

VISUAL MOTION PERCEPTION

Stephen Grossberg
Center for Adaptive Systems
Department of Cognitive and Neural Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University
677 Beacon Street, Boston, MA 02215

Invited chapter:
Encyclopedia of Human Behaviour 2nd Edition

Correspondence should be addressed to:
Professor Stephen Grossberg
Center for Adaptive Systems,
Department of Cognitive and Neural Systems, Boston University
677 Beacon Street, Boston, MA 02215
email: steve@bu.edu
fax: 617-353-7755

Running Title: Visual Motion Perception

Keywords:

motion integration, motion segmentation, motion capture, decision-making, aperture problem, feature tracking, formation, complementary computing, V1, V2, MT, MST, LIP, neural network

Acknowledgements

S. Grossberg was supported in part by CELEST, a National Science Foundation Science of Learning Center (NSF SBE-0354378) and the SyNAPSE program of DARPA (HRL Laboratories LLC subcontract #801881-BS under DARPA prime contract HR0011-09-C-0001).

Abstract

The brain carries out complementary motion integration and segmentation processes to compute unambiguous global motion percepts from ambiguous local motion signals. When, for example, an animal runs at variable speeds behind forest cover, the forest is an occluder that creates apertures through which fragments of the animal's motion signals are intermittently experienced. The brain groups these fragments into a trackable percept of the animal and its trajectory. Form and motion processes are needed to accomplish this using feedforward and feedback interactions both within and across cortical processing streams. All the cortical areas V1, V2, MT, and MST are involved in these interactions. Figure-ground processes in the form stream through V2, such as the separation of occluding boundaries of the forest cover from boundaries of the animal, select the motion signals which determine global object motion percepts in the motion stream through MT. Sparse, but unambiguous, feature tracking signals are amplified before they propagate across position and either select consistent motion signals, or are integrated with far more numerous ambiguous motion signals. Figure-ground and integration processes together determine the global percept. A neural model is used to clarify and organize many perceptual and brain data about form and motion interactions, including data about motion grouping across apertures in response to a wide variety of displays, and probabilistic decision making in parietal cortex in response to random dot displays.

Motion Integration and Segmentation

Aperture Problem. The brain's motion perception system solves the complementary problems of *motion integration* and *motion segmentation*. The former joins nearby motion signals into a single object, while the latter separates them into different objects. Wallach (1935; translated by Wuerger, Shapley and Rubin, 1996) first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: Given any real direction of motion, the perceived direction is perpendicular to the orientation of the line. This phenomenon was called the *aperture problem* by Marr and Ullman (1981). The aperture problem is faced by any localized neural motion sensor, such as a neuron in the early visual pathway, which responds to a moving local contour through an aperture-like receptive field. Only when the contour within an aperture contains features, such as line terminators, object corners, or high contrast blobs or dots, can a local motion detector accurately measure the direction and velocity of motion (Shimojo, Silverman and Nakayama, 1989). For example, when the aperture is rectangular, as during the barberpole illusion (Wallach, 1935), moving lines may appear to move in the direction parallel to the longer edges of the rectangle within which they move, even if their actual motion direction is not parallel to these edges. The brain must solve the aperture problem, despite the fact that every neuron's receptive field defines an "aperture," in order to detect the correct motion directions of important moving objects in the world. The examples of circular and rectangular apertures, or occluders, provide important cues about how the brain can often do this in the real world.

When an object moves behind multiple occluders, aperture ambiguities can again lead to non-veridical percepts of its real motion. Despite the fact that the object may be segmented into many visible parts by the occluders, the visual system can often integrate these parts into a percept of coherent object motion that crosses the occluders. Studying conditions under which the visual system can and cannot accomplish correct segmentation and integration provides important cues to the processes that are used by the visual system to create useful and predictive object motion percepts during normal viewing conditions.

Feature Tracking Signals. To solve the interlinked problems of motion integration and segmentation, the visual system uses the relatively few unambiguous motion signals arising from image features, called *feature tracking* signals, to select the ambiguous motion signals that are consistent with them, while suppressing the more numerous ambiguous signals that are inconsistent with them. For example, during the barberpole illusion, feature tracking signals from the moving line ends along the longer edges of the bounding rectangle of the barberpole compute an unambiguous motion direction. These feature tracking signals gradually propagate into the interior of the rectangle. This *motion capture* process selects the feature tracking motion direction from the possible ambiguous directions along the lines within the rectangle that are due to the aperture problem. Motion capture also suppresses the ambiguous motion signals (Ben-Av & Shiffrar, 1995; Bowns, 1996, 2001; Castet et al., 1993; Chey et al., 1997, 1998; Grossberg et al., 2001; Lorenceau & Gorea, 1989; Mingolla et al., 1992). When a scene does not contain any unambiguous motion signals, the ambiguous motion signals cooperate to compute a consistent object motion direction and speed (Grossberg et al., 2001; Lorenceau & Shiffrar, 1992).

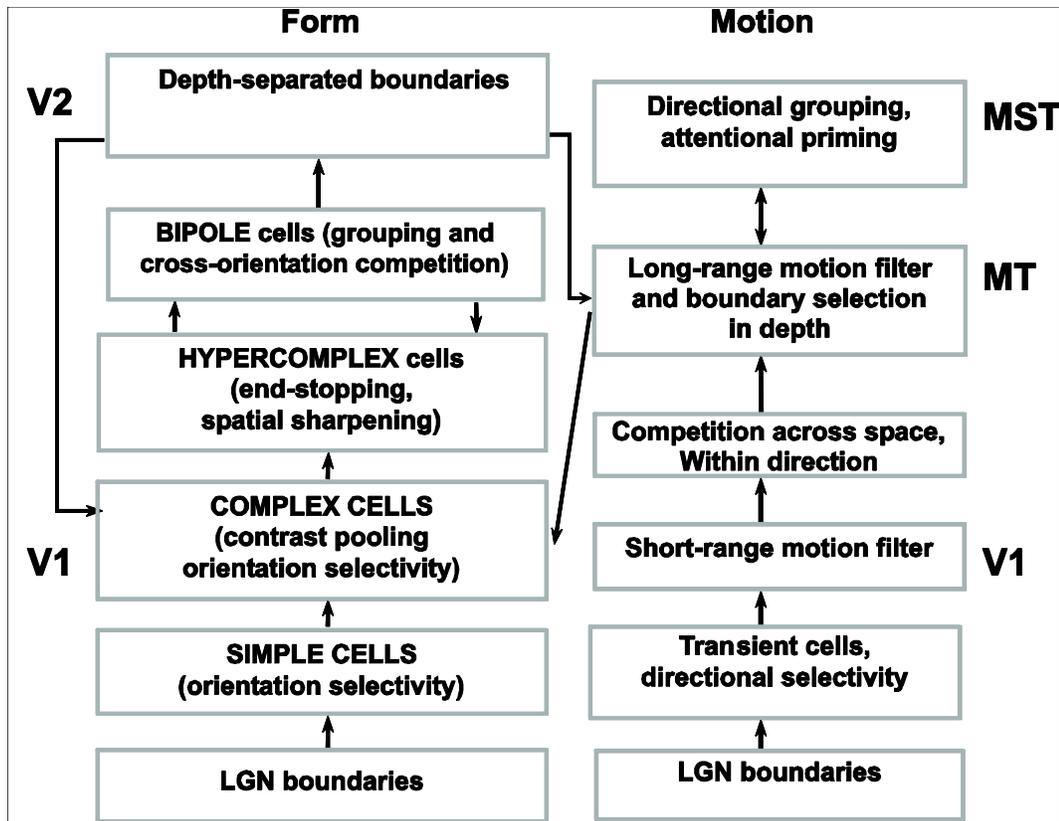


Figure 1. The 3D FORMOTION model processing stages. See text for details. [Reprinted with permission from Berzhanskaya et al. (2007)]

This chapter summarizes a neural model of the cortical form and motion processes that clarifies how such motion integration and segmentation processes occur (Figure 1). This 3D FORMOTION model has been progressively developed over the years to explain and predict an ever-larger set of data about motion perception; e.g., Baloch and Grossberg (1997), Berzhanskaya et al. (2007), Baloch et al. (1999), Chey, Grossberg, and Mingolla (1997, 1998), Grossberg et al. (2001), Grossberg and Pilly (2008), and Grossberg and Rudd (1989, 1992). Comparisons with related models and more data than can be summarized here are found in these archival articles.

Neurophysiological Support for Predicted Aperture Problem Solution. In addition to model explanations of known data, the model has predicted data that were subsequently reported. In particular, Chey, Grossberg, and Mingolla (1997) predicted how feature tracking estimates can gradually propagate across space to capture consistent motion directional signals, while suppressing inconsistent ones, in cortical area MT. Such motion capture was predicted to be a key step in solving the aperture problem. Pack and Born (2001) reported neurophysiological data that support this prediction. As simulated in the model, MT neurons initially respond primarily to the component of motion perpendicular to a contour's orientation, but over a period of approximately 60 ms the responses gradually shift to encode the true stimulus direction, regardless of orientation. Pack and Born also collected data which support the concept that motion signals are used for target tracking. Namely, the initial velocity of pursuit eye movements deviates in a direction perpendicular to local contour orientation, suggesting that the earliest neural responses influence the oculomotor response.

Many psychophysical data also illustrate how feature tracking signals can propagate gradually across space to capture consistent ambiguous signals. Castet, Lorenceau, Shiffrar, and Bonnet (1993) described a particularly clear illustration of this. Figure 2a summarizes their data. They considered the horizontal motion of both a vertical and a tilted line that are moving at the same speed. Suppose that the unambiguous feature tracking signals at the line ends capture the ambiguous motion signals near the line middle. The preferred ambiguous motion direction and speed are normal to the line's orientation. In the case of the vertical line, the speed of the feature tracking signals at the line ends equals that of the preferred ambiguous speed near the line's middle. For the tilted line, however, the preferred ambiguous speed is less than that of the feature tracking speed. If the speed of the line is judged using a weighted average of feature signals and ambiguous signals, then the tilted line will be perceived to move slower than the vertical line.

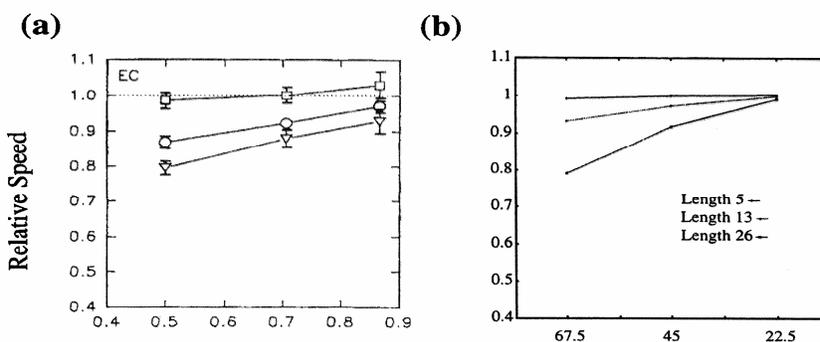


Figure 2. Effects of line length and orientation on perceived speed of horizontally moving lines. Relative perceived speed for three different line orientations and lengths are shown as percentages of the perceived speed of a vertical line of the same length. Part (A) shows data from Castet *et al.* (p. 1925). Each data line corresponds to a different line length (0.21, 0.88, and 1.76 deg). The horizontal axis shows the ratio of the speed normal to the line's orientation relative to the actual translation speed. The three data points from left to right for each line length correspond to line angles of 60, 45, and 30 deg from vertical, respectively. The horizontal dotted line indicates a veridical speed perception; results below this line indicate a bias toward the perception of slower speeds. Part (B) shows simulation results, also for three lengths and orientations. In both cases perceived relative speed decreases with line length and angle from vertical. Simulated lines use slightly different orientations from those in the experiments so that the simulated input conforms to the Cartesian grid. [Reprinted with permission from Chey *et al.* (1997).]

To further test this idea, Castet *et al.* (1993) also showed that the ambiguous speeds have a greater effect as line length increases when the line is viewed for a brief duration. These additional data strongly support the idea that feature tracking signals at the line ends propagate inwards along the line to capture the ambiguous motion speed and direction there. Since capture takes longer to complete when lines are longer, the ambiguous motion signals have a larger effect on longer lines. Chey, Grossberg, and Mingolla (1997) simulated these data, as shown in Figure 2b.

In addition to simulating data of Castet *et al.* (1993) on how the perceived speeds of moving lines are affected by their length and angle, Chey, Grossberg, and Mingolla (1997) used

similar model mechanisms to also simulate, among other percepts, how the barberpole illusion (Wallach, 1976) is produced, how it can be affected by various configurational changes, and how plaid patterns move both coherently and incoherently. In addressing plaid pattern motion, the model provides explanations of when plaid patterns cohere or do not (Adelson and Movshon, 1982; Kim and Wilson, 1993; Lindsey and Todd, 1995), how contrast affects the perceived speed and direction of moving plaids (Stone, Watson and Mulligan, 1990), and why the movement of so-called Type 2 patterns differs from those of Type 1 patterns (Ferrera and Wilson, 1990, 1991; Yo and Wilson, 1992). All of these data may be explained by an interaction of figure-ground separation mechanisms in the form cortical stream interacting with motion capture mechanisms in the motion cortical stream.

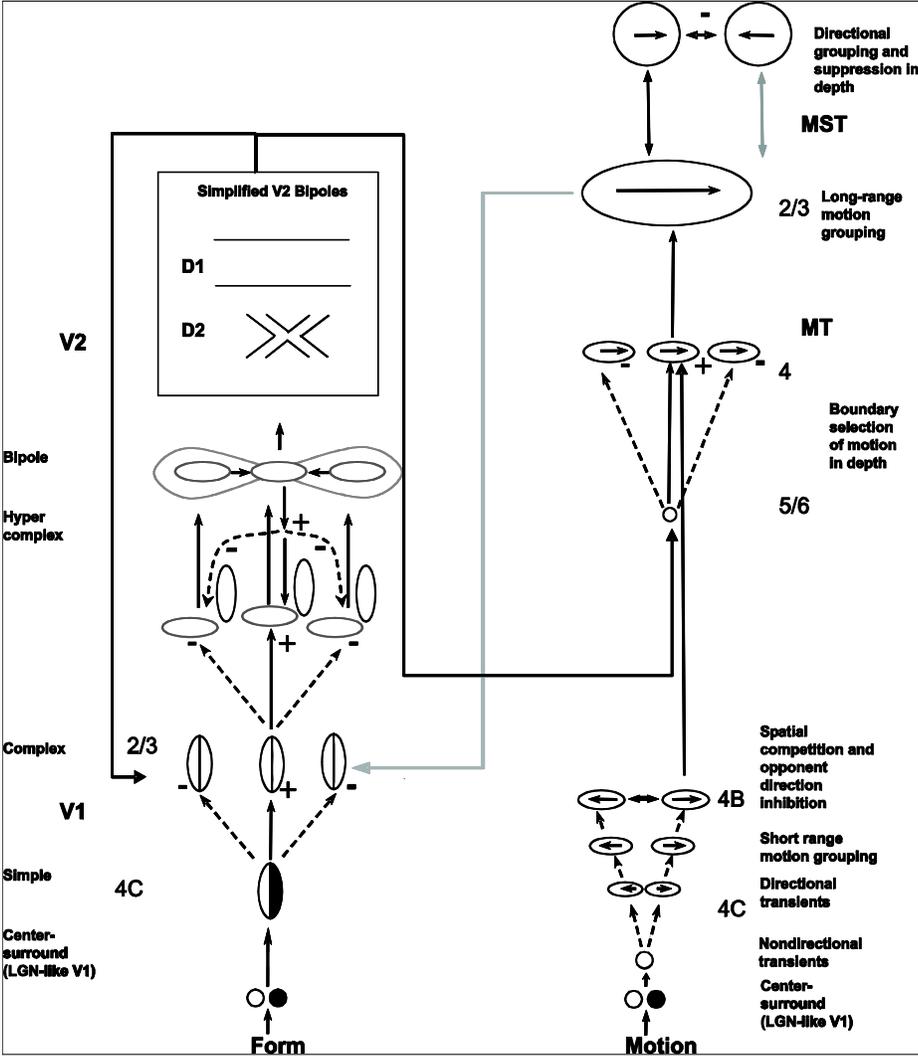


Figure 3. Laminar circuits of 3D FORMOTION model. See text for details. [Reprinted with permission from Berzhanskaya et al. (2007).]

Formotion Binding by Laminar Cortical Circuits. As the model name 3D FORMOTION suggests, it proposes how form and motion processes interact to generate coherent percepts of object motion in depth. Among the problems that the model analyses are the following form-motion (or *formotion*) binding issues, which go beyond the scope of other models: How do form-

based 3D figure-ground separation mechanisms in cortical area V2 interact with directionally selective motion grouping mechanisms in cortical areas MT and MST to preferentially bind together some motion signals more easily than others? In cases where form-based figure-ground mechanisms are insufficient, how do motion and attentional cues from cortical area MT facilitate figure-ground separation within cortical area V2 via MT-to-VI-to-V2 feedback? These processes help to explain and simulate many motion data, including the way in which the global organization of the motion direction field in areas MT and MST can influence whether the percept of an object's form looks rigid or deformable through time.

The model also goes beyond earlier motion models by proposing how *laminar* cortical circuits realize these mechanisms (Figure 3). These laminar circuits embody explicit predictions about the functional roles that are played by identified cells in the brain. The 3D FORMOTION model extends to the motion system laminar models of cortical circuits that have previously explained challenging perceptual and brain data about 3D form perception in cortical areas V1, V2, and V4 (e.g., Cao and Grossberg, 2005; Grossberg, 1999, 2003; Grossberg and Raizada, 2000; Grossberg and Seitz, 2003; Grossberg and Swaminathan, 2004; Grossberg and Williamson, 2001; Grossberg and Yazdanbakhsh, 2005), as well as about cognitive working memory, sequence learning, and variable-rate sequential performance (Grossberg and Pearson, 2006).

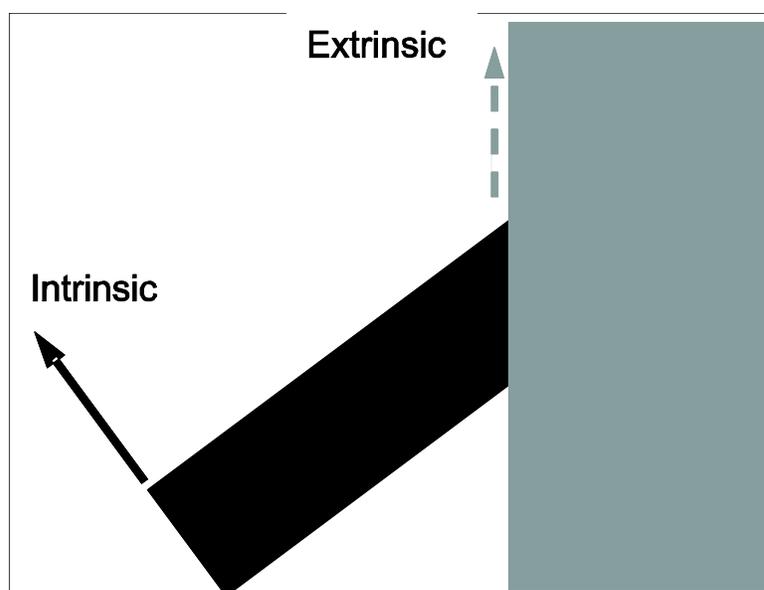


Figure 4. Extrinsic and intrinsic terminators: The local motion of the intrinsic terminator on the left reflects the true object motion, while the local motion of the extrinsic terminator on the right follows the vertical outline of the occluder.

Intrinsic and Extrinsic Terminators. A key issue in data and models about motion perception concerns the assignment of motion to an object boundary when it moves relative to an occluder. How does the brain prevent motion integration across both the occluder and its occluded objects? In the example in Figure 4, motion of the left line end corresponds to the real motion of the line. The right line end is formed by the boundary between the line and a stationary occluder. Its motion provides little information about the motion of the line. Bregman (1981) and Kanizsa (1979), and more recently Nakayama, Shimojo and Silverman (1989), have discussed this

problem. Nakayama et al. use the terminology of *intrinsic* and *extrinsic* terminators to distinguish the two cases. An intrinsic terminator belongs to the moving object, whereas an extrinsic one belongs to the occluder. Motion of intrinsic terminators is incorporated into an object's motion direction, whereas motion of extrinsic terminators is attenuated or eliminated (Duncan, Albright and Stoner, 2000; Lidén and Mingolla, 1998; Shimojo et al., 1989).

The FACADE (Form-And-Color-And-Depth) model (Grossberg, 1994, 1997; Kelly and Grossberg, 2000) of 3D form vision and figure-ground separation proposed how boundaries in 3D scenes or 2D images are assigned to different objects in different depth planes, and thereby offered a mechanistic explanation of the properties of extrinsic and intrinsic terminators. The 3D FORMOTION model (Berzhanskaya, Grossberg, and Mingolla, 2007; Grossberg, Mingolla, and Viswanathan, 2001) proposed how FACADE depth-selective figure-ground separation in cortical area V2, combined with depth-selective formotion interactions from area V2 to MT, enable intrinsic terminators to create strong motion signals on a moving object, while extrinsic terminators create weak ones. The model starts with motion signals in V1, where the separation in depth has not yet occurred, and predicts how V2-to-MT boundary signals can select V1-to-MT motion signals at the correct depths, while suppressing motion signals at the same visual locations but different depths.

Form and Motion are Complementary: What Sort of Depth does MT Compute? The prediction that V2-to-MT signals can capture motion signals at a given depth reflects the prediction that the form and motion streams compute *complementary* properties (Grossberg, 1991, 2000). The V1-V2 cortical stream, acting alone, is predicted to compute precise oriented depth estimates in the form of 3D boundary representations, but coarse directional motion signals. In contrast, the V1-MT cortical stream computes coarse depth estimates, but precise directional motion estimates. Overcoming the deficiencies of the form and motion cortical streams in computing precise estimates of form-and-motion-in-depth is predicted to occur via V2-to-MT inter-stream interactions, called *formotion* interactions. These interactions use depth-selective signals from V2 to capture motion signals in MT to lie at the correct depths. In this way, precise form-and-motion-in-depth estimates are achieved in MT, which can, in turn, be used to generate good target tracking estimates.

This analysis clarifies an otherwise vexing issue: Why does the brain bother to create a processing stream through cortical areas V1, MT, and MST to process motion, and a separate processing stream through cortical areas V1, V2, and V4 to process visual form? Indeed, individual cells in V1 already respond to properties of form, such as orientation, and properties of motion, such as direction. Given these shared properties at the earliest cortical stage V1, why could not a single processing stream compute both form and motion? In 1991 Grossberg predicted why this might be so, and in 2008, Ponce, Lomber, & Born have recently provided data that strongly support this prediction. This prediction was derived from a theoretical analysis of how the form cortical stream computes the *orientation* of form, whereas the motion stream computes the *direction* of motion.

Grossberg (1991) noted that the form stream through V1, V2, and V4 creates 3D boundaries and surfaces. In particular, 3D boundaries are computed by matching left and right eye images of visual features with similar *orientation*. The binocular disparity of these similarly oriented features is a key property that is used to compute boundaries whose interactions with surfaces can accurately represent objects in depth. This analysis predicted that the form stream can compute a precise measure of depth, but only a coarse measure of direction. In contrast, the motion stream pools over multiple orientations of an object's boundaries to compute the

direction in which the object is moving. Pooling over orientations sacrifices the key property that the form stream needs to precisely compute depth. This analysis predicted that the motion stream can compute a precise measure of direction, but only a coarse measure of depth.

These sets of properties (precise depth, coarse direction; coarse depth, precise direction) are computationally *complementary*. Grossberg (1991) predicted how an inter-stream interaction from V2 to MT could use the depth-precise boundaries in V2 to select compatible directional motion signals in MT, and thereby overcome these complementary deficiencies to generate a precise representation in MT of moving-form-in-depth. This is what various neurophysiologists, such as Livingstone and Hubel (1988) and DeAngelis et al. (1998) had earlier reported from their recordings in MT.

Neurophysiological Support for Form-Motion Complementarity and Formotion Capture. Ponce, Lomber, and Born (2008) reported neurophysiological data that are consistent with the prediction that V2 imparts finer disparity sensitivity onto MT: When V2 is reversibly cooled and V2-to-MT signals eliminated, depth selectivity, but not motion selectivity, is greatly impaired in MT. In other words, the predicted precise directional and coarse depth properties were hereby unmasked when V2 inputs were removed. These data do not support the alternative view that fine depth estimates are computed directly in MT.

Induced Motion and the Role of Form in Conscious Motion Perception. Many psychophysical data support this view of motion capture. Indeed, the V2-to-MT motion selection mechanism clarifies why we tend to perceive motion of visible objects and background features, but not of the intervening empty spaces between them. For example, consider an example of *induced motion* (Duncker, 1929/1937) wherein a frame moving to the right caused a stationary dot within the frame to appear to move to the left. Motion signals must propagate throughout the interior of the frame in order to reach and influence the dot. Despite this global propagation, the homogeneous space between the frame and the dot does not seem to move. The 3D FORMOTION model predicts that this occurs because there are no boundaries between the frame and the dot whereby to capture a motion signal. The model proposes that the formotion interaction whereby V2 boundaries select compatible MT motion signals may be necessary, if not sufficient, for a conscious percept of motion to occur.

Motion-to-Form Feedback and Figure-Ground Separation of Occluded Moving Objects. V2-to-MT formotion signals overcome one sort of uncertainty in cortical computation. Another sort of uncertainty is overcome by using MT-to-V1 feedback signals. These top-down modulatory signals can help to separate boundaries in V1 and V2 where they cross in feature-absent regions. Such feature-absent signals are illustrated by the *chopsticks illusion* (Anstis, 1990); see Figure 5. Supposed that attention or internal noise signals amplify motion signals of one chopstick more than the other via MST-MT interactions. This stronger chopstick can send its enhanced motion signals from MT to V1. The enhanced V1 signals can then use V1-to-V2 figure-ground separation mechanisms to separate the two chopsticks in depth. The form boundary of the attended chopstick will then be strengthened. The FACADE model explains how a stronger form boundary can appear to be nearer than a weaker one (e.g., Grossberg, 1994; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000). The nearer boundary can then be completed by perceptual grouping mechanisms over the ambiguous region where the two chopsticks cross, much like an illusory contour can be completed. FACADE mechanisms also explain how the intrinsic boundaries of the nearer chopstick can be detached from the farther chopstick, thereby enabling the farther chopstick boundaries to also be completed in depth behind those of the

occluding, nearer chopstick. As these boundaries are completed, they are injected back into MT from V2 to capture the corresponding motion direction signals and generate a percept of two separated figures moving in depth, one in front of the other.

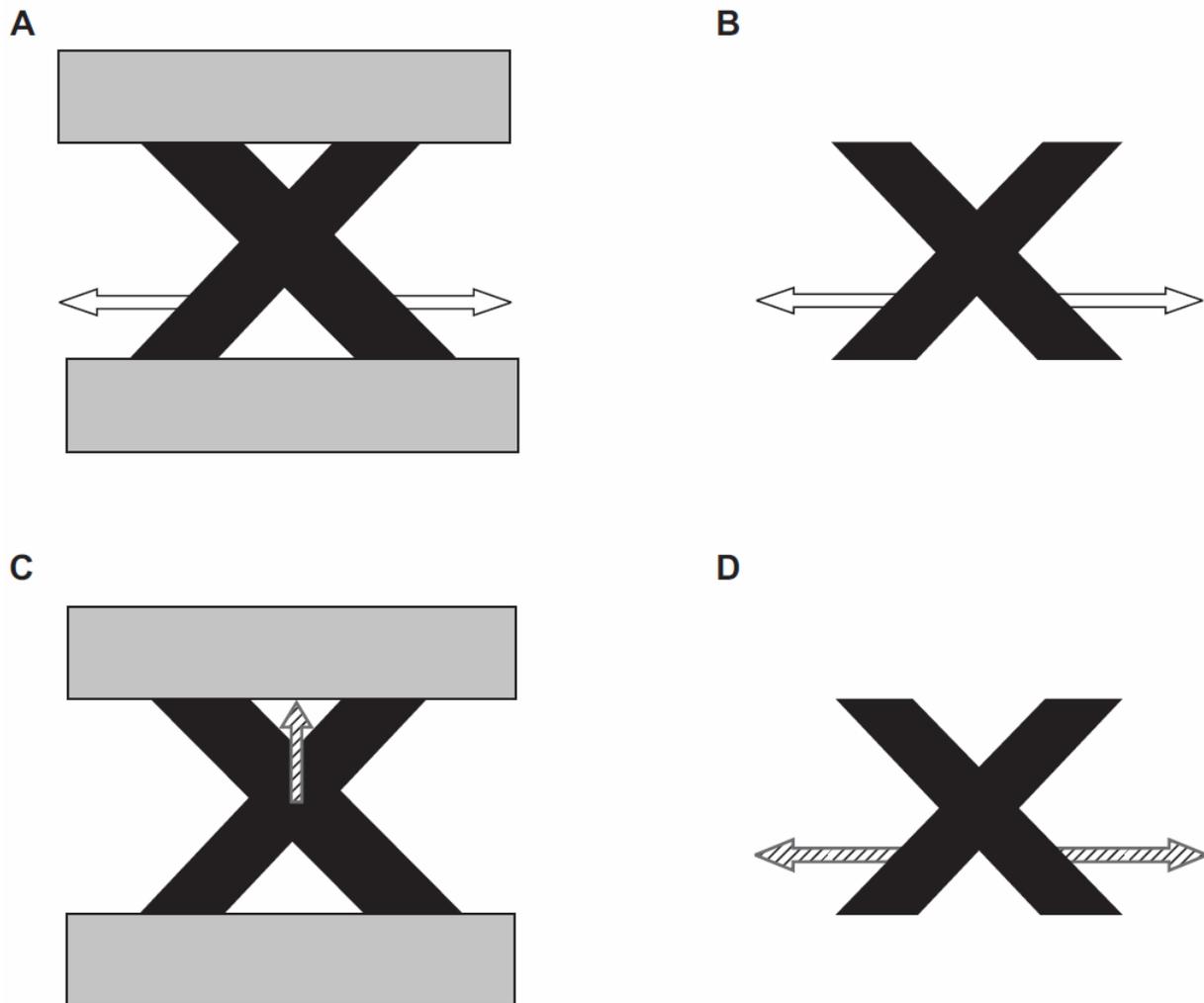


Figure 5. Chopsticks illusion: Actual chopsticks motion (clear arrows, top) is equivalent in (A) and (B), with visible and invisible occluders, respectively. Visible occluders result in a coherent vertical motion percept (C, hatched arrow). Invisible occluders result in the percept of two chopsticks sliding in opposite directions (D). [Reprinted with permission from Berzhanskaya et al. (2007).]

Adaptation and Bistable Motion. Another factor that influences motion perception is adaptation. This can be accomplished by a process of transmitter habituation, inactivation, or depression (Abbott et al., 1997; Grossberg, 1968, 1972; Francis, Grossberg, & Mingolla, 1994). For example, motion signals at the positions of a static extrinsic terminator can adapt and therefore become weaker through time. Moving intrinsic terminators, on the other hand, generate strong motion signals. The adaptation process hereby clarifies the computation of intrinsic motion signals on a relatively short time scale.

On a longer time scale, bistable motion percepts can occur due to the interaction of cooperative-competitive model mechanisms with habituated mechanisms when multiple moving objects overlap (Figures 1 and 3). For example, percepts of moving plaids or of pairs of random dot patterns can alternate between at least two possible perceptual outcomes (Ferrera and Wilson, 1987, 1990; Kim and Wilson, 1993; Snowden et al., 1991; Stoner and Albright, 1998; Stoner, Albright, and Ramachandran, 1990; Trueswell and Hayhoe, 1993). One possible outcome is a transparent motion percept, where two gratings or two dot-filled planes slide one over another in depth. Alternatively, if the directions of motions of the two gratings are compatible, then a percept of a unified plaid pattern may be seen, and no separation in depth occurs. Under prolonged viewing, the same display can perceptually alternate between coherent plaid motion and component motions separated in depth (Hupé and Rubin, 2003). The combination of 3D boundary and surface processes, augmented by habituated gating and spatial attention, can explain many bistable form and motion percepts. For example, Grossberg and Swaminathan (2004) have modeled how these processes may lead to bistable 3D percepts of the Necker cube.

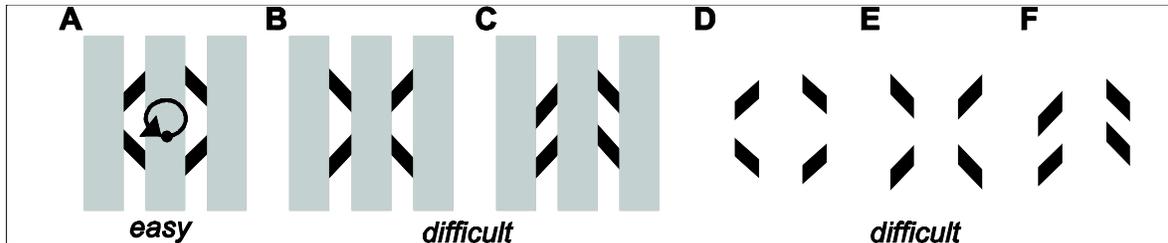


Figure 6. Lorenceau-Alais displays: Visible (A-C) and invisible (D-F) occluder cases. See text for details.

Shape and Motion Interactions Determine Object Motion Percepts Behind Apertures. Similar mechanisms can explain and simulate percepts of moving object shapes that are more complex than lines or dots. For example, Lorenceau and Alais (2001) studied different shapes moving in a circular-parallel motion behind occluders (Figure 6). Observers had to determine the direction of motion, clockwise or counterclockwise. The percent of correct responses depended on the type of shape, and on the visibility of the occluders. In the case of a diamond shape (Figure 6A), a single, coherent, circular motion of a partially occluded rectangular frame was easy to perceive across the apertures. In the case of an arrow shape (Figure 6C), two objects with parallel sides were seen to generate out-of-phase vertical motion signals in adjacent apertures. Local motion signals were identical in both displays, and only their spatial arrangement differed. Lorenceau and Alais suggested that certain shapes (such as arrows) “veto” motion integration across the display, while others (such as diamond) allow it.

The 3D FORMOTION model explains these data without using a veto process. The model proposes that the motion grouping process uses anisotropic direction-sensitive receptive fields (see Figure 3) that preferentially integrate motion signals within a given direction across gaps produced by the occluders. The explanation of percepts induced by the displays in Figures 4D-F follows in a similar way, with the additional factor that the ends of the bars activate intrinsic terminators that can strongly influence the perceived motion direction of the individual bars.

Rigid and Gelatinous Rotating Ellipses. Motion grouping also helps to explain percepts of rotational motion using the “gelatinous ellipses” display (Vallortigara et al., 1988, Weiss and Adelson, 2000). When “thin” (high aspect ratio) and “thick” (low aspect ratio) ellipses rotate

around their centers, percepts of their shapes differ markedly. The thin ellipse is perceived as a rigid rotating form, whereas the thick one is perceived as deforming non-rigidly through time. Here, the differences in 2D geometry result in differences of the spatiotemporal distribution of motion direction signals that are grouped together through time. When these motion signals are consistent with the coherent motion of a single object, then the motion grouping process within the model MT-MST processing stages (Figure 1) generates a percept of a rigid rotation. When the motion field decomposes after grouping into multiple parts, with motion trajectories incompatible with a rigid form, a non-rigid percept is obtained. Nearby “satellites” can convert the non-rigid percept into a rigid one. Weiss and Adelson (2000) proposed that such a percept can be explained via a global optimization process. The 3D FORMOTION model shows how motion grouping can provide a biologically more plausible explanation.

Once V2 boundaries capture MT motion signals, what are they used for? One major use is for visually-based navigation. For example, Browning, Grossberg, & Mingolla (2009) developed the ViSTARS neural model in which motion signals are processed through the cortical areas MT⁻ and MST^v to track moving targets, while cortical areas MT⁺ and MST^d use motion signals to compute heading from optic flow using computationally complementary interactions. Grossberg & Pilly (2008) showed how motion signals through cortical areas MT⁻, MST, and LIP can be used to probabilistically compute decisions about the direction of eye movements in response to moving dot patterns. These latter results will be discussed after the main mechanisms of the 3D FORMOTION model are summarized as a way to unify the discussion and understanding of many experimental data about the psychology and neurobiology of motion perception.

3D FORMOTION Model

The 3D FORMOTION model (Figure 1 and 3) proposes how six types of processes interact together in the brain’s form and motion systems. Because model processing stages are analogous to areas of the primate visual system, they are called by the corresponding anatomical names: (1) V1-to-MT filtering and cooperative-competitive processes begin to resolve the aperture problem by amplifying feature tracking signals and attenuating ambiguous motion signals so that the feature tracking signals have a chance to overwhelm numerically superior ambiguous motion signals. (2) 3D boundary representations, in which figures are separated from their backgrounds, are formed in cortical area V2. (3) These depth-selective V2 boundaries select motion signals at the appropriate boundary positions and depths in MT via V2-to-MT signals. (4) A spatially anisotropic motion grouping process propagates across perceptual space via MT-MST feedback to integrate veridical feature tracking signals with ambiguous motion signals and to thereby determine a global object motion percept. This motion capture process solves the aperture problem. (5) MST-MT feedback can convey an attentional priming signal from higher brain areas that can influence the motion capture process, and thereby influence form processing via MT-to-V1 feedback. (6) Motion signals in MT can hereby disambiguate locally incomplete or ambiguous boundary signals in V2 via MT-to-V1-to-V2 signals.

These interactions provide a functional explanation of many neurophysiological data. Table 1 summarizes anatomical connections and neuron properties that occur in the model, alongside experimental references which support those connections or functional properties. Table 1 also lists the model’s key physiological predictions that remain to be tested. As illustrated in Figures 1 and 3, these interactions are naturally understood as part of a *form processing stream* and a *motion processing stream*.

Connection/Functional property	Selected references
Functional projections	
V1 4Ca to 4B	Yabuta et al., 2001, Yabuta & Callaway, 1998
V1 to MT	Anderson et al., 1998; Rockland, 2002; Sincich & Horton 2003, Movshon & Newsome, 1996
V1 to V2	Rockland, 1992, Sincich & Horton, 2002
V2 to MT	Anderson & Martin, 2002; Rockland 2002; Shipp & Zeki 1985; DeYoe & Van Essen 1985
MT to V1 feedback	Shipp & Zeki 1989; Callaway 1998; Movshon & Newsome 1996; Hupé et al., 1998
V2 to V1 feedback	Rockland & Pandya, 1981; Kennedy & Bullier 1985
Properties	
V1 adaptation	Abbott et al., 1997; Chance et al., 1998; (rat); Carandini & Ferster, 1997, (cat)
V1(4ca) transient nondirectional cells	Livingstone & Hubel, 1984
V1 spatially offset inhibition	Livingstone, 1998; Livingstone & Conway, 2003; Murthy & Humphrey, 1999 (cat)
V2 figure-ground separation	Zhou et al., 2000; Bakin et al., 2000
MT figure-ground separation and disparity sensitivity	Bradley et al., 1998, Grunewald et al., 2002; Palanca & DeAngelis 2003
MT center- surround receptive fields	Bradley & Andersen, 1998; Born, 2000; DeAngelis & Uka, 2003
Some MT receptive fields elongated in preferred direction of motion	Xiao et al., 1997
Attentional modulation in MT	Treue & Maunsell, 1999
Predictions	
Short-range anisotropic filter in V1 (motion stream)	
Long-range anisotropic filter in MT (motion)*	
V2 to MT projection carries figure-ground completed-form-in-depth separation signal	
MT to V1 feedback carries figure-ground separation signal from motion to form stream	
MST to MT feedback helps solve aperture problem by selecting consistent motion directions	

*Although Xiao et al, 1997 found that some MT neurons have receptive fields that are elongated along the preferred direction of motion, there is no direct evidence that these neurons participate preferentially in motion grouping.

Table 1. Functional projections and properties of model cell types and predictions. [Reprinted with permission from Berzhanskaya et al. (2007)]

The Form Processing System

The model's form processing system comprises six stages that are displayed on the left sides of Figures 1 and 3. Input are processed by distinct ON and OFF cell networks whose cells obey membrane, or shunting, equations while they undergo on-center off-surround and off-center on-

surround network interactions, respectively, that are similar to those of LGN cells. These cells excite simple cells in cortical area V1 to register boundary orientations in cells that are sensitive to a particular polarity-of-contrast. Then complex and hypercomplex cells pool across simple cells tuned to opposite contrast polarities to detect object boundaries even if their contrast with respect to a background reverses along the perimeter of the object. Then several mechanisms work together to enhance feature tracking signals while down-regulating ambiguous motion signals: divisive normalization reduces cell activities where there are multiple ambiguous orientations in a region, end-stopping enhances activity at features like line-ends where feature tracking signals often occur, and spatial sharpening enhances the feature tracking advantage. These cells input to a perceptual grouping circuit in layer 2/3 of V2. Here bipole cells receive signals via long-range horizontal interactions from approximately collinear cells whose orientational preferences lie along, or near, the collinear axis. These cells are indicated by and-gates the figure-8 shape in Figure 3. They act like statistical and-gates that permit grouping only when there is sufficient evidence from pairs or greater numbers of inducers on both sides of the cell body (Grossberg, 1994; Grossberg and Mingolla, 1985a, 1985b). Grouping completes and sharpens object boundaries. It includes a stage of cross-orientation competition that reinforces boundary signals with greater support from neighboring boundaries while weakening spatially overlapping boundaries of non-preferred orientations. Boundaries are assigned into different depths, as follows:

Perceptual Grouping and Figure-Ground Separation of 3D Form. The FACADE model shows how the boundary completion process within the pale stripes of V2 can automatically initiate separation of extrinsic vs. intrinsic boundaries in depth without positing separate mechanisms to compute T-junctions (Grossberg, 1994, 1997; Kelly and Grossberg, 2000). Indeed, one cue for extrinsic vs. intrinsic boundaries is occlusion in a 2D image at a T-junction, as illustrated in Figure 4, where a moving black bar intersects a stationary gray rectangular occluder. The top of the T belongs to the occluding gray rectangle, while the stem belongs to the occluded black bar. Bipole long-range excitatory horizontal interactions can strengthen the boundary of the gray occluder where it intersects the black bar, while short-range competition (Figure 3) weakens, or totally inhibits, the boundary of the black occluded bar where it touches the gray occluder. This *end gap* in the black boundary initiates the process of separating occluding and occluded boundaries. In other words, perceptual grouping properties are predicted to initiate the separation of figures from their backgrounds, without the use of explicit T-junction operators. This prediction has received support from psychophysical experiments; e.g., Dresch, Durand, and Grossberg (2002) and Tse (2005). The figure-ground separation process enables the model to distinguish extrinsic from intrinsic terminators, and to thereby select motion signals at the correct depths.

Qualitative explanations and quantitative computer simulations of 3D figure-ground separation during perception of static images were provided using the laminar cortical circuits of the 3D LAMINART model by Fang and Grossberg (2009), who simulated 3D surface perception in response to dense, sparse, and occluding stereograms. Grossberg and Yazdanbaksh (2005) simulated stable and bistable transparency and 3D neon color spreading. The non-laminar circuits of the FACADE model were earlier used by Kelly and Grossberg (2000) to simulate such percepts as Bregman-Kanizsa figure-ground perception, Kanizsa stratification, and lightness percepts such as the Munker-White, Benary cross, and checkerboard percepts. In 3D FORMOTION analyses of figure-ground separation of moving targets, a complete simulation of all form and motion computations was computationally prohibitive. Instead, to reduce the

simulation computational load, Berzhanskaya et al. (2007) (see also Grossberg et al., 2001) used the following approximation: As soon as T-junctions were detected by the model dynamical equations, V2 boundaries were algorithmically assigned the depths that a complete figure-ground simulation would have assigned to them. In particular, static occluders were assigned to the near depth (D1 in Figure 3) and lines with extrinsic terminators were assigned to the far depth (D2 in Figure 3). These V2 boundaries were used to provide both V2-to-MT motion selection signals and V2-to-V1 depth-biasing feedback. While V2-to-V1 feedback is orientation-specific, the V2-to-MT projection sums boundary signals over all orientations, just as motion signals do at MT (Albright, 1984).

Motion Induction of Figure-Ground Separation. When form cues are not available to initiate figure-ground separation, motion cues may be able to do so via feedback projections from MT to V1 (Figures 1 and 3). Such a feedback projection has been studied both anatomically and electrophysiologically (Bullier, 2001; Jones, Grieve, Wang and Sillito, 2001; Movshon and Newsome, 1996) and it can benefit from attentional biasing within MT/MST (Treue and Maunsell, 1999). As explained above, this mechanism can help to separate chopsticks in depth (see Figure 5B). Focusing spatial attention at one end of a chopstick enhanced that chopstick's direction of motion in MT and MST. Enhanced MT-to-V1 feedback selectively strengthened the boundary signals of the corresponding chopstick in Figure 5B enough to trigger its boundary completion across the other chopstick via V1-to-V2 interactions. This operation also initiated figure-ground separation that assigned the now occluded chopstick to a farther depth. Then, by closing the V2-to-MT loop, these two overlapping but depth-separated bars can support depth-selective motions by the chopsticks in opposite directions (Bradley, Chang and Andersen, 1998).

The Motion Processing System

The model's motion processing stream consists of six stages that represent cell dynamics homologous to LGN, V1, MT, and MST (Figures 1 and 3, right).

Level 1: ON and OFF cell inputs from Retina and LGN, which are lumped into a single processing stage, activate model V1 (Xu, Bonds and Casagrande, 2002). These inputs are not depth-selective. In response to a 2D picture, this depth-selectivity will come from figure-ground separated V2 boundaries when they select consistent motion signals in MT.

Both ON and OFF cells have a role to play. For example, if a bright chopstick moves to the right on a dark background, ON cells respond to its leading edge, but OFF cells respond to its trailing edge. Likewise, when the chopstick reverses direction and starts to move to the left, its leading edge now activates ON cells and its trailing edge OFF cells. By differentially activating ON and OFF cells in different parts of this motion cycle, these cells have more time to recover from habituation, so that the system remains more sensitive to repetitive motion signals.

Level 2: Transient cells. The second stage of the motion processing system consists of non-directional transient cells, inhibitory directional interneurons, and directional transient cells. The non-directional transient cells respond briefly to a change in the image luminance, irrespective of the direction of movement. Such cells respond well to moving boundaries and poorly to a static occluder because of the habituation of the process that activates the transient signal. Adaptation is known to occur at several stages in the visual system, including retinal Y cells (Enroth-Cuggell and Robson, 1966; Hochstein and Shapley, 1976a, 1976b) and cells in V1 (Abbott, Sen, Varela and Nelson, 1997; Carandini and Ferster, 1997; Chance, Nelson and Abbott, 1998; Varela, Sen, Gibson, Fost, Abbott and Nelson, 1997) and beyond.

The non-directional transient cells send signals to inhibitory directional interneurons and directional transient cells, and the inhibitory interneurons interact with each other and with the directional transient cells (Figure 7). The directional inhibitory interneuronal interaction enables the directional transient cells to maintain their directional selectivity at a wide range of speeds (Grossberg, Mingolla, and Viswanathan, 2001). This selectivity is predicted to be due to the inhibitory interneurons, and goes beyond the capabilities of the classical Barlow and Levick (1965) model of transient cell activation. The predicted interactions in this model circuit are consistent with retinal data concerning how bipolar cells interact with inhibitory starburst amacrine cells and direction-selective ganglion cells, and how starburst cells interact with each other and with ganglion cells (Fried, Münch, and Werblin, 2002). The possible role of starburst cell inhibitory interneurons in ensuring directional selectivity at a wide range of speeds has not yet been tested.

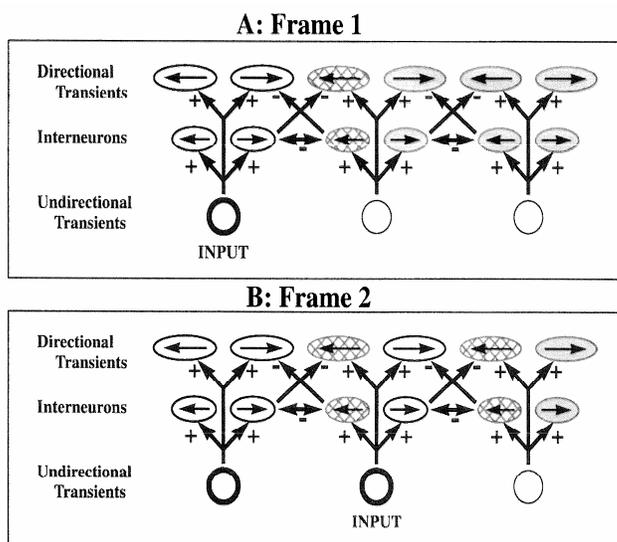


Figure 7. Schematic diagram of a 1D implementation of the transient cell network showing the first two frames of the motion sequence. Thick circles represent active unidirectional transient cells while thin circles are inactive unidirectional transient cells. Ovals containing arrows represent directionally selective neurons. Unfilled ovals represent active cells, cross-filled ovals are inhibited cells and gray-filled ovals depict inactive cells. Excitatory and inhibitory connections are labeled by ‘+’ and ‘-’ signs respectively. [Reprinted with permission from Grossberg et al. (2001).]

A directionally selective neuron fires vigorously when a stimulus is moved through its receptive field in one direction (called the preferred direction), while motion in the reverse direction (called the null direction) evokes little response. This type of directional selectivity was first modeled by Barlow and Levick (1965). Mechanisms of direction selectivity include asymmetric inhibition along the preferred cell direction, notably an inhibitory veto of null-direction signals.

As noted above, after the transient cells adapt to a static boundary, then boundary segments that belong to a static occluder—that is, extrinsic terminators—in the chopsticks display with visible occluders (Figure 5A) produce weaker signals than those that belong to the continuously moving parts of the chopstick. On the other hand, in the invisible occluder chopsticks display (Figure 5B), the horizontal motion signals at the chopstick ends will

continually move, hence will be strong, and can thus significantly influence the conscious motion percept.

Level 3: Short-range filter. A short-range directional filter (Figure 3) helps to selectively strengthen unambiguous feature tracking signals, relative to ambiguous motion signals. Cells in this filter *accumulate evidence* from directional transient cells of similar directional preference within a spatially anisotropic region that is oriented along the preferred direction of the cell; cf., Braddick (1980). Short-range filter cells amplify feature tracking signals at unoccluded line endings, object corners, and other scenic features. The short-range spatial filter, followed by competitive selection, eliminates the need to solve the *feature correspondence problem* that various other motion models use, such as Reichardt and Elaborated Reichardt models (Reichardt, 1961; van Santen and Sperling, 1985).

Level 4: Spatial competition and opponent direction competition. Two kinds of competition further enhance the relative advantage of feature tracking signals. These competing cells are proposed to occur in layer 4B of V1 (Figure 3, bottom-right). Spatial competition among like-directional cells of the same spatial scale further boosts the amplitude of feature tracking signals relative to those of ambiguous signals. This happens because feature tracking locations are often found at motion discontinuities, and thus get less inhibition than ambiguous motion signals that lie within an object's interior. Opponent-direction competition also occurs at this processing stage, with properties similar to V1 cells that may play this functional role (Rust, Majaj, Simoncelli and Movshon, 2002).

Data of Pack, Gartland, and Born (2004) support properties of cells at this model stage. In their data, V1 cells exhibit suppression of responses to motion along visible occluders. Suppression occurs in the model due to the adaptation of transient inputs to static occluding boundaries. In addition, V1 cells in the middle of a grating, where ambiguous motion signals occur, respond more weakly than cells at the edge of the grating, where intrinsic terminators occur. Model spatial competition between motion signals emulate this property through its properties of divisive normalization and endstopping. Together these properties amplify directionally unambiguous feature tracking signals at line ends relative to the strength of aperture-ambiguous signals along line interiors, which compete among themselves for normalized activity at their position.

Level 5: Long-range filter and formotion selection. The long-range directional filter pools together motion signals with the same, or similar, directional preference from moving features with different orientations, contrast polarities, and eyes. These motion signals may be carried from model layer 4B of V1 input to model area MT. Its cell targets may be considered true directional cells which combine evidence from multiple informational sources. These cells have directionally-sensitive properties in the motion stream through MT that are computationally homologous to those of orientationally-sensitive complex cells in the form stream through V2. It will be interesting to see if future studies of cortical evolution and development support the idea that these cells share an underlying computational design that is later specialized for motion or form processing.

Area MT also receives a projection from V2 (Anderson and Martin, 2002; Rockland, 1995). As described above, this V2-to-MT formotion projection is predicted to carry depth-specific figure-ground separated boundary signals. These V2 form boundaries selectively assign to different depths the motion signals coming into MT from layer 4B of V1.

Formotion selection is proposed to occur via a modulatory on-center, off-surround projection from V2 to layer 4 of MT. For example, in response to the chopsticks display with

visible occluders (Figure 5A), motion signals which lie along the visible occluder boundaries are selected in the near depth and are suppressed by the off-surround at other locations at that depth. The selected signals will be weak because the bottom-up input from V1 habituates along the selected occluder boundary positions. The V2 boundary signals that correspond to the moving boundaries select strong motion signals at the farther depth. The on-center of the V2-to-MT interaction is modulatory to prevent it from creating motion signals at MT where none are received from V1. This type of circuit is predicted by Adaptive Resonance theory, or ART, to enable fast learning with self-stabilizing memory, and to focus attention on salient information (e.g., Grossberg, 1999b, 2003). In the present case, the V2-to-MT interaction between correlated form and motion information is predicted to be tuned by learning during perceptual experience.

Boundary-gated signals from layer 4 of MT input to the upper layers of MT (Figure 3, top-right), where they activate a directionally-selective, spatially *anisotropic* long-range filter via long-range horizontal connections. The hypothesis that the long-range filter uses an anisotropic filter is consistent with data showing that approximately 30 % of the cells in MT show a preferred direction of motion that is aligned with the main axis of their receptive fields (Xiao, Raiguel, Marcar and Orban, 1997).

The long-range directional filter cells in layer 2/3 of MT are proposed to play a role in motion grouping that is homologous to the role played by bipole cells during form grouping within layer 2/3 of the pale stripes of cortical area V2 (Grossberg 1999a; Grossberg and Raizada, 2000). As noted above, the anisotropic long-range motion filter allows motion signals to be selectively integrated across occluders in a manner that naturally explains the percepts generated by the Lorenceau-Alais displays of Figure 6.

Long-Range Apparent Motion and Target Tracking. The long-range directional filter can also help to explain many data about long-range apparent motion, which creates a percept of continuous motion between discrete flashes of light that are separated in space and time (e.g., Kolers, 1972). The evolutionary utility of apparent motion may be appreciated by considering an animal running at variable speeds behind forest cover. The forest is an occluder that creates apertures through which fragments of the animal's motion signals are intermittently experienced. The mechanisms of long-range apparent motion enable the brain to group these fragments into a continuous motion percept that facilitates predictive tracking of the animal and its trajectory. Because of the Gaussian shape of the long-range filter, it can interpolate discrete flashes into a continuous trajectory that has many of the properties of long-range apparent motion, notably its ability to speed up as the distance between inducing flashes increases without a change in their delay, or as the delay between the two flashes decreases without a change in their distance (Grossberg and Rudd, 1992; Grossberg, 1998). The depth-selective separation of extrinsic vs. intrinsic boundaries in V2, followed by V2-to-MT depth-selective motion binding, together help to explain how the forest cover appears closer to an observer than the animal running behind it. In summary, the long-range filter helps to group motion signals across both space and time.

Level 6: Directional grouping. The first five model stages can amplify feature tracking signals and assign motion signals to the correct depths. However, they do not explain how feature tracking signals propagate across space to select consistent motion directions from ambiguous motion directions, suppress inconsistent motion directions, all the while without distorting their speed estimates,. They also cannot explain how motion integration can compute a vector average of ambiguous motion signals across space to determine the perceived motion direction when feature tracking signals are not present at that depth. The model's final stage accomplishes these properties by using a motion grouping network that is interpreted to exist in

ventral MST (MSTv), which is known to be important for target tracking (Bherezovskii and Born, 1999; Born and Tootell, 1992; Eifuku and Wurtz, 1998; Pack, Grossberg, and Mingolla, 2001; Tanaka et al., 1993). This motion grouping network is predicted to determine the coherent motion direction of discrete moving objects.

During motion grouping, cells that code the same, or similar, directions in MT send convergent inputs to cells in model MSTv via the motion grouping network. Within MSTv, directional competition at each position determines a winning motion direction. This winning directional cell then feeds back to its source cells in MT. Just as in the case of V2-to-MT signaling, this MSTv-to-MT feedback is defined by a modulatory on-center, off-surround network. It selects activities of MT cells that code the winning direction, while suppressing activities of cells that code other directions. Using this broad feedback kernel, the motion grouping network enables feature tracking signals to select similar directions at nearby ambiguous motion positions, while suppressing other directions there. In other words, motion capture occurs and disambiguates ambiguous motion positions. The use of a modulatory on-center, off-surround top-down circuit from MSTv-to-MT enables bottom-up learning from MT-to-MSTv and top-down learning from MSTv-to-MT to create and self-stabilize directionally selective motion grouping receptive fields during perceptual experience, using general ART properties.

As the grouping process cycles bottom-up and top-down between MT and MSTv, directional motion capture propagates laterally across space: Each cycle of top-down MSTv-to-MT feedback creates a region of cells in MT whose directional preference is consistent with that of the MSTv cells that activated the feedback. Then this newly unambiguous region of directional preference in MT can use the bottom-up MT-to-MSTv filter to select directionally consistent MSTv grouping cells at positions near them, and the cycle continues. In other words, motion capture emerges automatically using the feedback process that enabled stable development of directional tuning to occur. Chey et al. (1997) and Grossberg et al. (2001) used this process to simulate psychophysical data showing how the aperture problem may be solved via a gradual process of motion capture, and Pack and Born (2001) provided supportive neurophysiological data by directly recording from MT cells, as noted above.

Ubiquitous Circuit Design for Selection, Attention, and Learning. It is worth emphasizing that both the V2-to-MT and the MSTv-to-MT signals carry out their selection processes using modulatory on-center, off-surround interactions. The V2-to-MT signals select motion signals at the locations and depth of a moving boundary. The MSTv-to-MT signals select motion signals in the direction and depth of a motion grouping. Adaptive Resonance Theory predicted that such a modulatory on-center, off-surround network would be used to carry out attentive selection and modulation of adaptive tuning within all brain circuits wherein fast learning and self-stabilizing memory of appropriate features is needed. In the V2-to-MT circuit, a formotion association is learned. In the MST-to-MT circuit, directional grouping cells are learned. Grossberg (2003) and Raizada and Grossberg (2003) review behavioral and neurobiological data that support this prediction in several brain systems. The Ponce, Lomber, and Born (2006) study supports the V2-to-MT prediction, but does not study how this association may be learned. There do not seem to be any direct neurophysiological tests of the MSTv-to-MT prediction.

Converting Motion into Action during Perceptual Decision-Making

Motion Capture in Perceptual Decision-Making. The 3D FORMOTION model sheds new light on how the brain may make motion-based movement decisions, in particular saccadic eye

movements, in response to probabilistically defined motion stimuli. This example illustrates the general theme that motion may be used to drive a variety of target tracking behaviors.

It is well known that speed and accuracy of perceptual decisions and the movements that they control covary with certainty in the input, and correlate with the rate of evidence accumulation in parietal and frontal cortical neurons. Data concerning such motion-based movement decisions can be explained and quantitatively simulated by pre-processing motion stimuli with the V1-MT-MST mechanisms that are articulated in the 3D FORMOTION model before MST activates the parietal cortex, indeed LIP, where motion direction is converted into a directional movement command that is gated by the basal ganglia (Figure 8). The model in which these processing stages are implemented is called the MOTion DEcision, or MODE, model because it clarifies challenging data about probabilistically defined motion perception and action (Grossberg and Pilly, 2008).

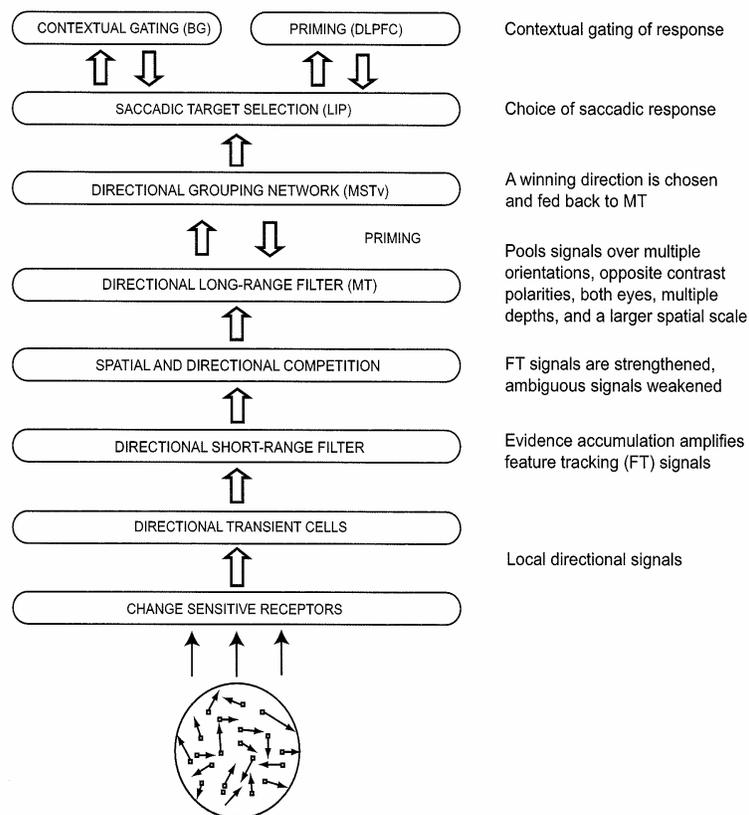


Figure 8. Retina/LGN-V1-MT-MST-LIP-BG model processing stages. See text for details. The random dot motion stimuli are preprocessed by the model Retina/LGN and processed by the model cortical V1-MT-MST stream. They contextually transform locally ambiguous motion signals into unambiguous global object motion signals with a rate, amplitude, and direction that covaries with the amount of dot coherence. These spatially distributed global motion signals then feed into model area LIP to generate appropriate directional saccadic eye movement commands, which are gated by the model basal ganglia (BG). [Rerpinted with permission from Grossberg and Pilly (2008).]

The MODE model quantitatively simulates dynamic properties of decision-making in response to the types of ambiguous visual motion stimuli that have been studied in LIP neurophysiological

recordings by Newsome, Shadlen, and colleagues. The most important circuits of this enhanced model already lie within the 3D FORMOTION model. since the rate of motion capture in the MT-MST grouping network covaries with the activation rate and amplitude of LIP cells that control a monkey's observable behavior in the experiment. The model hereby clarifies how brain circuits that solve the aperture problem, notably the circuits that realize motion capture, may control properties of probabilistic decision making in real time. This is not surprising when one interprets the motion capture process as a resolution of ambiguity that selects the best consensus movement that is compatible with motion data.

Are the Brain's Decisions Bayesian? These results are of particular interest because distinguished perceptual and brain scientists, including Newsome and Shadlen, have proposed that perception and decision-making can be described as Bayesian inference, which estimates the optimal interpretation of the stimulus given priors and likelihoods. However, Bayesian concepts do not, in themselves, provide a way to discover the neocortical mechanisms that make decisions. The present model explains data that Bayesian models have heretofore failed to explain, does so without an appeal to Bayesian inference and, unlike other existing models of these data, generates perceptual representations in response to the moving experimental visual stimuli. In particular, the MODE model quantitatively simulates the time course of LIP neuronal dynamics, as well as behavioral accuracy and reaction time properties, during both correct and error trials at different levels of input ambiguity in both fixed duration and reaction time tasks. Model MST computes the global direction of random dot motion stimuli as part of the motion capture process, while model LIP converts a distributed motion representation in MST into a directional movement decision that leads to a saccadic eye movement. MODE hereby trades accuracy against speed, and illustrates how cortical dynamics go beyond Bayesian concepts, while clarifying why probability theory ideas are initially so appealing.

Concerning the appeal of statistical, in particular Bayesian, concepts, it should be noted that the shunting on-center off-surround networks (Grossberg, 1973, 1980) that occur ubiquitously in the brain, and also in the 3D FORMOTION model, tend to normalize the activities across a neural network. The spatially distributed pattern of these normalized activities may be viewed as a type of real-time probability distribution. In addition, any filtering operation, such as the short-range and long-range directional filters, may be interpreted as a prior (namely, the current neural signal) multiplied by a conditional probability or likelihood (namely, the filter connection strength to the target cell). Likewise, a contrast-enhancing operation, such as the LIP recurrent on-center off-surround network that selects a winning direction from filter inputs, may be viewed as maximizing the posterior. These insights have been known in the neural modeling literature for a long time (Grossberg, 1978). However, as Figures 1, 3, and 8 illustrate, such local processes do not embody the computational intelligence of an entire neural system that has emerged through evolution to realize particular behavioral competences, such as motion perception and movement decision-making.

Two Movement tasks. Newsome, Shadlen, and colleagues studied neural correlates of perceptual decision-making in macaques which were trained to discriminate motion direction. Random dot motion displays, covering a 5° diameter aperture centered at the fixation point, were used to control motion coherence; namely, the fraction of dots moving non-randomly in a particular direction from one frame to the next in each of three interleaved sequences. Varying motion coherence provided a quantitative way to control the ambiguity of directional information

that a monkey used to make a saccadic eye movement to a peripheral choice target in the perceived motion direction, and thus the task difficulty.

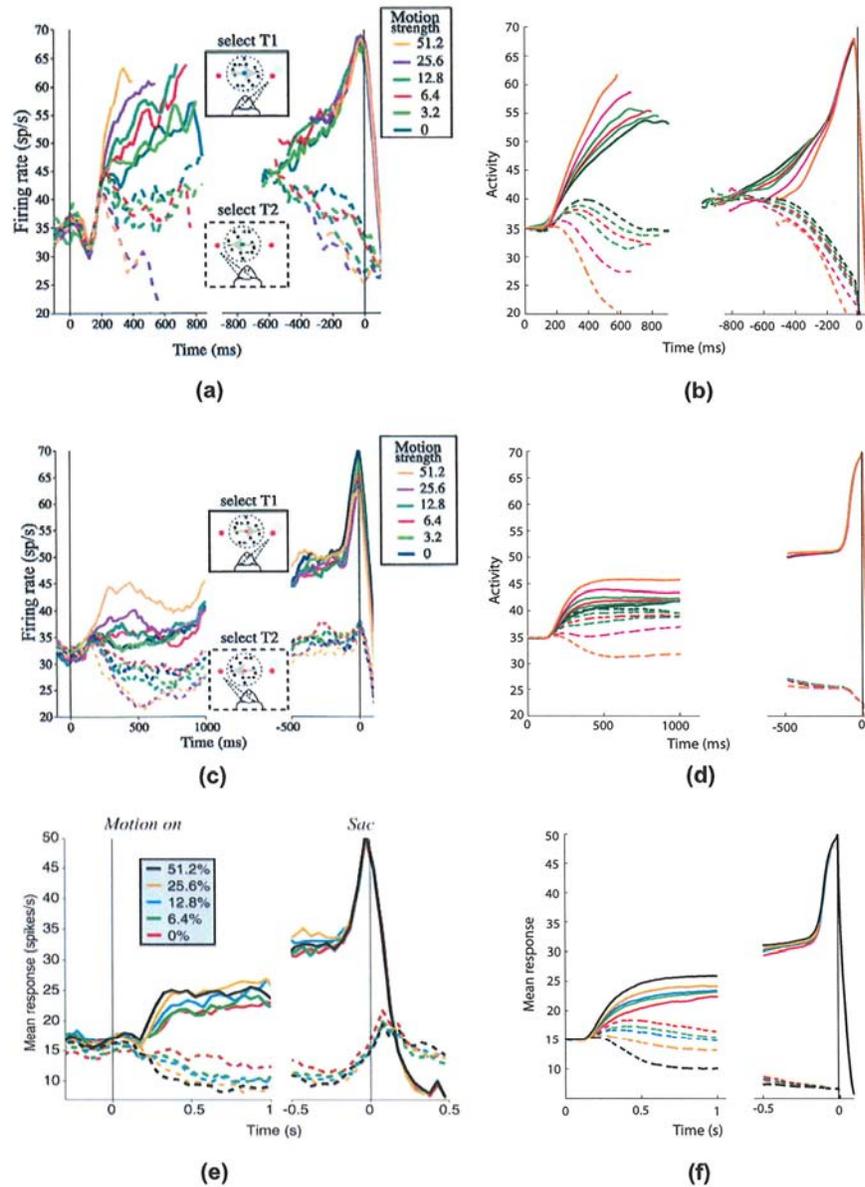


Figure 9. Temporal dynamics of LIP neuronal responses during the fixed duration (FD) and reaction time (RT) tasks. (A) Average responses of a population of 54 LIP neurons among correct trials during the RT task (Roitman and Shadlen, 2002). The left part of the plot is time-aligned to the motion onset, and includes activity only up to the median RT, and excludes any activity within 100 ms backward from saccade initiation (which corresponds to presaccadic enhancement). The right part of the plot is time-aligned to the saccade initiation, and excludes any activity within 200 ms forward from motion onset (which corresponds to initial transient dip and rise). (B) Model simulations replicate LIP cell recordings during the RT task. In both data and simulations for the RT task, the average responses were smoothed with a 60 ms running mean. (C) Average responses of

a population of 38 LIP neurons among correct trials during the 2002 FD task (Roitman and Shadlen, 2002), during both the motion viewing period (1 s) and a part (0.5 s) of the delay period before the saccade is made. (D) Model simulations mimic LIP cell recordings during the 2002 FD task. (E) Average responses of a population of 104 LIP neurons among correct trials during the 2001 FD task (Shadlen and Newsome, 2001), during both the motion viewing period (1 s) and a part (0.5 s) of the delay period before the saccade is made. (F). Model simulations emulate LIP cell recordings during the 2001 FD task. In (A-F), solid and dashed curves correspond to trials in which the monkey correctly chose the right target (T1) and the left target (T2), respectively. Cell dynamics (rate of rise or decline, and response magnitude) reflect the incoming sensory ambiguity (note the different colors; the color code for the various coherence levels is shown in the corresponding data panels), and the perceptual decision (note the two line types). For 0% coherence, even though there is no correct choice per se, the average LIP response rose or declined depending on whether the monkey chose T1 or T2, respectively. [Reprinted with permission from Grossberg and Pilly (2008).]

Two kinds of tasks were employed, namely *fixed duration* (FD) and *reaction time* (RT) tasks. In the FD task (Roitman and Shadlen, 2002; Shadlen and Newsome, 2001), monkeys viewed the moving dots for a fixed duration of 1 s, and then made a saccade to the target in the judged direction after a variable delay. In the RT task (Roitman and Shadlen, 2002), monkeys had theoretically unlimited viewing time, and were trained to report their decision as soon as the motion direction was perceived. The RT task allowed measurement of how long it took the monkey to make a decision, which was defined as the time from the onset of the motion until when the monkey initiated a saccade.

Neurophysiological recordings were done in LIP while the monkeys performed these tasks. The recorded neurons had receptive fields that encompassed just one target, and did not include the circular aperture in which the moving dots were displayed. Also, they were among those that showed sustained activity during the delay period of a memory-guided saccade task. Even though there is no motion stimulus within their classical receptive fields, these neurons still respond with directional-selectivity, probably because of extensive training on the task during which an association was learned (Bichot et al., 1996). This property has also been observed for neurons in superior colliculus whose movement fields contain just one target (Horwitz et al., 2004a; Horwitz and Newsome, 2001a).

The recorded LIP neurons show visuo-motor responses. On correct trials during the decision-making period, more coherence in the favored direction causes faster LIP cell activation, on average, in both the tasks (Figure 9), and also higher maximal cell activation in the FD task (Figures 9(C-F)). More coherence in the opposite direction causes faster cell inhibition in both the tasks, and also lower minimal cell activation in the FD task.

Comparing Trackable Features with Coherently Moving Dots. There are many details that need to be carefully discussed to quantitatively explain data from this paradigm. These details should not, however, obscure the main point, which is that a clear mechanistic homolog exists between sparse feature tracking signals and sparse but coherent moving dots.

We have already discuss that, in order to solve the aperture problem, the brain needs to ensure that a sparse set of unambiguous feature tracking motion signals can gradually capture a vastly greater number of ambiguous motion signals to determine the global direction and speed of object motion. In the case of random dot motion discrimination tasks, the signal dots at any

coherence level produce unambiguous, though short-lived, motion signals. Thus, individual dots do not experience an aperture problem. Despite this difference, the MODE model shows how the same mechanisms that help resolve the aperture problem can also enable a small number of coherently moving dots to capture the motion directions of a large number of unambiguous, but incoherently moving, dots. Thus, although individual dots do not experience an aperture problem, a sufficient number of randomly moving dots create informational uncertainty that creates a similar computational problem for the visual system to solve. The intuitive idea is that the MT-MST feedback loop needs more time to capture the incoherent motion signals when there are more of them, and cannot achieve as high a level of asymptotic response magnitude when more of them compete with the emerging winning direction. In other words, the effectiveness of the motion capture process depends on input coherence. LIP then converts the distributed motion directional inputs from MST into a directional eye movement command, and thereby enables the monkey to report its decision via a saccade.

References

- Abbott, L.F., Sen, K., Varela, J.A. & Nelson, S.B. (1997). Synaptic depression and cortical gain control. *Science*, 275, 220-222.
- Adelson, E.H. & Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523-525.
- Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, 52, 1106-1130.
- Anderson, J.C., Binzegger, T., Martin, K.A. & Rockland., K.S. (1998). The connection from cortical area V1 to V5: a light and electron microscopic study. *Journal of Neuroscience*, 18, 10525-10540.
- Anderson, J.C. and Martin, K.A. (2002). Connection from cortical area V2 to MT in macaque monkey. *Journal of Comparative Neurology*, 443, 56-70.
- Anstis, S. M. (1990). Imperceptible intersections: the chopstick illusion. In *AI and the Eye*. New York, NY: Wiley.
- Bakin, J.S., Nakayama, K. & Gilbert, C.D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience*, 20, 8188-8198.
- Baloch, A.A. & Grossberg, S. (1997). A neural model of high-level motion processing: line motion and formotion dynamics. *Vision Research*, 37, 3037-3059.
- Baloch, A.A., Grossberg, S., Mingolla, E. & Nogueira, C.A.M. (1998). A neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America A*, 16, 953-978.
- Barlow, H.B. & Levick, W.R. (1965). The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology*, 178, 477-504.
- Berezovskii, V. & Born, R.T. (1999). Specificity of projections from wide-field and local motion-processing regions within the middle temporal visual area of the owl monkey. *The Journal of Neuroscience*, 20, 1157-1169.
- Berzhanskaya, J., Grossberg, S. & Mingolla, E. (2007). Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. *Spatial Vision*, 20, 337-395.
- Bichot, N.P. & Schall, J.D. (1996). Thompson KG Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, 381, 697-699.
- Born, R.T. (2000). Center-surround interactions in the middle temporal visual area of the owl monkey. *Journal of Neurophysiology*, 84, 2658-2669.
- Born, R. & Tootell, R. (1992). Segregation of global and local motion processing in macaque middle temporal cortex. *Nature*, 357, 497-499.
- Braddick, O.J. (1980). Low-level and high-level processes in apparent motion. *Philosophical Transactions of the Royal Society of London Series B*, 290, 137-151.
- Bradley, D.C. and Andersen, R.A. (1998) Center-surround antagonism based on disparity in primate area MT. *Journal of Neuroscience*, 18, 552-565.
- Bradley, D.C., Chang G.C. & Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature*, 392, 714-717.
- Bregman, A.L. (1981). Asking the "what for" question in auditory perception. In M. Kubovy and J. R. Pomerantz (eds.), *Perceptual organization*. Hillsdale, NJ: Earlbaum Associates.
- Browning, A., Grossberg, S., and Mingolla, M. (2009). Cortical dynamics of navigation and steering in natural scenes: Motion-based object segmentation, heading, and obstacle avoidance. *Neural Networks*, 22, 1383-1398.

- Bullier, J. (2001). Integrated model of visual processing. *Brain Research - Brain Research Reviews*, 36, 96-107.
- Callaway, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, 21, 47-74.
- Cao, Y. & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, 18, 515-578.
- Carandini, M. & Ferster, D. (1997). Visual adaptation hyperpolarizes cells of the cat striate cortex. *Science*, 276, 949.
- Castet, E., Lorenceau, J., Shiffrar, M. & Bonnet, C. (1993) Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, 33, 1921-1936.
- Chance, F.S., Nelson, S.B. & Abbott, L.F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience*, 18, 4785-4799.
- Chey, J., Grossberg, S. & Mingolla, E. (1997). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769-2786.
- Chey, J., Grossberg, S. & Mingolla, E. (1998). Neural dynamics of motion processing and speed discrimination. *Vision Research*, 38, 2769-2786.
- DeAngelis, G.C., Cumming, B.G., & Newsome, W.T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature*, 394, 677-680.
- DeAngelis, G.C. & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology*, 89, 1094-1111.
- DeYoe, E.A. & Van Essen, D.C. (1985). Segregation of efferent connections and receptive field properties in visual area V2 of the macaque. *Nature*, 317, 58-61.
- Dresp, B., Durand, S. & Grossberg, S. (2002). Depth perception from pairs of overlapping cues in pictorial displays. *Spatial Vision*, 15, 255-276.
- Duncan, R.O., Albright, T.D. & Stoner, G.R. (2000). Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *Journal of Neuroscience*, 20, 5885-5897.
- Duncker, K. (1937). Induced motion. In W.E. Ellis (Ed.), *A Sourcebook of Gestalt Psychology*. London: Routledge and Kegan Paul. (Original work published in 1929).
- Eifuku, S. & Wurtz, R.H. (1998). Response to motion in extrastriate area MSTl: Center-surround interactions. *Journal of Neurophysiology*, 80, 282-296.
- Enroth-Cugell, C. & Robson, J. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, 187, 517-552.
- Fang, L. & Grossberg, S. (2007). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45-82.
- Ferrera, V.P. & Wilson, H.R. (1987). Direction specific masking and the analysis of motion in two dimensions. *Vision Research*, 27, 1783-1796.
- Ferrera, V.P. & Wilson, H.R. (1990). Perceived direction of moving two-dimensional patterns. *Vision Research*, 30, 273-387.
- Ferrera, V.P. & Wilson, H.R. (1991). Perceived speed of moving two-dimensional patterns. *Vision Research*, 31, 877-893.
- Francis, G., Grossberg, S., Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089-1104.
- Fried, S.I., Münch, T.A. & Werblin, F.S. (2002). Mechanisms and circuitry underlying directional selectivity in the retina. *Nature*, 420, 411-414.
- Grossberg, S. (1968). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, 60, 758-765.

- Grossberg, S. (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences*, 15, 253-285.
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213-257.
- Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen and F. Snell (Eds.), *Progress in theoretical biology, Volume 5* (p. 233-374). New York, NY: Academic Press.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1-51.
- Grossberg, S. (1991). Why do parallel cortical systems exist for the perception of static form and moving form? *Perception and Psychophysics*, 49, 117-141.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55, 48-121.
- Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review*, 104, 618-658.
- Grossberg, S. (1998). How is a moving target continuously tracked behind occluding cover? In T. Watanabe (Ed.), *High level motion processing: Computational, neurobiological, and psychophysical perspectives*, pp. 3-52. Cambridge, MA : MIT Press.
- Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12, 163-185.
- Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition*, 8, 1-44.
- Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4, 233-246.
- Grossberg, S., Mingolla, E. & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, 41, 2351-2553.
- Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, 2, 47-76.
- Grossberg, S. & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92, 173-211
- Grossberg, S. & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, 38, 141-171.
- Grossberg, S. & Pearson, L. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Towards a unified theory of how the cerebral cortex works. *Psychological Review*, 115, 677-732.
- Grossberg, S. & Pilly, P. (2008). Temporal dynamics of decision-making during motion perception in the visual cortex. *Vision Research*, 48, 1345-1373.
- Grossberg, S. & Raizada, R.D. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413-1432.
- Grossberg, S. & Swaminathan, G. (2004). A laminar cortical model for 3D perception of slanted and curved surfaces and of 2D images: development, attention, and bistability. *Vision Research*, 44, 1147-1187.
- Grossberg S. & Williamson J.R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex*, 11, 37-58.
- Grossberg, S. & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception:

- Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725-1743.
- Grunewald, A., Bradley, D.C. & Andersen, R.A. (2002). Neural correlates of structure-from-motion perception in macaque V1 and MT. *Journal of Neuroscience*, 22, 6195-6207.
- Hochstein, S. & Shapley, R.M. (1976a). Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *Journal of Physiology*, 262, 265-284.
- Hochstein, S. & Shapley R.M. (1976b). Quantitative analysis of retinal ganglion cell classifications. *Journal of Physiology*, 262, 237-264.
- Horwitz, G.D. & Newsome, W.T. (2001a). Target Selection for Saccadic Eye Movements: Direction-Selective Visual Responses in the Superior Colliculus. *Journal of Neurophysiology*, 86, 2527-2542.
- Horwitz, G.D., Batista, A.P. & Newsome, W.T. (2004a). Direction-selective visual responses in macaque superior colliculus induced by behavioral training. *Neuroscience Letters*, 366, 315-319.
- Hupé, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P. & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394, 784-787.
- Hupé, J.M. & Rubin, N. (2003). The dynamics of bi-stable alternation in ambiguous motion displays: a fresh look at plaids. *Vision Research*, 43, 531-548.
- Jones, H.E., Grieve, K.L., Wang, W. & Sillito, A.M. (2001). Surround suppression in primate V1. *Journal of Neurophysiology*, 86, 2011-2028.
- Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception*. New York, NY: Praeger Press.
- Kelly, F. J. & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics*, 62, 1596-1619.
- Kennedy, H. & Bullier, J. (1985). A double-labeling investigation of the afferent connectivity to cortical areas V1 and V2 of the macaque monkey. *Journal of Neuroscience*, 5, 2815-2830.
- Kim, J. & Wilson, H.R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research*, 33, 2479-2489.
- Kolers, P. (1972). *Aspects of motion perception*. Oxford, UK: Pergamon Press.
- Lidén, L. & Mingolla, E. (1998). Monocular occlusion cues alter the influence of terminator motion in the barber pole phenomenon. *Vision Research*, 38, 3883-3898.
- Lindsey, D.T. & Todd, J.T. (1996). On the relative contributions of motion energy and transparency to the perception of moving plaids. *Vision Research*, 36, 207-222.
- Livingstone, M.S. (1998). Mechanisms of direction selectivity in macaque V1. *Neuron*, 20, 509-526.
- Livingstone, M.S. & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, 240, 740-749.
- Livingstone, M.S. & Conway, B.R. (2003). Substructure of direction-selective receptive fields in macaque V1. *Journal of Neurophysiology*, 89, 2743-2759.
- Livingstone, M.S. & Hubel, D.H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, 4, 309-356.
- Lorenceau, J. & Alais, D. (2001). Form constraints in motion binding. *Nature Neuroscience*, 4, 745-751.
- Marr, D. & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London B*, 211, 151-180.
- Movshon, J.A. & Newsome, W.T. (1996). Visual response properties of striate cortical neurons

- projecting to area MT in macaque monkeys. *Journal of Neuroscience*, 16, 7733-7741.
- Murthy, A. & Humphrey, A.L. (1999). Inhibitory contributions to spatiotemporal receptive-field structure and direction selectivity in simple cells of cat area 17. *Journal of Neurophysiology*, 81, 1212-1224.
- Nakayama, K., Shimojo, S. & Silverman, G.H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55-68.
- Pack, C.C. & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, 409, 1040-1042.
- Pack, C.C., Gartland, A.J. & Born, R.T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *Journal of Neuroscience*, 24, 3268-3280.
- Pack, C., Grossberg, S. & Mingolla, E. (2001). A neural model of smooth pursuit control and motion perception by cortical area MST. *Journal of Cognitive Neuroscience*, 13, 102-120.
- Palanca, B.J. & DeAngelis, G.C. (2003). Macaque middle temporal neurons signal depth in the absence of motion. *Nature Neuroscience*, 11, 216-223.
- Ponce, C.R., Lomber, S.G. & Born R.T. (2008). Integrating motion and depth via parallel pathways. *Society for Neuroscience abstracts*, Atlanta, GA.
- Raizada RD. & Grossberg S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, 13, 100-113.
- Reichardt W. (1961). In W.A. Rosenblith (ed.), *Sensory Communication* (p. 303). New York, NY: Wiley.
- Rockland, K.S. (1992) Laminar distribution of neurons projecting from area V1 to V2 in macaque and squirrel monkeys. *Cerebral Cortex*, 2, 38-47.
- Rockland, K.S. (1995). Morphology of individual axons projecting from area V2 to MT in the macaque. *Journal of Comparative Neurology*, 355, 15-26.
- Rockland, K.S. (2002). Visual cortical organization at the single axon level: a beginning. *Neuroscience Research*, 42, 155-166.
- Rockland, K.S. & Pandya, D.N. (1981). Cortical connections of the occipital lobe in the rhesus monkey: interconnections between areas 17, 18, 19 and the superior temporal sulcus. *Brain Research*, 212, 249-270.
- Roitman, J.D. & Shadlen, M.N. (2002). Response of Neurons in the Lateral Intraparietal Area during a Combined Visual Discrimination Reaction Time Task. *Journal of Neuroscience*, 22, 9475-9489.
- Rust, N.C., Majaj, N.J., Simoncelli, E.P. & Movshon, J.A. (2002). Gain control in Macaque area MT is directionally selective. *Society for Neuroscience Abstracts*, Orlando, Florida.
- Shadlen, M.N. & Newsome, W.T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *Journal of Neurophysiology*, 86, 1916-1936.
- Shimojo, S., Silverman, G.H. & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619-626.
- Shipp, S. & Zeki, S. (1985). Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex. *Nature*, 315, 322-325.
- Shipp, S. & Zeki, S. (1989). The organization of connections between areas V5 and V1 in Macaque visual cortex. *European Journal of Neuroscience*, 1, 309-332.
- Sincich, L.C. & Horton, J.C. (2002). Divided by cytochrome oxidase: a map of the projections from V1 to V2 in macaques. *Science*, 295, 1734-1737.

- Sincich, L.C. & Horton, J.C. (2003). Independent projection streams from macaque striate cortex to the second visual area and middle temporal area. *Journal of Neuroscience*, *23*, 5684-5692.
- Snowden, R.J., Treue, S., Erickson R.G. & Andersen R.A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, *11*, 2768-2785.
- Stone, L.S., Watson, A.B. & Mulligan, J.B. (1990). Effect of contrast on the perceived direction of a moving plaid. *Vision Research*, *30*, 1049-1067.
- Stoner, G.R., Albright, T.D. & Ramachandran, V.S. (1990). Transparency and coherence in human motion perception. *Nature*, *344*, 153-155.
- Stoner, G.R. & Albright, T.D. (1998). Luminance contrast affects motion coherency in plaid patterns by acting as a depth-from-occlusion cue. *Vision Research*, *38*, 387-401.
- Tanaka, K., Sugita, Y., Moriya, M. & Saito, H. (1993). Analysis of object motion in the ventral part of the medial superior temporal area of the macaque visual cortex. *Journal of Neurophysiology*, *69*, 128-142.
- Treue, S. & Maunsell, J.H. (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.
- Trueswell, J.C. & Hayhoe, M.M. (1993). Surface segmentation mechanisms and motion perception. *Vision Research*, *33*, 313-323.
- Tse, P.U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research*, *45*, 1095-1098.
- Vallortigara, G., Bressan, P. & Bertamini, M. (1988). Perceptual alternations in stereokinesis. *Perception*, *17*, 4-31.
- van Santen, J.P. & Sperling, G. (1985) Elaborated Reichardt detectors. *Journal of the Optical Society of America A.*, *2*, 300-321.
- Varela, J.A., Sen, K., Gibson, J., Fost, J., Abbott, L.F. & Nelson., S.B. (1997). A quantitative description of short-term plasticity at excitatory synapses in Layer 2/3 of rat primary visual cortex. *Journal of Neuroscience*, *17*, 7926-7940.
- Wallach, H. (1935). On the visually perceived direction of motion. *Psychologische Forschung*, *20*, 325-380.
- Wallach, H. (1976). *On Perception*. New York: Quadrangle.
- Weiss, Y. & Adelson, E.H. (2000). Adventures with gelatinous ellipses—constraints on models of human motion analysis. *Perception*, *29*, 543-566.
- Wuerger, S., Shapley, R. & Rubin, N. (1996). ‘On the visually perceived direction of motion’ by Hans Wallach: 60 years later. *Perception*, *25*, 1317-1367.
- Xiao, D.K., Raviguel, S., Marcar, V. & Orban, G.A. (1997). The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cerebral Cortex*, *7*, 662-677.
- Xu, X., Bonds, A.B. & Casagrande, V.A. (2002). Modeling receptive-field structure of koniocellular, magnocellular, and parvocellular LGN cells in the owl monkey (*Aotus trivigatus*). *Visual Neuroscience*, *19*, 703-711.
- Yabuta, N.H., Sawatari, A. & Callaway, E.M. (2001). Two functional channels from primary visual cortex to dorsal visual cortical areas. *Science*, *292*, 297-300.
- Yabuta, N.H. & Callaway, E.M. (1998) Functional streams and local connections of layer 4C neurons in primary visual cortex of the macaque monkey. *Journal of Neuroscience*, *18*, 9489-9499.

- Yo, C. & Wilson, H.R. (1991). Perceived speed of moving two-dimensional patterns depends on duration, contrast and eccentricity. *Vision Research*, 32, 877-893.
- Zhou, H., Friedman, H.S. & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, 20, 6594-6611