neurons (Fuster, 1990; Miyashita & Chang, 1988), Miller et al. found increased baseline activity during the retention interval after the sample display (*delay activity*). The baseline shifts were stimulus specific: A neuron fired more strongly if the target object was a preferred stimulus for the neuron (such that the cell would normally respond to the target when it was actually presented). These findings are readily interpreted as reflecting a mental image of the sample stimulus. However, Miller et al. tested the effect of presenting several intervening stimuli between the sample and the probe stimulus. This procedure revealed that the stimulus-specific baseline shift was eliminated after the first intervening (nonmatching) stimulus. Therefore, the baseline shift in IT cells could not have been maintaining the memory of the sample stimulus throughout the trial. Instead, this memory must be represented in a different part of the brain; the following study points to PF cortex.

Miller, Erickson, and Desimone (1996). Miller et al. (1996) extended the investigation of delay activity to neurons in PF cortex. They also used a match-to-sample task with multiple items intervening between the sample and the probe. Unlike IT neurons (which were also tested in this study), the PF neurons continued to respond with stimulus-specific delay activity across intervening objects. One way to interpret this finding is to assume that immediately after the presentation of the sample stimulus, the sample was represented by a comparatively concrete mental image (in PF, IT, and possibly lower visual areas), but later on only representations of comparatively abstract features of the sample were kept active (in PF). This might help in the setting of attentional parameters after the sample presentation but free the visual system (up to IT) to process the stimuli later in the sequence.

In a further exploration of delay activity in PF, Rainer et al. (1999) showed that neurons can code for an anticipated stimulus that is different from the presented sample (i.e., generate a mental image from long-term memory). Rainer et al. used a task in which the sample was to be matched not with an identical stimulus but with a different, paired associate stimulus. Initially after presentation of the sample, the baseline shift in PF reflected the sample, but soon it changed to reflect the features of the paired associate.

Luck, Chelazzi, Hillyard, and Desimone (1997). Luck et al. (1997) studied V1, V2, and V4 neurons when attention was directed to a specific location. The design was similar to that used by Reynolds et al. (1999, Experiment 2, 2000). The task was to monitor a target location in a sequence of displays, reacting when a target object was shown. Attention was directed to the target location by means of instruction trials, and the monkey then had to remember the location for a block of trials. Stimuli could also appear at another location, but this was irrelevant to the task. The target object, a square, was the same throughout the whole study. Before the target appeared, the monkey was presented with from one to six brief displays, each containing either one or two distractors in the attended or the unattended location. The stimuli were rectangles of different orientations and colors, some of which were effective and some of which were ineffective at driving the cell.

One main result of Luck et al.'s (1997) study has already been mentioned: a confirmation of the filtering mechanism discovered by Moran and Desimone (1985; see the *Filtering by Location* section above). Another important finding was a change in baseline firing with attention. Luck et al. found that 54% of the V4 neurons had significantly higher firing rates during the last 100 ms before stimulus exposure when attention was directed to stimuli inside their RFs rather than being directed to stimuli outside their RFs. Seventy-five percent of V2 neurons showed the same pattern, whereas V1 neurons were unaffected. The increase in baseline firing rate was large, about 30%–40%. This effect was also found when location marker boxes were removed from the display, so the effect could not have been due to a sustained sensory response to these background stimuli. Instead, the increase in baseline firing appeared to have been due to top-down input to the cells.

Luck et al. (1997) dismissed the idea that the top-down input reflected an internal template or mental image of the target stimulus. They argued that if such were the case, the baseline shift should be found only when the target stimulus was an effective stimulus for the cell but not when it was an ineffective stimulus (cf. Chelazzi et al., 1998). Only eight cells were tested for this effect, but their responses generally pointed to an increase in baseline firing rate independent of whether the target stimulus was a preferred stimulus for the cell. Luck et al. (1997) suggested that it was the direction of attention into the RF per se that caused the shift in baseline firing, but they also mentioned the possibility that the baseline shift reflected "a memory that specifies only the location of the target" (p. 36).

Within the framework of NTVA, the hypothesis that the baseline shift in the study by Luck et al. (1997) manifested a mental image representing just the target location seems highly plausible. Presumably, stimulus selection (filtering) was based mainly on spatial location, and the main components of attentional weights were based on η values computed by matching the stimuli against a neural representation of the target location. It seems plausible that this neural representation was a mental image—that is, a representation kept in VSTM. The fact that the target object (square) was kept constant throughout the experiment presumably eliminated any need to keep an image specifying the color and shape of the target in VSTM (cf. Li, Miller, & Desimone, 1993; Miller et al., 1993).

Using similar experimental paradigms, Reynolds et al. (1999, 2000) also found baseline shifts when attention was directed into the RF. The effect was not further characterized in these studies, but we assume the results can be explained in the same way as the data of Luck et al. (1997).

McAdams and Maunsell (1999a). In McAdams and Maunsell's (1999a) match-to-sample task (see the *Attentional Effects With a Complex Stimulus in the RF* section above), no shift in the baseline activity of V4 neurons was observed when attention was directed to a location in their RFs. However, McAdams and Maunsell only measured the response before the sample was presented. It is possible that the monkey used a mental image in the delay interval after sample presentation as preparation for the upcoming matching task with the probe stimulus.

Chelazzi, Duncan, Miller, and Desimone (1998) versus Chelazzi, Miller, Duncan, and Desimone (2001). Recall from the *Filtering by Nonspatial Categories* section that Chelazzi et al. (1998) found a baseline shift in IT neurons, but Chelazzi et al. (2001) found no such change in V4 neurons, although they used very similar experiments. The baseline shift was stimulus specific: An IT cell fired more strongly if the target object was a preferred stimulus for the cell (cf. Miller et al., 1993). This happened even though the target's location was not known in advance. The target was only expected to fall somewhere inside the RF. However, if the monkey used a mental image to retain the cue stimulus, it presumably imagined the target at a particular location inside the large RF of the IT neuron, possibly at the central location where the cue stimulus had been shown. This may explain why Chelazzi et al. (2001) found no baseline effect in V4: In this case, the central location of the cue stimulus fell outside the RF of the recorded neuron. The same was the case for the cue stimulus in Motter's (1994a) study of V4 neurons, which also showed no cue-specific baseline shift.

Note that by a generalization of the argument presented above, attentional effects on baseline firing should be more widespread at higher than at lower levels of the visual system. Because RFs are larger at higher levels than at lower levels, the likelihood that an imagined object is located in the RF of a randomly chosen neuron is greater at the higher levels. Therefore, imagining an object in the visual field is likely to involve a higher proportion of the cells at higher than at lower levels of the visual system.

The fact that the two studies by Chelazzi et al. (1998, 2001) showed similar effects of attention in the recorded cells in IT and V4, but only the IT study showed baseline shifts, speaks against the notion that baseline shifts reflect "bias signals" needed for controlling the competition between stimuli in the RF (cf. Desimone, 1999). Instead, baseline shifts or mental images seem to have more indirect influences on the competition between stimuli for neural representation. In particular, the competition can be "biased" by attentional weighting on the basis of η values computed by matching the stimuli against a mental image of a particular type of target.

Summary

We tested the explanatory power of NTVA against 16 important studies in the single-cell visual attention literature. Using the filtering mechanism of NTVA, we could straightforwardly explain the strongest and most consistent effect in the literature: the change in firing rate when attention is reallocated across multiple stimuli in the same RF (see the *Attentional Effects With Multiple Stimuli in the RF* section above). The finding of linear weighting of mean responses to individual RF stimuli (e.g., Reynolds et al., 1999) followed readily from NTVA's Equation 1. The long unselective processing stage in experiments requiring nonspatial selection (see the *Filtering by Nonspatial Categories* section above) and the effect of varying contrast luminance between two RF stimuli (see the *Filtering of Stimuli With Different Contrast* section above) also fit closely with NTVA.

NTVA's filtering mechanism further accounted for many findings with a single experimental stimulus in the RF (see the *Attentional Effects With a Single Stimulus in the RF* section above). NTVA explained why filtering has little or no effect when the stimulus is sufficiently simple and strong. NTVA also explained how filtering becomes effective when the stimulus is faint (so that internal random noise must be considered) or complex (so that individual parts of the stimulus must be considered). In experimental conditions in which attention was directed to a particular feature across spatial locations, NTVA's pigeonholing mechanism could explain the modulation of firing rates. Both mechanisms are compatible with the common finding of multiplicative modulation of neural activation (across a stimulus dimension) with attention.

Finally, extant findings on shifts in baseline firing rates with attention (see the *Attentional Effects on Baseline Firing* section above) seemed to be in general agreement with NTVA's notion of representations in VSTM (mental images). Among other findings, NTVA explained why baseline shifts are more widespread at higher than at lower levels of the cortical visual system.

Conclusion

Being a neural interpretation of TVA, NTVA provides quantitative accounts of human performance (reaction times and error rates) in a broad range of experimental paradigms of single-stimulus recognition and attentional selection from multiobject displays. By use of the same basic equations as TVA, NTVA also accounts for a broad range of attentional effects observed in firing rates of single cells in the primate visual system. Thus, NTVA provides a mathematical framework to unify the two fields of research.

References

Albrecht, D. G., & Hamilton, D. B. (1982). Striate cortex of monkey and cat: Contrast response functions. *Journal of Neurophysiology, 48,* 217–237.

Allport, D. A. (1977). On knowing the meaning of words we are unable to report: The effects of visual masking. In S. Dornic (Ed.), *Attention and performance VI* (pp. 505–534). Hillsdale, NJ: Erlbaum.

Anderson, C. H., & Van Essen, D. C. (1987). Shifter circuits: A computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences, USA, 84,* 6297–6301.

Bricolo, E., Gianesini, T., Fanini, A., Bundesen, C., & Chelazzi, L. (2002). Serial attention mechanisms in visual search: A direct behavioral demonstration. *Journal of Cognitive Neuroscience, 14,* 980–993.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica, 50,* 253–290.

Bundesen, C. (1982). Item recognition with automatized performance. *Scandinavian Journal of Psychology, 23,* 173–192.

Bundesen, C. (1987). Visual attention: Race models for selection from multielement displays. *Psychological Research/Psychologische Forschung, 49,* 113–121.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547.

Bundesen, C. (1991). Visual selection of features and objects: Is location special? A reinterpretation of Nissen's (1985) findings. *Perception & Psychophysics, 50,* 87–89.

Bundesen, C. (1993). The relationship between independent race models and Luce's choice axiom. *Journal of Mathematical Psychology, 37,* 446–471.

Bundesen, C. (1996). Formal models of visual attention: A tutorial review. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 1–43). Washington, DC: American Psychological Association.

Bundesen, C. (1998a). A computational theory of visual attention. *Philosophical Transactions of the Royal Society of London, Series B, 353,* 1271–1281.

Bundesen, C. (1998b). Visual selective attention: Outlines of a choice model, a race model, and a computational theory. *Visual Cognition, 5,* 287–309.

Bundesen, C., & Habekost, T. (2005). Attention. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 105–129). London: Sage.

Bundesen, C., & Harms, L. (1999). Single-letter recognition as a function of exposure duration. *Psychological Research/Psychologische Forschung, 62,* 275–279.

Bundesen, C., Kyllingsbæk, S., & Larsen, A. (2003). Independent encoding of colors and shapes from two stimuli. *Psychonomic Bulletin & Review, 10,* 474–479.

Bundesen, C., Larsen, A., Kyllingsbæk, S., Paulson, O. B., & Law, I. (2002). Attentional effects in the visual pathways: A whole-brain PET study. *Experimental Brain Research, 147,* 394–406.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics, 33,* 487–493.

Bundesen, C., Pedersen, L. F., & Larsen, A. (1984). Measuring efficiency of selection from briefly exposed visual displays: A model for partial report. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 329–339.

Bundesen, C., Shibuya, H., & Larsen, A. (1985). Visual selection from multielement displays: A model for partial report. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 631–649). Hillsdale, NJ: Erlbaum.

Bushnell, M. C., Goldberg, M. E., & Robinson, D. L. (1981). Behavioral enhancement of visual responses in monkey cerebral cortex: I. Modulation in posterior parietal cortex related to selective visual attention. *Journal of Neurophysiology, 46,* 755–772.

Cave, K. R. (1999). The FeatureGate model of visual selection. *Psychological Research/Psychologische Forschung, 62,* 182–194.

Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22,* 225–271.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology, 80,* 2918–2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993, May 27). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345–347.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cerebral Cortex, 11,* 761–772.

Cohen, J. D., Romero, R. D., Farah, M. J., & Servan-Schreiber, D. (1994). Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience, 6,* 377–387.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22,* 319–349.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology, 75,* 1306–1308.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience, 17,* 3201–3214.

Corbetta, M., & Shulman, G. L. (1998). Human cortical mechanisms of visual attention during orienting and search. *Philosophical Transactions of the Royal Society of London, Series B, 353,* 1353–1362.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998, February 27). An area specialized for spatial working memory in human frontal cortex. *Science, 279,* 1347–1351.

Crick, F. (1984). Function of the thalamic reticular complex: The searchlight hypothesis. *Proceedings of the National Academy of Sciences, USA, 81,* 4586–4590.

Danziger, S., Ward, R., Owen, V., & Rafal, R. (2001). The effects of unilateral pulvinar damage in humans on reflexive orienting and filtering of irrelevant information. *Behavioral Neurology, 13,* 95–104.

Desimone, R. (1999). Visual attention mediated by biased competition in extrastriate cortex. In G. W. Humphreys, J. Duncan, & A. Treisman (Eds.), *Attention, space, and action* (pp. 13–31). Oxford, England: Oxford University Press.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In E. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). Amsterdam: Elsevier.

Desimone, R., Wessinger, M., Thomas, L., & Schneider, W. (1990).

Attentional control of visual perception: Cortical and subcortical mechanisms. *Cold Spring Harbor Symposia on Quantitative Biology, 55,* 963–971.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87,* 272–300.

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

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI: Information integration in perception and communication* (pp. 549–578). Cambridge, MA: MIT Press.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Chavda, S., & Shibuya, H. (1999). Systematic analysis of deficits in visual attention. *Journal of Experimental Psychology: General, 128,* 450–478.

Duncan, J., Bundesen, C., Olson, A., Humphreys, G., Ward, R., Kyllingsbæk, S., et al. (2003). Attentional functions in dorsal and ventral simultanagnosia. *Cognitive Neuropsychology, 20,* 675–701.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

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

Estes, W. K., & Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences, USA, 52,* 446–454.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neuroscience, 5,* 671–676.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36,* 193–202.

Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory. *Journal of Neurophysiology, 64,* 681–697.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121–134.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87,* 1–51.

Guillery, R. W., Feig, S. L., & Lozsádi, D. A. (1998). Paying attention to the thalamic reticulate nucleus. *Trends in Neurosciences, 21,* 28–32.

Habekost, T., & Bundesen, C. (2003). Patient assessment based on a theory of visual attention (TVA): Subtle deficits after a right frontal-subcortical lesion. *Neuropsychologia, 41,* 1171–1188.

Haenny, P. E., & Schiller, P. H. (1988). State dependent activity in monkey visual cortex: I. Single cell activity in V1 and V4 on visual tasks. *Experimental Brain Research, 69,* 225–244.

Hebb, D. O. (1949). *Organization of behavior.* New York: Wiley.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: Simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). *Psychological Review, 110,* 29–87.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 77–109). Cambridge, MA: MIT Press.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Perception & Psychophysics, 23,* 1–11.

Hume, D. (1896). *A treatise of human nature.* In L. A. Selby-Bigge (Ed.), *Hume's treatise.* Oxford, England: Clarendon Press. (Original work published 1739)

Humphreys, G. W., & Müller, H. J. (1993). SEarch via Recursive Rejection (SERR): A connectionist model of visual search. *Cognitive Psychology, 25,* 43–110.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Kanwisher, N., & Wojciulik, E. (2000). Visual attention: Insights from brain imaging. *Nature Reviews Neuroscience, 1,* 91–100.

Karnath, H., Himmelbach, M., & Rorden, C. (2002). The subcortical anatomy of human spatial neglect: Putamen, caudate nucleus and pulvinar. *Brain, 125,* 350–360.

Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience, 23,* 315–341.

Keele, S. W. (1973). *Attention and human performance.* Pacific Palisades, CA: Goodyear.

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

Kyllingsbæk, S., Schneider, W. X., & Bundesen, C. (2001). Automatic attraction of attention to former targets in visual displays of letters. *Perception & Psychophysics, 63,* 85–98.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review, 96,* 101–124.

LaBerge, D., & Buchsbaum, M. S. (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. *Journal of Neuroscience, 10,* 613–619.

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1–20.

Larsen, A., & Bundesen, C. (1998). Effects of spatial separation in visual pattern matching: Evidence on the role of mental translation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 719–731.

Lee, D. K., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience, 2,* 375–381.

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology, 69,* 1918–1929.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Logan, G. D. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review, 103,* 603–649.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review, 109,* 376–400.

Logan, G. D. (2004). Cumulative progress in formal theories of attention. *Annual Review of Psychology, 55,* 207–234.

Logan, G. D., & Bundesen, C. (1996). Spatial effects in the partial report paradigm: A challenge for theories of visual spatial attention. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 35, pp. 243–282). San Diego, CA: Academic Press.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 575–599.

Logan, G. D., & Bundesen, C. (2004). Very clever homunculus: Compound stimulus strategies for the explicit task-cuing procedure. *Psychonomic Bulletin & Review, 11,* 832–840.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108,* 393–434.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush,

& E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York: Wiley.

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

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology, 77,* 24–42.

Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281.

Martínez-Trujillo, J. C., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron, 35,* 365–370.

McAdams, C. J., & Maunsell, J. H. R. (1999a). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *Journal of Neuroscience, 19,* 431–441.

McAdams, C. J., & Maunsell, J. H. R. (1999b). Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron, 23,* 765–773.

McAdams, C. J., & Maunsell, J. H. R. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *Journal of Neurophysiology, 83,* 1751–1755.

McGill, W. J. (1963). Stochastic latency mechanisms. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 309–360). New York: Wiley.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology, 10,* 309–325.

Mesulam, M.-M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28,* 597–613.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16,* 5154–5167.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience, 13,* 1460–1478.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action.* Oxford, England: Oxford University Press.

Miyashita, Y., & Chang, H. S. (1988, January 7). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331,* 68–70.

Moran, J., & Desimone, R. (1985, August 23). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782–784.

Moray, N. (1969). *Attention: Selective processes in vision and hearing.* London: Hutchinson.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70,* 909–919.

Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience, 14,* 2178–2189.

Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience, 14,* 2190–2199.

Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach.* Cambridge, MA: MIT Press.

Mozer, M. C. (2002). Frames of reference in unilateral neglect and visual perception: A computational perspective. *Psychological Review, 109,* 156–185.

Mozer, M. C., & Sitton, M. (1998). Computational modeling of spatial attention. In H. Pashler (Ed.), *Attention* (pp. 341–393). London: Psychology Press.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience, 5,* 1203–1209.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience, 13,* 4700–4719.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). Cambridge, England: Cambridge University Press.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Page, M. P. A. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23,* 443–467.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research, 40,* 1227–1268.

Petersen, S. E., Robinson, D. L., & Keys, W. (1985). Pulvinar nuclei of the behaving rhesus monkey: Visual responses and their modulations. *Journal of Neurophysiology, 54,* 867–886.

Petersen, S. E., Robinson, D. L., & Morris, D. (1987). Contributions of the pulvinar to visual spatial attention. *Neuropsychologia, 25,* 97–105.

Phaf, R. H., van der Heijden, A. H. C., & Hudson, P. T. W. (1990). SLAM: A connectionist model for attention in visual selection tasks. *Cognitive Psychology, 22,* 273–341.

Posner, M. I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M. I., Nissen, M. J., & Ogden, W. C. (1978). Attended and unattended processing modes: The role of set for spatial location. In H. L. Pick & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp. 137–157). Hillsdale, NJ: Erlbaum.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160–174.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience, 4,* 1863–1874.

Rafal, R. D., & Posner, M. I. (1987). Deficits in human visual spatial attention following thalamic lesions. *Proceedings of the National Academy of Sciences, USA, 84,* 7349–7353.

Rainer, G., Rao, S. C., & Miller, E. K. (1999). Prospective coding for objects in primate prefrontal cortex. *Journal of Neuroscience, 19,* 5493–5505.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319–343.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience, 19,* 1736–1753.

Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron, 37,* 853–863.

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron, 26,* 703–714.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code.* Cambridge, MA: MIT Press.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3,* 1199–1204.

Robinson, D. L., & Cowie, R. J. (1997). The primate pulvinar: Structural, functional, and behavioral components of visual salience. In M. Steriade, E. G. Jones, & D. A. McCormick (Eds.), *Thalamus* (Vol. 2, pp. 53–92). Amsterdam: Elsevier.

Robinson, D. L., & Petersen, S. E. (1992). The pulvinar and visual salience. *Trends in Neurosciences, 15,* 127–132.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2003). Taking the MAX from neuronal responses. *Trends in Cognitive Sciences, 7,* 99–102.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7,* 191–218.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5,* 631–632.

Sapir, A., Rafal, R., & Henik, A. (2002). Attending to the thalamus: Inhibition of return and nasal–temporal asymmetry in the pulvinar. *NeuroReport, 13,* 693–697.

Sato, T. (1988). Effects of attention and stimulus interaction on visual responses of inferior temporal neurons in macaque. *Journal of Neurophysiology, 60,* 344–364.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research, 77,* 23–30.

Schall, J. D., & Thompson, K. G. (1999). Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience, 22,* 241–259.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics, 31,* 160–168.

Schneider, W. X. (1995). VAM: A neuro-cognitive model for visual attention, control of segmentation, object recognition, and space-based motor action. *Visual Cognition, 2,* 331–375.

Sherman, S. M., & Guillery, R. W. (2001). *Exploring the thalamus.* San Diego, CA: Academic Press.

Shibuya, H. (1991). Comparison between stochastic models for visual selection. In J.-P. Doignon & J.-C. Falmagne (Eds.), *Mathematical psychology: Current developments* (pp. 337–356). New York: Springer-Verlag.

Shibuya, H. (1993). Efficiency of visual selection in duplex and conjunction conditions in partial report. *Perception & Psychophysics, 54,* 716–732.

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 591–600.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Shih, S.-I., & Sperling, G. (2002). Measuring and modeling the trajectory of visual spatial attention. *Psychological Review, 109,* 260–305.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33,* 5–42.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*(11, Whole No. 498).

Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica, 27,* 285–292.

Sperling, G., & Weichselgartner, E. (1995). Episodic theory of the dynamics of spatial attention. *Psychological Review, 102,* 503–532.

Spitzer, H., Desimone, R., & Moran, J. (1988, April 15). Increased attention enhances both behavioral and neuronal performance. *Science, 240,* 338–340.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research, 41,* 414–419.

Townsend, J. T., & Ashby, F. G. (1982). Experimental test of contemporary mathematical models of visual letter recognition. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 834–864.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, England: Cambridge University Press.

Townsend, J. T., & Landon, D. E. (1982). An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion. *Journal of Mathematical Psychology, 25,* 119–162.

Treisman, A. M. (1964a). The effect of irrelevant material on the efficiency of selective listening. *American Journal of Psychology, 77,* 533–546.

Treisman, A. M. (1964b). Verbal cues, language, and meaning in selective attention. *American Journal of Psychology, 77,* 206–219.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treisman, A. M., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15–48.

Treue, S., & Martínez-Trujillo, J. C. (1999, June 10). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575–579.

Treue, S., & Maunsell, J. H. R. (1999). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience, 19,* 7591–7602.

Tsotsos, J. K., Culhane, S. M., Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence, 78,* 507–545.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Usher, M., & Cohen, J. D. (1999). Short term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphreys, & A. Olson (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). Berlin, Germany: Springer-Verlag.

Usher, M., & Niebur, E. (1996). Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience, 8,* 311–327.

van der Heijden, A. H. C. (1981). *Short-term visual information forgetting.* London: Routledge & Kegan Paul.

van der Heijden, A. H. C. (1992). *Selective attention in vision.* London: Routledge.

van der Heijden, A. H. C. (2004). *Attention in vision: Perception, communication and action.* Hove, England: Psychology Press.

van der Heijden, A. H. C., La Heij, W., & Boer, J. P. A. (1983). Parallel processing of redundant targets in simple visual search tasks. *Psychological Research/Psychologische Forschung, 45,* 235–254.

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups. *Memory & Cognition, 10,* 396–404.

van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition, 16,* 341–348.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 419–433.

Zeki, S. (1993). *A vision of the brain.* Oxford, England: Blackwell.

Appendix A

Neural Computation of Attentional Weights

The physiological mechanisms by which the receptive field (RF) of a cortical neuron can be adjusted to a particular stimulus are not known, nor is the time course of the process. For simplicity and specificity, we make the following assumptions. Let x be a stimulus object at a particular location in the visual field, and consider a feature-j neuron in whose classical RF x is present during the wave of unselective processing. The *effective* RF of the neuron can vary in position and size within the boundaries of the classical RF. We assume that the neuron represents object x if, and only if, the effective RF fits the position and scale of object x. Let p_{xj} be the probability that the effective RF of the feature-j neuron fits the position and scale of object x such that the neuron represents object x rather than representing any other object in the classical RF of the feature-j neuron. We assume that (a) the effective RF is determined before the stimulus array is presented, and (b) the effective RF does not change during the wave of unselective processing. Given that a subject has no prior knowledge of the stimulus array, probability p_{xj} should be independent of the behavioral importance of object x. Disregarding cases of overlapping objects at the same scale, p_{xj} also should be independent of the number and the nature of other objects in the visual field.

Consider the set of all those feature-j neurons in whose classical RFs object x is present. Let the neurons in the set be numbered from 1 up to $c(x, j)$, and let $n_k(x, j)$ be the kth ($1 \leq k \leq c[x, j]$) feature-j neuron in the set. For convenience, define

$$N(x, j) = \{1, 2, \ldots, c(x, j)\},$$

so that the set of all those feature-j neurons in whose RFs x is present can be written as

$$\{n_k(x, j) | k \in N(x, j)\}.$$

Let $\eta_k(x, j)$ be the activation of neuron $n_k(x, j)$ when the neuron represents object x rather than representing any other object.

As illustrated in Figure 8 in the main text, the attentional weight of stimulus x is computed on the basis of activations of neurons representing stimulus x at different cortical levels (values of $\eta_k[x, j]$ for different features j and neurons k). Specifically, for each neuron $n_k(x, j)$ representing stimulus x, the activation $\eta_k(x, j)$ is weighted by multiplication with a nonnegative factor, π'_j, that reflects the importance of attending to objects with feature j, and the product of $\eta_k(x, j)$ and π'_j is added to similar products for other feature-j neurons representing stimulus x and for neurons representing stimulus x with respect to features other than j (features $i, k, l \ldots$). The sum of these products, s_x, is given by

$$s_x = \sum_{j \in R} \sum_{k \in M(x, j)} \eta_k(x, j) \pi'_j, \tag{A1}$$

where R is the set of all visual features, and $M(x, j)$ is the set of all those values of k for which $n_k(x, j)$ represents stimulus x (thus, $M[x, j] \subseteq N[x, j]$). The sum s_x is assumed to be stored as a level of activation in a unit (a neuron or a population of similar neurons) in a saliency map.

Consider the summation

$$\sum_{k \in M(x, j)} \eta_k(x, j)$$

over all values of k for which neuron $n_k(x, j)$ represents stimulus x. For any $k \in N(x, j)$, neuron $n_k(x, j)$ represents stimulus x (i.e., $k \in M[x, j]$) with probability p_{xj}, so the contribution from neuron $n_k(x, j)$ to the sum

$$\sum_{k \in M(x, j)} \eta_k(x, j)$$

equals $\eta_k(x, j)$ with probability p_{xj} and 0 with probability $1 - p_{xj}$. Accordingly, the expected contribution from neuron $n_k(x, j)$ to the sum equals $\eta_k(x, j)\, p_{xj}$, whence the expected value of the sum can be written as

$$\mathrm{E}[\sum_{k \in M(x, j)} \eta_k(x, j)] = \sum_{k \in N(x, j)} \eta_k(x, j)\, p_{xj}. \tag{A2}$$

Treating π' values as constants, Equations A1 and A2 imply that the expectation of s_x is given by

$$\begin{aligned}\mathrm{E}(s_x) &= \mathrm{E}[\sum_{j \in R} \sum_{k \in M(x, j)} \eta_k(x, j)\, \pi'_j] \\ &= \sum_{j \in R} \mathrm{E}[\sum_{k \in M(x, j)} \eta_k(x, j)]\, \pi'_j \\ &= \sum_{j \in R} \sum_{k \in N(x, j)} \eta_k(x,j)\, p_{xj}\, \pi'_j. \end{aligned} \tag{A3}$$

Consider processing of a set of stimuli that is homogeneous in the sense that for any pair of stimuli x and y and any feature j, $p_{xj} = p_{yj}$ (i.e., the probability that x is represented in a feature-j neuron in whose RF x is present equals the probability that y is represented in a feature-j neuron in whose RF y is present). For such a set of stimuli, there is a constant p_j such that for any stimulus x,

$$p_{xj} = p_j. \tag{A4}$$

By defining

$$\pi_j = p_j \pi'_j, \tag{A5}$$

$$\eta(x, j) = \sum_{k \in N(x, j)} \eta_k(x, j), \tag{A6}$$

and

$$w_x = \mathrm{E}(s_x), \tag{A7}$$

Equation A3 reduces to Equation 2 of TVA:

$$\begin{aligned} w_x &= \mathrm{E}(s_x) && \text{(by Equation A7)} \\ &= \sum_{j \in R} \sum_{k \in N(x,j)} \eta_k(x, j)\, p_{xj}\, \pi'_j && \text{(by Equation A3)} \\ &= \sum_{j \in R} \eta(x, j)\, \pi_j. && \text{(by Equations A4–A6)} \end{aligned}$$

(*Appendixes continue*)

Appendix B

Procedures for Mediate Perception

Visual identification of an object consists in making perceptual categorizations of the object. Encoding a categorization into visual short-term memory (VSTM) is one way of making the categorization (viz., making the categorization by *immediate perception*). However, mutually contradictory categorizations (e.g., "x is an A," "x is an H") can be made by immediate perception, and decision procedures for resolving contradictions (procedures for *mediate perception)* are needed. Below, we describe three types of procedures for mediate perception (procedures for making mediate perceptual categorizations): a simple (exponential) race procedure, a (Poisson) counter procedure, and a (Poisson) random-walk procedure. We also consider the ways that decision making by use of the procedures is affected when the number of neurons used to represent the object to be identified is increased by attentional filtering (distribution of processing resources in accordance with attentional weights). The decision procedures for mediate perception may possibly be executed by the so-called frontoparietal network (see, e.g., Corbetta & Shulman, 1998, 2002), but the description below is restricted to a general functional level.

Perceptual categorizations of an object are based on activations (v values) in the set of cortical neurons that represent the object. Generally speaking, the higher the attentional weight of an object, the larger the set of neurons representing the object (see the *Dynamic Remapping of RFs* section in the main text), and the more neurons used for representing an object to be identified, the better the perceptual categorizations of the object in terms of speed and accuracy. Thus, the effect of distributing processing resources (cortical neurons) among visual objects on the basis of attentional weights is to improve perceptual categorizations of behaviorally important objects at the expense of categorizations of less important ones.

More formally, the effect of increasing the set of cortical neurons that represent an object can be described as follows. A typical neuron behaves approximately as a Poisson generator. To a first approximation, the latency measured from an arbitrary point of time (corresponding to the starting time of a race) to the first firing of the neuron is exponentially distributed with a certain rate parameter, v (see, e.g., McGill, 1963; Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1997). In steady-state conditions, v is a constant equal to the mean rate of firing. A set of n independent neurons (Poisson generators) working in parallel may be regarded as a single Poisson generator with a rate parameter equal to the sum of the rate parameters of the n neurons. In essence, the type and amount of processing done by n independent, identical neurons working in parallel for 1 ms is the same as the type and amount of processing done by a single one of the neurons during a period of n ms. Other things equal, if the number of neurons representing an object is multiplied by n, so also is the speed at which the object is processed. Thus, if the number of neurons representing an object is multiplied by n, the accuracy of a perceptual categorization of the object based on t ms of processing reaches a level as high as the accuracy previously reached after nt ms of processing.

Consider a perceptual categorization task in which object x must be assigned to one of m mutually exclusive categories, $i_1, i_2, \ldots,$ or i_m. The task can be performed by connecting a feature-i_1 neuron, a feature-i_2 neuron, . . ., and a feature-i_m neuron representing object x to a winner-take-all (WTA) cluster for recording the neuron that fires first (the winner of the race). If the neurons are Poisson generators with rate parameters $v_1, v_2, \ldots,$ and v_m, respectively, the categorization is made in accordance with an *exponential race model* (Bundesen, 1987). In this case, the probability that x is classified as a member of category i_1 is given by the Luce choice rule, $v_1/(v_1 + v_2 + \ldots + v_m)$, and the time taken by the winner to complete the race is exponentially distributed with a rate parameter equal to $v_1 + v_2 + \ldots + v_m$.

If object x is allocated n independent, identical neurons for each of the m categories, the perceptual categorization can be done by connecting each of the feature-i_1 neurons to the unit for category i_1 in the WTA cluster, each of the feature-i_2 neurons to the unit for category i_2, and so on. Because the n neurons for a given category can be regarded as a single Poisson generator with a rate parameter n times as high as the rate parameter for each of the n individual neurons, the categorization is again made in accordance with an exponential race model. The probability that x is classified as a member of a given category is the same as before, but the time taken by the winner to complete the race is exponentially distributed with a rate parameter equal to $nv_1 + nv_2 + \ldots + nv_m$, so the race is speeded up by a factor of n.

Response criteria based on accumulation of evidence can also be used (cf. Bundesen & Harms, 1999). Instead of responding on the basis of the first spike (firing) arriving at the WTA cluster from the set of feature-i_1, feature-i_2, . . ., and feature-i_m neurons representing object x, the number of spikes arriving from the feature-i_1, feature-i_2, . . ., and feature-i_m neurons, respectively, can be counted, and x can be categorized as a member of category i_1, if, and only if, the count of feature-i_1 spikes is the first among the m counts to reach a categorization threshold at a value of r spikes. In case $r = 1$, the model is identical to the exponential race model described above. In case $r > 1$, the model is called a *Poisson counter model* (for reviews, see Luce, 1986; Townsend & Ashby, 1983, chap. 9; see also Logan, 1996). Because the latency measured from an arbitrary point of time (corresponding to the starting time of the counting race) to the rth spike from a Poisson generator is gamma distributed, the model is also called a *gamma race model* (Bundesen, 1987). By raising of the categorization threshold, speed can be traded for accuracy. However, if the threshold is kept constant at a value of r spikes, but the numbers of independent (and otherwise identical) feature-i_1, feature-i_2, . . ., and feature-i_m neurons representing object x are increased by a given factor, the categorization process is speeded up by the same factor without affecting the accuracy of the process.

A related possibility is to categorize x as a member of category i_1 if, and only if, the count of feature-i_1 spikes is the first among the m counts that is at least r spikes greater than any of the other $m - 1$ counts. Again, in case $r = 1$, the model is identical to the exponential race model. For $r > 1$, the model is a *random-walk model* with exponential interstep times (for reviews, see Luce, 1986; Townsend & Ashby, 1983, chap. 10; see also Bundesen, 1982; Logan, 2002; Logan & Gordon, 2001; Nosofsky & Palmeri, 1997; Ratcliff & McKoon, 1997). As before, by raising of the categorization threshold, speed can be traded for accuracy. If the threshold is kept constant at a value of r spikes, but the numbers of feature-i_1, feature-i_2, . . ., and feature-i_m neurons representing object x are increased by a given factor, the categorization process is speeded up by the same factor without affecting the accuracy of the process.

Received March 29, 2004
Revision received September 21, 2004
Accepted September 24, 2004 ■