



Review

Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support

Stephen Grossberg *

*Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA**Graduate Program in Cognitive and Neural Systems, Departments of Mathematics & Statistics, Psychological & Brain Sciences, and Biomedical Engineering Boston University, 677 Beacon Street, Boston, MA 02215, USA*

ARTICLE INFO

Article history:

Received 2 August 2016

Received in revised form 21 October 2016

Accepted 20 November 2016

Available online 6 December 2016

Keywords:

Consciousness

Adaptive resonance

Attention

Vision

Audition

Emotion

ABSTRACT

The hard problem of consciousness is the problem of explaining how we experience qualia or phenomenal experiences, such as seeing, hearing, and feeling, and knowing what they are. To solve this problem, a theory of consciousness needs to link brain to mind by modeling how emergent properties of several brain mechanisms interacting together embody detailed properties of individual conscious psychological experiences. This article summarizes evidence that Adaptive Resonance Theory, or ART, accomplishes this goal. ART is a cognitive and neural theory of how advanced brains autonomously learn to attend, recognize, and predict objects and events in a changing world. ART has predicted that “all conscious states are resonant states” as part of its specification of mechanistic links between processes of consciousness, learning, expectation, attention, resonance, and synchrony. It hereby provides functional and mechanistic explanations of data ranging from individual spikes and their synchronization to the dynamics of conscious perceptual, cognitive, and cognitive-emotional experiences. ART has reached sufficient maturity to begin classifying the brain resonances that support conscious experiences of seeing, hearing, feeling, and knowing. Psychological and neurobiological data in both normal individuals and clinical patients are clarified by this classification. This analysis also explains why not all resonances become conscious, and why not all brain dynamics are resonant. The global organization of the brain into computationally complementary cortical processing streams (complementary computing), and the organization of the cerebral cortex into characteristic layers of cells (laminar computing), figure prominently in these explanations of conscious and unconscious processes. Alternative models of consciousness are also discussed.

© 2016 The Author. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Contents

1. Towards solving the hard problem of consciousness	40
1.1. What is the hard problem?	40
1.2. To what extent can a scientific theory clarify the Hard Problem?	40
1.3. Establishing a linking hypothesis between resonant brain dynamics and the conscious mind.....	41
1.4. Adaptive Resonance Theory links resonant dynamics and the conscious mind	42
2. The stability-plasticity dilemma and rapid learning throughout life	42
3. Equations, modules, and modal architectures	44
4. Complementary processing streams for perception/cognition and space/action	44
5. Laminar neocortical circuits to represent higher-order biological intelligence	45
6. An example of complementary computing: Visual boundaries and surfaces.....	45
6.1. Neon color spreading, end gaps, and end cuts	46
6.2. Simple cells and hierarchical resolution of uncertainty.....	47
6.3. Recognizing without seeing: Complex cells imply that all boundaries are invisible	47

* Correspondence to: Center for Adaptive Systems, Boston University, 677 Beacon Street, Boston, MA 02215, USA. Fax: +617 353 7755.

E-mail address: steve@bu.edu.

URL: <http://cns.bu.edu/~steve>.

6.4.	Boundary completion and bipole grouping cells	48
6.5.	Neon color spreading: Boundary-surface interactions and hierarchical resolution of uncertainty	48
6.6.	Why resonance?	50
7.	An example of complementary computing: Attention vs. orienting during category learning	50
7.1.	Object attention by a top-down, modulatory on-center, off-surround network	50
7.2.	ART hypothesis testing and learning of a predictive recognition category	51
7.3.	Complementary PN and N200 ERPs during attention and memory search	52
7.4.	Combining paradigms to disclose finer details about category learning	52
7.5.	Resonance between distributed features and compressed categories solves symbol grounding	53
7.6.	Feature-category resonances support object attention, learning, and recognition	53
8.	Linking brain to behavior: All conscious states are resonant states, but not conversely	53
9.	The varieties of brain resonances and the conscious experiences that they support	53
10.	Surface-shroud resonances for conscious seeing of visual qualia	54
10.1.	Surface-shroud resonances are generated between V4 and PPC	54
10.2.	V2 and V4: Recognizing occluded objects, seeing unoccluded surfaces, and transparency	54
10.3.	Conscious seeing, orienting, and reaching	55
10.4.	A link between object learning, recognition, and seeing	56
10.5.	Explaining data about visual neglect: Coordinates, competition, grouping, and action	57
10.6.	Explaining data about visual crowding: A unified view of crowding, neglect, and search	58
10.7.	Explaining data about change blindness and motion-induced blindness	59
10.8.	Explaining many data with the same model mechanisms	60
10.9.	Shrouds have complex internal structure that influences properties of visual search	60
11.	From view- and position-specific categories to view- and position-invariant categories	60
11.1.	Conscious visual qualia are computed in retinotopic coordinates	60
11.2.	Visual imagery: Basal ganglia volition converts top-down modulation to driving inputs	60
12.	Learning invariant object categories using Where-to-What stream interactions	60
12.1.	From complementary consistency to figure-ground separation: Surface contours	60
12.2.	The main problem: Why inhibiting view categories does not inhibit invariant categories	62
12.3.	Attentional shroud inhibits reset of invariant object category during object search	62
12.4.	Human and monkey data support shroud reset properties: Explanations and predictions	62
12.5.	A surface-shroud resonance enables the eyes to explore multiple object views	63
12.6.	Predictive remapping: Gain fields maintain shroud stability	63
12.7.	Both retinotopic and spatial coordinates are needed during active vision	63
12.8.	Perceptual stability: Binocular fusion during eye movements	63
12.9.	Two types of perceptual stability cooperate during active conscious vision	64
13.	Seeing and knowing: Synchronous surface-shroud and feature-category resonances	64
14.	What and Where/How streams for audition	64
14.1.	Ventral sound-to-meaning vs. dorsal sound-to-action: Complementary invariants	65
14.2.	Circular reactions in vision and audition: Babbling to reach and to speak	66
14.3.	Motor-equivalent reaching and speech production: Tool use and coarticulation	66
14.4.	From eating to speaking	67
14.5.	Auditory neglect and speech production deficits	68
15.	Stream-shroud resonances for conscious hearing of visual qualia	68
15.1.	Auditory scene analysis: Tracking sound sources through noise and overlapping frequencies	68
15.2.	Auditory continuity illusion and spectral-pitch resonances	69
15.3.	ARTSTREAM model and stable learning of pitch, timbre, and voice categories	69
15.4.	Stream-shroud resonances: Spatial attention and neglect in the auditory stream	70
16.	Item-list resonances for recognition of speech and language	71
16.1.	From streaming to speech: Item-and-Order working memories and list chunks	71
16.2.	Top-down attentive matching and item-list resonances	72
16.3.	Phonemic restoration: How the future can influence what is consciously heard in the past	72
16.4.	A shared design for linguistic, motor, and spatial working memories: LTM Invariance and Normalization	72
16.5.	Phonemic restoration in a laminar cortical model of speech perception	73
17.	Learning and storage of speaker-invariant and rate-invariant working memories	73
17.1.	Speaker normalization in the What stream: Circular reactions and imitative maps	74
17.2.	From auditory streaming to speaker normalization: A shared circuit design	75
17.3.	From speaker normalization to rate normalization	75
17.4.	Transient and sustained processing streams in audition and speech	76
17.5.	PHONET: Asymmetric T-to-S gain control creates within-syllable rate-invariance	76
17.6.	ARTPHONE: Rate-sensitive gain control creates rate-invariant working memories	76
17.7.	From masked priming to resonance in lexical decision and word frequency tasks	77
18.	From invariant working memory storage to volitionally-variant productions and percepts	78
18.1.	LIST PARSE: Volitionally-controlled variable-rate sequential performance	78
18.2.	lisTELOS: Basal ganglia volitional control and Item-Order-Rank working memory	78
18.3.	From learning of invariant lists to rhythmic intonated performance	79
18.4.	From speaking to singing	80
19.	The feeling of what happens: Cognitive-emotional resonances for conscious feelings	80
19.1.	Resonance between object, value, and object-value categories: Beyond Chomsky and Skinner	80
19.2.	Reinforcement learning: Conditioned reinforcer and incentive motivational learning	81
19.3.	Interactions between sensory cortices, thalamus, amygdala, and orbitofrontal cortex	82
19.4.	The feeling of what happens, core consciousness, dual competition, and survival circuits	82
19.5.	Breakdowns during mental disorders: Theory of Mind, autism, and schizophrenia	83

19.6.	ARTSCAN Search: Solving the Where's Waldo problem	83
19.7.	Conscious vs. non-conscious emotions	84
19.8.	Conditioning and consciousness: Trace conditioning, hippocampus, and time cells	85
19.9.	Why a Weber law? Distinguishing expected and unexpected disconfirmations.....	86
19.10.	Effects of hippocampal, amygdala, and orbitofrontal lesions on learning and memory.....	86
20.	Some other approaches to consciousness	86
21.	Concluding remarks	88
21.1.	Unified experiences of perceiving, knowing, and feeling	88
21.2.	Using self-organizing and self-stabilizing expertise to control changing bodies.....	89
	References.....	89

1. Towards solving the hard problem of consciousness

1.1. What is the hard problem?

This article summarizes an emerging theory of the events that occur in our brains when we have conscious experiences. Before turning to the scientific results themselves, it may be helpful to summarize the explanatory goals of the article, as well as the limits to what such a theory can achieve.

First, what is the “hard problem of consciousness”? Wikipedia (https://en.wikipedia.org/wiki/Hard_problem_of_consciousness) says: “The **hard problem of consciousness** is the problem of explaining how and why we have qualia or phenomenal experiences—how sensations acquire characteristics, such as colors and tastes”. David Chalmers, who introduced the term ‘hard problem’ of consciousness, contrasts this with the ‘easy problems’ of explaining the ability to discriminate, integrate information, report mental states, focus attention, etc. As Chalmers (1995) has noted: “The really hard problem of consciousness is the problem of *experience*. When we think and perceive, there is a whir of information-processing, but there is also a subjective aspect. As Nagel (1974) has put it, there is *something it is like* to be a conscious organism. This subjective aspect is experience. When we see, for example, we *experience* visual sensations: the felt quality of redness, the experience of dark and light, the quality of depth in a visual field. Other experiences go along with perception in different modalities: the sound of a clarinet, the smell of mothballs. Then there are bodily sensations, from pains to orgasms; mental images that are conjured up internally; the felt quality of emotion, and the experience of a stream of conscious thought. What unites all of these states is that there is something it is like to be in them. All of them are states of experience”.

The Internet Encyclopedia of Philosophy (<http://www.iep.utm.edu/hard-con/>) goes on to say: “The hard problem of consciousness is the problem of explaining why any physical state is conscious rather than nonconscious. It is the problem of explaining why there is “something it is like” for a subject in conscious experience, why conscious mental states “light up” and directly appear to the subject. The usual methods of science involve explanation of functional, dynamical, and structural properties—explanation of what a thing does, how it changes over time, and how it is put together. But even after we have explained the functional, dynamical, and structural properties of the conscious mind, we can still meaningfully ask the question, *Why is it conscious?* This suggests that an explanation of consciousness will have to go beyond the usual methods of science. Consciousness therefore presents a hard problem for science, or perhaps it marks the limits of what science can explain. Explaining why consciousness occurs at all can be contrasted with so-called “easy problems” of consciousness: the problems of explaining the function, dynamics, and structure of consciousness. These features *can* be explained using the usual methods of science. But that leaves the question of why there is something it is like for the subject when these functions, dynamics, and structures are present. This is the hard problem...There seems to be an unbridgeable explanatory

gap between the physical world and consciousness. All these factors make the hard problem hard”.

Philosophical opinions about the Hard Problem are so diverse and strongly felt by their supporters, that even attempting a possible scientific approach to solving the Hard Problem may sometimes feel like stepping on a philosophical “third rail”. This is because philosophers vary passionately in their views between the claim that no Hard Problem remains once it is explained how the brain generates experience, as in the writings of Daniel Dennett, to the claim that it cannot in principle be solved by the scientific method, as in the writings of David Chalmers. See the above reference for a good summary.

1.2. To what extent can a scientific theory clarify the Hard Problem?

The present article presents a theory that proposes answers, however incomplete, that respond to various of these philosophical concerns by showing how this theory scientifically clarifies various of the distinctions that philosophers have discussed. Looking ahead, let me note the following:

First, it is fair to ask what kind of “event” occurs in the brain during a conscious experience that is anything more than just a “whir of information-processing”? What happens when conscious mental states “light up” and directly appear to the subject? This article proposes that, over and above “just” information processing, our brains sometimes go into a context-sensitive *resonant state* that can involve multiple brain regions. The article summarizes some of the abundant experimental evidence that “all conscious states are resonant states”. Not all brain dynamics are “resonant”, and thus consciousness is not just a “whir of information-processing”.

Second, it is fair to ask: When does a resonant state embody a conscious experience? “Why is it conscious”. And how do different resonant states support different kinds of conscious qualia? The other side of the coin is equally important: When does a resonant state fail to embody a conscious experience? In order to provide even partial answers to these questions, the theory will be used to explain how various evolutionary challenges that advanced brains face in order to adapt to changing environments in real time have been met with particular conscious states, and how these conscious states form part of larger adaptive behavioral capabilities. In this regard, the theory argues that humans are not conscious just to Platonically contemplate the beauty of the world. Humans are conscious in order to enable them to better adapt to the world’s changing demands. To illustrate these claims, the theory proposes how the brain generates resonances that support particular conscious experiences of seeing, hearing, feeling, and knowing. In so doing, it suggests how resonances for conscious seeing help to ensure effective reaching, resonances for conscious hearing help to ensure effective speaking, and resonances for conscious feeling help to ensure effective goal-directed action (Table 1). The theory also proposes why different resonances are triggered at the processing stages that they are, and suggests how several functionally different kinds of resonances can interact with one another, so that when we consciously see a familiar

Table 1

The visual, auditory, and emotional conscious percepts and recognition events that are explained in the text, along with the kinds of responses that they control further downstream.

VISUAL seeing, knowing, and reaching

AUDITORY hearing, knowing, and speaking

EMOTIONAL feeling, knowing, and acting

Table 2

Types of resonances and the conscious experiences that they embody.

TYPE OF RESONANCE	TYPE OF CONSCIOUSNESS
surface-shroud	see visual object or scene
feature-category	recognize visual object or scene
stream-shroud	hear auditory object or stream
spectral-pitch-and-timbre	recognize auditory object or stream
item-list	recognize speech and language
cognitive-emotional	feel emotion and know its source

valued object, we can also know some things about it, and have appropriate feelings that are directed towards it.

Towards this end, the article will summarize and explain six different types of neural representations of conscious qualia (Table 2), and will provide enough theoretical background and data explanations based on these representations to illustrate their explanatory and predictive power. The theory's explanations and predictions, many of them stated in this article for the first time, also shed further light on these issues by suggesting multiple kinds of experiments to deepen our mechanistic understanding of the brain mechanisms for generating conscious resonances.

1.3. Establishing a linking hypothesis between resonant brain dynamics and the conscious mind

Before presenting this theory, it is important to also ask: How far can any scientific theory go towards achieving this goal? Let us suppose that a theory exists whose neural mechanisms interact to generate dynamical states with properties that mimic the parametric properties of the individual qualia that we consciously experience, notably the spatio-temporal patterning and dynamics of the resonant neural representations that represent these qualia. Suppose that these resonant dynamical states, in addition to mirroring properties of subjective reports of these qualia, generate observable data about these experiences that are collected in psychological and non-invasive neurobiological experiments on humans, and are consistent with psychological, multiple-electrode neurophysiological data, and other types of neurobiological data that are collected from monkeys who experience the same stimulus conditions.

Given such detailed correspondences with experienced qualia and multiple types of data, it can be argued that these dynamical resonant states are not just "neural correlates of consciousness" (Chalmers, 2000; Mormann & Koch, 2007). Rather, they are mechanistic representations of the qualia that embody individual

conscious experiences on the psychological level. If such a correspondence between detailed brain representations and detailed properties of conscious qualia occurs for a sufficiently large body of psychological data, then it would provide strong evidence that these brain representations create and support these conscious experiences. A theory of this kind would have provided a *linking hypothesis* between brain dynamics and the conscious mind. Such a linking hypothesis between brain and mind must be demonstrated before one can claim to have a "theory of consciousness".

If, despite such a linking hypothesis, a philosopher or scientist claims that, unless one can "see red" or "feel fear" in a theory of the Hard Problem, then it does not contribute to solving that problem, then no scientific theory can ever hope to solve the Hard Problem. This is true because science as we know it cannot do more than to provide a mechanistic theoretical description of the dynamical events that occur when individual conscious qualia are experienced. However, as such a principled, albeit incrementally developing, theory of consciousness becomes available, including increasingly detailed psychological, neurobiological, and even biochemical processes in its explanations, it can dramatically shift the focus of discussions about consciousness, just as relativity theory transformed discussions of space and time, and quantum theory of how matter works. As in quantum theory, there are measurement limitations in understanding our brains. We can no more personally ride an electron than we can enter a neuron that is participating in a conscious experience. Such an empathic limitation has not deterred physicists from believing that they have acquired an effective understanding of the physical world, based on their ability to explain and predict enough facts about it. This article subscribes to the view that it will not deter psychologists and neurobiologists from believing that they have acquired an effective understanding of the Hard Problem, based on our own explanatory and predictive successes.

1.4. Adaptive Resonance Theory links resonant dynamics and the conscious mind

To this end, the current article summarizes a theory, called Adaptive Resonance Theory, or ART, that explains neural mechanisms and psychological functions of dynamical brain states whose properties mimic subjective properties of individual conscious experiences, even as they explain parametric properties of many psychological data about these experiences, including their spatio-temporal dynamics. ART also proposes in what brain regions these dynamical states may occur, and how these dynamical states may all arise from a unified set of shared neural mechanisms, even when they carry out different functional roles and adaptive behaviors (Table 2) using different parts of the brain.

Such a dynamical state is called an *adaptive resonance*. As explained in greater detail in subsequent sections, a *resonance* is a dynamical state during which neuronal firings across a brain network are amplified and synchronized when they interact via reciprocal excitatory feedback signals during a matching process that occurs between bottom-up and top-down pathways. Often the activities of these synchronized cells oscillate in phase with one another. Resonating cell activities also focus attention upon a subset of cells, thereby clarifying how the brain can become conscious of attended events. It is called an *adaptive resonance* because the resonant state can trigger learning within the adaptive weights, or long-term memory (LTM) traces, that exist at the synapses of these pathways. ART explains how an adaptive resonance, when it occurs in different specialized anatomies that receive different types of environmental and brain inputs, can give rise to different conscious qualia as emergent, or interactive, properties of specialized networks of neurons. In particular, ART proposes how brain resonances mirror parametric properties of conscious experiences of seeing, hearing, and feeling, and of knowing what these experiences are (Table 2).

Not all cells that sustain an adaptive resonance may represent conscious contents of an experience. ART explains why only cells that are sensitive to the kinds of psychological discriminations (e.g., color, fear) that are embodied within the conscious experience can do this, and why these discriminations can occur only at prescribed stages of brain processing. Neurophysiological data about perceptual experiences like binocular rivalry illustrate this fact (e.g., Logothetis, 1998) and neural models have explained, and indeed predicted, key properties of these data (Grossberg, 1987; Grossberg, Yazdanbakhsh, Cao, & Swaminathan, 2008). In this regard, ART predicts that “all conscious states are resonant states”, but the converse statement that “all resonant states are conscious states” is not true. If, in fact, the concept of resonance is more general than that of consciousness, then such a theory needs to explain why that is so, and how conscious states arise from that more general function (see Section 8). Brain dynamics can also occur that do not include any possibility of resonance, notably during the generation of spatial and motor representations, and thus may not support any conscious experience (see Section 4).

The current article proposes answers to these questions. In addition, by explaining neural mechanisms and psychological functions of adaptive resonances, and specifying the functional, anatomical, and neurophysiological differences between resonances that can support conscious experiences of seeing, hearing, feeling, and knowing, ART provides, as noted above, a guide to how many different kinds of future experiments may be designed to provide more information about the brain resonances that support conscious experiences of qualia.

Before explaining how ART can clarify the brain substrates of conscious experiences, it is useful to realize that ART was not developed to explain consciousness. Rather, it was developed to explain and predict large psychological and neurobiological databases about learning, perception, cognition, and emotion. In order to achieve a self-contained exposition, notably to enable new

data explanations and predictions to be understood about ART, it is therefore necessary to review aspects of how ART explains data whose properties naturally lead to neural representations of conscious qualia. It also needs to be emphasized that the current article explains much more data about consciousness than has previously been done using ART or any other theory. It extends ART, rather than just reviewing it. The article also unifies parallel streams of modeling activity that have been separately developed, but whose synthesis is needed to draw many of the article's conclusions about consciousness. Key properties of these separate modeling streams thus need to be reviewed, so that their synthesis can be described in a self-contained way. This synthesis and extension of the ART model is called the *conscious ART*, or cART, model to distinguish it from its precursors.

Even if one's personal definition of the Hard Problem excludes any scientific theory, or if one is not interested in the Hard Problem, one can benefit from how the current cART theory explains the functional meaning of many different kinds of psychological and neurobiological data, both normal and clinical, that no other theory has yet explained, and makes predictions to further test these explanations.

2. The stability–plasticity dilemma and rapid learning throughout life

Humans are able to rapidly learn enormous amounts of new information, on their own, throughout life and to integrate all this information into unified conscious experiences that cohere into a sense of self. One has only to see an exciting movie just once to marvel at this capacity, since we can describe many details about it later on, including consciously experienced moments of seeing, hearing, feeling, and knowing, even though the individual scenes flashed by quickly. Humans can rapidly learn new facts without being forced to just as rapidly forget what they already know, even if no one tells us how the rules of each environment differ or change through time. When such forgetting does occur, it is often called *catastrophic forgetting*.

Grossberg (1980) called the problem whereby the brain learns quickly and stably without catastrophically forgetting its past knowledge the *stability–plasticity dilemma*. ART was introduced to explain how brains solve the stability–plasticity dilemma. Since its introduction in Grossberg (1976a, 1976b), ART has been incrementally developed into a cognitive and neural theory of how the brain autonomously learns to attend, recognize, and predict objects and events in a changing world, without experiencing catastrophic forgetting. ART currently has the broadest explanatory and predictive range of available cognitive and neural theories.

How does consciousness enter the ART story? To solve the stability–plasticity dilemma, ART specified mechanistic links between brain processes of learning, expectation, attention, resonance, and synchrony to explain psychological data about seeing, hearing, feeling, and knowing. As these discoveries were made, it became clear that the resonances hereby discovered also exhibit parametrical properties of individual conscious experiences, as embodied within the spatio-temporal patterning of cell activities across networks of feature detectors. Indeed, early mathematical results about the brain's functional units of short-term memory (STM) and long-term memory (LTM) proved that the functional units of both STM and LTM are *distributed patterns* across networks of feature-selective cells (Grossberg, 1968a, 1968b, 1973). Later results showed how these distributed patterns form part of synchronous resonant states that focus attention upon the critical features that represent predictive neural information, including those feature patterns that embody the parametric properties of individual conscious experiences (Fig. 1). Such discoveries, repeated multiple times over a period of decades, led to a growing understanding of links between processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony. These processes are called the CLEARs processes.

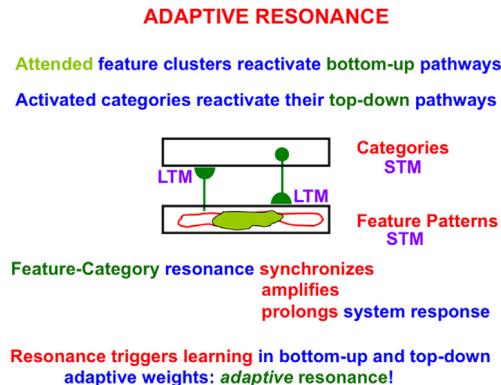


Fig. 1. During an adaptive resonance, attended feature patterns interact with recognition categories, both stored in short-term memory (STM), via positive feedback pathways that can synchronize, amplify, and prolong the resonating cell activities. Such a resonance can trigger learning in the adaptive weights, or long-term memory (LTM) traces, within both the bottom-up adaptive filter pathways and the top-down learned expectation pathways. In the present example, the resonance is a *feature-category resonance* (see Table 2).

ART predicts that all brain representations that solve the stability-plasticity dilemma use variations of CLEARs mechanisms (Grossberg, 1978a, 1980, 2007, 2013a). The CLEARs mechanisms clarify why many animals are intentional beings who pay attention to salient objects, why “all conscious states are resonant states”, and how brains can learn both *many-to-one maps* (representations whereby many object views, positions, and sizes all activate the same invariant object category; see Section 12) and *one-to-many maps* (representations that enable us to expertly know many things about individual objects and events; e.g., see Section 19.2).

ART accomplishes these properties by proposing how top-down expectations focus attention on salient combinations of cues, called *critical feature patterns* (Fig. 1); characterizes how attention may operate via a form of self-normalizing “biased competition” (Bhatt, Carpenter, & Grossberg, 2007; Carpenter & Grossberg, 1987, 1991; Desimone, 1998; Reynolds & Heeger, 2009); and explains how top-down attentive matching may help to solve the stability-plasticity dilemma. In particular, when a good enough match occurs between a bottom-up input pattern and a top-down expectation, a synchronous resonant state emerges that embodies an attentional focus. Such a resonance is capable of driving fast learning that incorporates the attended critical feature pattern into both the bottom-up adaptive filters that activate recognition categories, and the top-down expectations that are read-out by them – hence the name *adaptive resonance* – while suppressing unattended outliers that could have caused catastrophic forgetting.

The processes that are summarized by the CLEARs mnemonic are related to one another in specific ways that psychological and neurobiological data have supported. For example, attention and learning are linked in a manner consistent with ART predictions about visual perceptual learning (e.g., Ahissar and Hochstein (1993, 1997)), Ito, Westheimer, and Gilbert (1998), and Lu and Dosher (2004), auditory learning (e.g., Gao & Suga, 1998), and somatosensory learning (e.g., Krupa, Ghazanfar, and Nicolelis (1999) and Parker and Dostrovsky (1999)), including an explanation (Grossberg, 2003a) of when and how perceptual learning can occur without focused attention or conscious awareness (e.g., Seitz and Watanabe (2003) and Watanabe, Nanez, and Sasaki (2001)). ART also predicted links between attention and synchronous oscillations that were subsequently experimentally reported (e.g., Buschman and Miller (2007), Engel, Fries, and Singer (2001), Gregoriou, Goots, Zhou, and Desimone (2009), Grossberg (2009), and Pollen (1999)), and between synchronous oscillations and consciousness (e.g., Lamme (2006), Llinas, Ribary, Contreras, and Pedroarena (1998), and Singer (1998)). Although

several distinguished neurophysiologists have proposed a linkage between synchronous oscillations and consciousness, without a detailed computational theory to link mechanistic properties of resonant brain states to parametrical properties of individual conscious psychological experiences, and an explanation of where and why particular kinds of resonances occur in the brain to support different subjective experiences, this proposal does not rise to the level of a true theory of conscious awareness. Section 20 provides a comparative discussion of several approaches to explaining consciousness.

In the discussions below, links between attention and consciousness play an important role, and include several functionally different kinds of attention, such as *prototype attention* during recognition learning, as well as *boundary attention* and *surface attention* during visual perception. Some authors have nonetheless argued that visual attention and consciousness can be dissociated. For example, Koch and Tsuchiya (2007) note that “subjects can attend to a location for many seconds and yet fail to see one or more attributes of an object at that location...In lateral masking (*visual crowding*), the orientation of a peripherally-presented grating is hidden from conscious sight but remains sufficiently potent to induce an orientation-dependent aftereffect...” In fact, the data that Koch and Tsuchiya (2007) describe can be explained by how ART links attention and consciousness, notably by using properties of a *surface-shroud resonance* (Table 2). Properties of conscious visual crowding experiences follow immediately from properties of the surface-shroud resonances that are formed in response to visual crowding stimuli, as explained in Section 10.6.

Since ART was introduced in Grossberg (1976a, 1976b), it has undergone continual development to explain and predict increasingly large behavioral and neurobiological databases, ranging from normal and abnormal aspects of human and animal perception and cognition, to the spiking and oscillatory dynamics of hierarchically-organized laminar thalamocortical and corticocortical networks in multiple modalities. Indeed, some ART models individually explain, simulate, and predict behavioral, anatomical, neurophysiological, biophysical, and even biochemical data. ART currently provides functional and mechanistic explanations of data concerning such diverse topics as laminar cortical circuitry; invariant object and scenic gist learning and recognition; prototype, surface, and boundary attention; gamma and beta oscillations; learning of entorhinal grid cells and hippocampal place cells; computation of homologous spatial and temporal mechanisms in the entorhinal–hippocampal system; vigilance breakdowns during autism and medial temporal amnesia; cognitive–emotional interactions that focus attention on valued objects in an adaptively timed way; Item-Order-Rank working memories and learned list chunks for the planning and control of sequences of linguistic, spatial, and motor information; conscious speech percepts that are influenced by future context; auditory streaming in noise during source segregation; and speaker normalization. Brain regions whose functions are clarified by ART include visual and auditory neocortex; specific and nonspecific thalamic nuclei; inferotemporal, parietal, prefrontal, entorhinal, hippocampal, parahippocampal, perirhinal, and motor cortices; frontal eye fields; supplementary eye fields; amygdala; basal ganglia; cerebellum; and superior colliculus. These unified explanations of many different types of data increased confidence in the emerging classification of resonances.

The article first describes resonances that support (Table 1) visual seeing, knowing, and reaching (Sections 6, 7, 10–13); then auditory hearing, knowing, and speaking (Sections 14–18); and finally emotional feeling, knowing, and acting (Section 19), while also proposing how these various kinds of resonances may interact together. A comparative discussion of some other prominent contributions to consciousness studies is then provided (Section 20), followed by concluding remarks (Section 21).

Table 3
Some complementary brain processes and their anatomical substrates.

SOME COMPLEMENTARY PROCESSES	
Visual Boundary Interblob Stream V1-V4	Visual Surface Blob Stream V1-V4
Visual Boundary Interblob Stream V1-V4	Visual Motion Magno Stream V1-MT
WHAT Stream Perception & Recognition Inferotemporal and Prefrontal areas	WHERE Stream Space & Action Parietal and Prefrontal areas
Object Tracking MT+ Interbands and MSTv	Optic Flow Navigation MT+ Bands and MSTd
Motor Target Position Motor and Parietal Cortex	Volitional Speed Basal Ganglia

Much of this text tries to explain *how* key design principles and neural mechanisms control the brain's processing of visual, auditory, cognitive, and affective signals. This includes a frontal attack on how the brain learns to attend, recognize, and predict such events. As noted above, *how* such a capability works involves linked processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony in each of these domains (the CLEARS processes). The text also explains *what* the visual, auditory, cognitive, and affective representations are that emerge from these processes, because if one does not understand a lot about how vision, audition, affect, and recognition actually work, then one cannot say anything deep about how, when, and why their representations of individual experiences become conscious.

Along the way, the text proposes explanations of *how* and *why* the resonances that involve these processes *embody* the subjective properties of conscious qualia, not just correlate with them, including *how* and *why* these resonances occur at the processing stages that they do.

3. Equations, modules, and modal architectures

Before describing how these resonances work, it is helpful to review the kind of mind-brain theory that enables links between brain mechanisms and psychological functions to be established, and how similar organizational principles and mechanisms, suitably specialized, can support conscious qualia across modalities.

One reason for this inter-modality unity is that a small number of equations suffice to model all modalities. These include equations for short-term memory, or STM; medium-term memory, or MTM; and long-term memory, or LTM, that were introduced in Grossberg (1968c, 1969); see Grossberg (2013b) for a review. These equations are used to define a somewhat larger number of modules, or microcircuits, that are also used in multiple modalities where they can carry out different functions within each modality. These modules include shunting on-center off-surround networks, gated dipole opponent processing networks, associative learning networks, spectral adaptively timed learning networks, and the like. Each of these types of modules exhibits a rich, but not universal, set of useful computational properties. For example, shunting on-center off-surround networks can carry out

properties like contrast-normalization, including discounting the illuminant; contrast-enhancement, noise suppression, and winner-take-all choice; short-term memory and working memory storage; attentive matching of bottom-up input patterns and top-down learned expectations; and synchronous oscillations and traveling waves. These equations and modules are specialized and assembled into *modal* architectures, where “modal” stands for different modalities of biological intelligence, including architectures for vision, audition, cognition, cognitive-emotional interactions, and sensory-motor control.

An integrated self is possible because it builds on a shared set of equations and modules within modal architectures that can interact seamlessly together.

Modal architectures are *general-purpose*, in that they can process any kind of inputs to that modality, whether from the external world or from other modal architectures. They are also *self-organizing*, in that they can autonomously develop and learn in response to these inputs. Modal architectures are thus less general than the von Neumann architecture that provides the mathematical foundation of modern computers, but much more general than a traditional AI algorithm. ART networks form part of several different modal architectures, including modal architectures that enable seeing, hearing, feeling, and knowing.

4. Complementary processing streams for perception/cognition and space/action

How are modal architectures organized with respect to one another? This is a question about the global organization of the brain and mind. Modal architectures illustrate at least two general computational paradigms.

The first paradigm is called Complementary Computing (Grossberg, 2000a). Complementary Computing describes how the brain is organized into complementary parallel processing streams whose interactions generate biologically intelligent behaviors (Table 3). A single cortical processing stream can individually compute some properties well, but cannot, by itself, process other computationally complementary properties. Pairs of complementary cortical processing streams interact, using multiple processing

stages, to generate emergent properties that overcome their complementary deficiencies to compute complete information with which to represent or control some faculty of intelligent behavior.

For example, the category learning, attention, recognition, and prediction circuits of ART are part of the ventral, or What, cortical processing stream for perception and cognition (Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983). The ventral stream exhibits properties that are often computationally complementary to those of the dorsal, or Where and How, cortical processing stream for spatial representation and action (Goodale & Milner, 1992; Mishkin, 1982; Mishkin et al., 1983). One reason for this What–Where complementarity is that the What stream learns object recognition categories that are substantially invariant under changes in an object's view, size, and position (see Section 12). These invariant object categories enable our brains to recognize valued objects without experiencing a combinatorial explosion. They cannot, however, locate and act upon a desired object in space. Where stream spatial and motor representations can locate objects and trigger actions towards them, but cannot recognize them. By interacting together, the What and Where streams can recognize valued objects and direct appropriate goal-oriented actions towards them (see Section 19.6).

Table 4 summarizes basic complementary properties of the What and Where cortical streams. As will be discussed in greater detail below, perceptual/cognitive processes in the What stream often use ART-like excitatory matching and match-based learning to create self-stabilizing categorical representations of objects and events that solve the stability–plasticity dilemma. They thereby enable increasing expertise, and an ever-expanding sense of self, to emerge throughout life. As further discussed in Section 7.1, excitatory matching by the ART Matching Rule enables us to prime our expectations to anticipate objects and events before they occur, and to focus attention upon expected objects and events when they do occur. No less important is the fact that good enough matches between expected and actual events trigger resonant states that can support learning of new recognition categories and refinement of old ones, while also triggering conscious recognition of the critical feature patterns that are attended as part of these percepts (see Fig. 1). Excitatory matching also controls reset of the attentional focus when bottom-up inputs significantly mismatch currently active top-down expectations. As will be described below in greater detail (see Section 7), cycles of resonance and reset underlie much of the brain's perceptual and cognitive category learning dynamics.

Complementary spatial/motor processes in the Where stream often use inhibitory matching and mismatch-based learning to continually update spatial maps and sensory-motor gains that can effectively control our changing bodies throughout life. As discussed in Section 14.3, inhibitory matching can take place between representations of where we want to move and where we are now, so that when we arrive at where we want to be, the match equals zero (see Fig. 27). Inhibitory matching by the Vector Associative Map, or VAM, Matching Rule thus cannot solve the stability–plasticity dilemma. Instead, spatial maps and motor gains experience catastrophic forgetting as they learn how to accurately control our changing bodies throughout life. Together these complementary processes create a self-stabilizing perceptual/cognitive front end in the What stream for learning about the world and becoming conscious of it, while it intelligently commands more labile spatial/motor processes in the Where stream that control our changing bodies.

Thus, if only due to the complementary organization of the brain, ART is not “a theory of everything”. Likewise, because excitatory matching is needed to generate resonances that support conscious internal representations, spatial and motor processes (“procedural memories”; Cohen & Squire, 1980; Mishkin, 1982; Scoville & Milner, 1957; Squire & Cohen, 1984) that use inhibitory matching cannot generate conscious internal representations.

5. Laminar neocortical circuits to represent higher-order biological intelligence

The second computational paradigm is called Laminar Computing. Laminar Computing describes how the cerebral cortex is organized into layered circuits whose specializations support all higher-order biological intelligence. Indeed, the laminar circuits of cerebral cortex seem to realize a revolutionary computational synthesis of the best properties of feedforward and feedback processing, digital and analog processing, and data-driven bottom-up processing and hypothesis-driven top-down processing (Grossberg, 2007, 2013a). ART mechanisms have, to the present, been naturally embodied in laminar cortical models of vision, speech, and cognition, specifically in the 3D LAMINART model of 3D vision and figure-ground separation (Cao & Grossberg, 2005, 2012; Fang & Grossberg, 2009; Grossberg, 1999, 2016a; Grossberg & Swaminathan, 2004; Grossberg & Versace, 2008; Grossberg & Yazdankhsh, 2005; Grossberg et al., 2008; Raizada & Grossberg, 2003), the cARTWORD model of speech perception (Grossberg & Kazerounian, 2011, 2016; Kazerounian & Grossberg, 2014), and the LIST PARSE model of cognitive working memory and chunking (Grossberg & Pearson, 2008; Silver, Grossberg, Bullock, Histed, & Miller, 2011). Each model uses variations of the same canonical laminar cortical circuitry, thereby providing additional insights into how specialized resonances use similar types of circuits to support different conscious experiences.

6. An example of complementary computing: Visual boundaries and surfaces

In order to understand how a surface-shroud resonance can support conscious seeing of visual qualia (Table 2), it is necessary to summarize basic cortical processes whereby the brain sees. These explanations will clarify why surface-shroud resonances are predicted to be triggered by interactions between prestriate visual cortical area V4 and the posterior parietal cortex (PPC), before propagating both top-down to lower cortical areas such as V2 and V1, and bottom-up to higher cortical areas such as prefrontal cortex (PFC). The basic idea is that multiple visual preprocessing stages are needed before the brain can construct, in V4, a 3D visual surface representation that is complete and stable enough to control effective behaviors, notably reaching to desired objects. This summary unifies for the first time three parallel streams of neural modeling development: A stream that proposes how visual boundaries and surfaces work; a stream that proposes how surface-shroud resonances work to control the learning of invariant object categories; and a stream that proposes how surface-shroud resonances support conscious seeing of visual qualia. This synthesis helps to explain why cART predicts that surface-shroud resonances are triggered between cortical areas V4 and PPC (Section 10.3). New data explanations are also provided using this synthesis, including data about visual neglect (Section 10.5), visual crowding (Section 10.6), and change blindness and motion-induced blindness (Section 10.7).

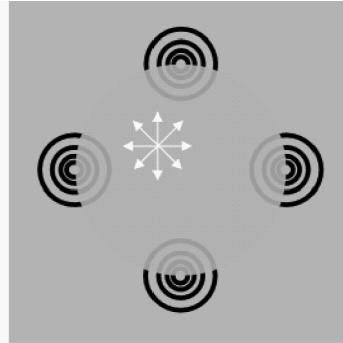
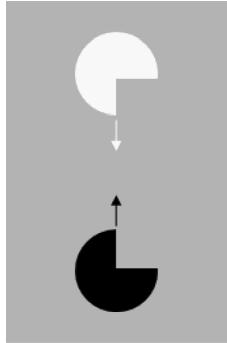
Perhaps the most basic fact about 3D vision and figure-ground perception is that its functional units are 3D *boundaries* and *surfaces*, where these words need to be properly understood. These processes were first modeled in Grossberg (1984a) and have provided a foundation for subsequent explanations and predictions about many data, including how looking at 2D pictures can generate conscious 3D percepts of occluding and occluded objects (e.g., Cao & Grossberg, 2005; Dresp, Durand, & Grossberg, 2002; Dresp-Langley & Grossberg, 2016; Fang & Grossberg, 2009; Grossberg, 1994, 1997, 2016a; Grossberg & Yazdankhsh, 2005; Kelly & Grossberg, 2000; Tanca, Grossberg, & Pinna, 2010).

Visual boundaries and surfaces are computed by the blob and interblob cortical processing streams that occur within and

Table 4

Complementary What and Where cortical stream properties. What cortical stream perceptual and cognitive representations can solve the stability-plasticity dilemma, using brain regions like inferotemporal (IT) cortex, where recognition categories are learned. These processes carry out excitatory matching and match-based learning. Where cortical stream spatial and motor processes, in contrast, do not solve the stability-plasticity dilemma, but rather adapt to changing bodily parameters, using brain regions like posterior parietal cortex (PPC). Whereas the recognition categories in the What cortical stream become increasingly invariant at higher cortical levels with respect to object views, positions, and sizes, the Where cortical stream elaborates spatial representations of object positions and mechanisms whereby to act upon them. Together the two streams can learn to recognize and become conscious of valued objects and scenes, while directing appropriate actions towards them.

WHAT		WHERE	
IT	PPC	IT	PPC
Spatially-invariant object learning and recognition		Spatially-variant reaching and movement	
Fast learning without catastrophic forgetting		Continually update sensory-motor maps and gains	
MATCHING		WHAT WHERE	
LEARNING		EXCITATORY	INHIBITORY
ART		MATCH	MISMATCH
VAM			

**Complementary Properties of Boundaries and Surfaces****Boundary Completion**

Inward Oriented
Insensitive to direction-of-contrast

Surface Filling-in

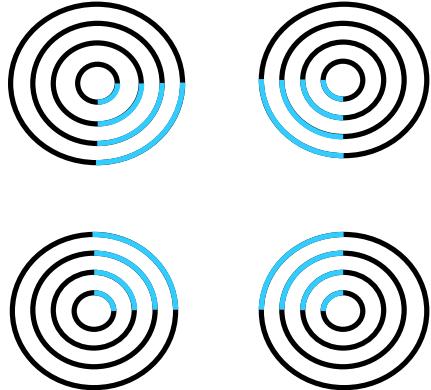
Outward Unoriented
Sensitive to direction-of-contrast

Fig. 2. Complementary computational properties of visual boundaries and surfaces. See text for details.

between cortical areas V1, V2, and V4, and provide a vivid example of Complementary Computing (Table 3 and Fig. 2). The visual illusion of neon color spreading dramatically illustrates the complementary properties of how boundaries are completed and surface brightnesses and colors are filled-in (Fig. 3).

6.1. Neon color spreading, end gaps, and end cuts

Neon color spreading was reported in Varin (1971), who studied a “chromatic spreading” effect that was induced when viewing an image similar to the one in Fig. 3. van Tuijl (1975) introduced images that gave rise to percepts that he called “neon-like color spreading”. Each black arc and blue arc in Fig. 3 generates boundaries in our brains. At the positions where these boundaries

**Fig. 3.** An example of neon color spreading. The image consists of black and blue circular arcs. The percept of the color blue filling a square is a visual illusion that is constructed by the brain. See text for details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

touch, the boundaries caused by black arcs cause small breaks, called *end gaps*, to occur in the boundaries caused by blue arcs, if the contrast of the black arcs with respect to the white background is larger than the contrast of the blue arcs with respect to the white background.

End gaps are created as follows: The boundary cells that are activated by the image in Fig. 3 are contrast-sensitive and orientationally tuned. The boundary cells that are activated by the black-white image contrasts thus become more active than the cells that are activated by blue-white contrasts. These contrast-sensitive boundary signals excite boundary cells at their own positions at the next processing stage, while inhibiting neighboring cells at this stage via a competitive network that acts across space; that is, by a short-range *spatial competition* (Fig. 4(a)). Due to the contrast-sensitivity of boundary cell responses, the stronger black-white boundary signals inhibit nearby blue-white boundary cells more than con-

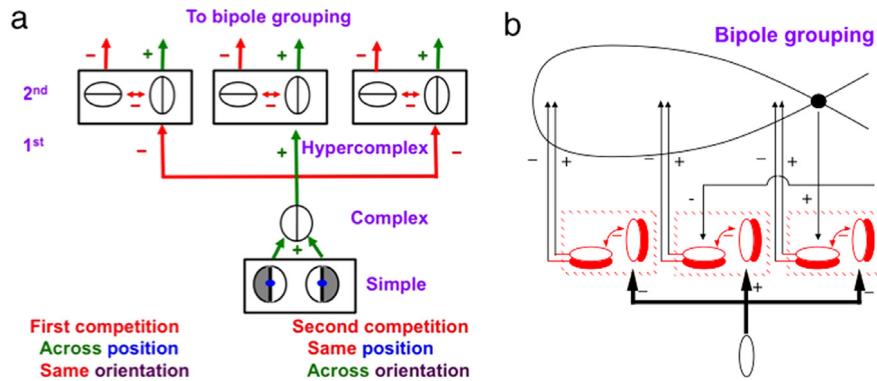


Fig. 4. A network of simple, complex, hypercomplex, and bipole cells begins the processing of perceptual boundaries. (a) Pairs of like-oriented simple cells with opposite contrast polarities at each position add their inputs to complex cells. Complex cells input to hypercomplex cells through a short range spatial competition (1st), followed by an orientational competition at each position (2nd). The spatial competition can cause end gaps in boundaries. The orientational competition can cause end cuts in boundaries. (b) The hypercomplex cells in the second competitive stage input to cooperative bipole grouping cells. ON cells (open red ellipses) at each position excite like-oriented bipole cells at that position, whereas OFF cells (solid red ellipses) at each position inhibit like-oriented bipole cells at that position. Bipole cells use long-range oriented interactions to complete boundaries across parts of an image that are occluded, including the blind spot, retinal veins, and various incomplete representations of visual images and scenes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

versely, thereby creating a hole in the blue–white boundary; that is, an *end gap*.

The boundary cells at the processing level where the end gap forms are tonically active and inhibit other boundary cells that are tuned to different orientations at the same position (Fig. 4(a)); that is, by an *orientational competition*. In the absence of external inputs, the tonic activity of these cells is held in check by their mutual competition. When blue–white boundary cells are inhibited, the competitive balance is upset, causing cells that are tuned to other orientations, notably the perpendicular orientation, to be disinhibited and to thereby create an extra boundary segment that is called an *end cut*.

In summary, end gaps and end cuts are formed as a result of two successive stages of spatial and orientational competition between contrast-sensitive and orientationally tuned boundary cells (Grossberg, 1984a; Grossberg & Mingolla, 1985).

6.2. Simple cells and hierarchical resolution of uncertainty

Why does the brain compute end cuts and end gaps? Surely not just to cause a visual illusion like neon color spreading! In fact, spatial and orientational competition are needed to achieve effective vision, for the following reasons.

An early stage of boundary processing uses contrast-sensitive cells with orientationally-tuned receptive fields to begin to estimate boundary orientations at each position. These orientationally-tuned *simple cells* (Hubel & Wiesel, 1968) can respond to an oriented distribution of contrasts in response to scenic lines, edges, textures, and shading, not just edges alone (Fig. 4(a)). If the brain did use specialized detectors like edge detectors, then it would require many different types of specialized detectors, followed by complicated subsequent processing to try to fuse together all their information. Such an endeavor would fail if only because, in many natural scenes, lines, edges, textures, and shading are all overlaid.

Instead, the brain begins with less precise oriented estimates using simple cells. Simple cells cannot, however, respond at the ends of sufficiently thin lines. Without additional boundary processing, gaps would exist in boundaries at line ends. Brightness and color could flow through these gaps via surface filling-in (see Section 6.5). Every scene that contains lines would thus be awash in spurious brightness and color.

Spatial and orientational competition prevent this catastrophe from occurring by closing boundary gaps at line ends using end cuts (Grossberg & Mingolla, 1985). These competitive stages thus

illustrate *hierarchical resolution of uncertainty*: They overcome the spatial uncertainty at line ends that is caused by using simple cell receptive fields. End cuts do not, however, prevent all spreading of brightness and color from occurring beyond the edges of a scene, as neon color spreading illustrates (Section 6.5). However, events like neon color spreading are rare and do not interfere with the brain's ability to effectively represent the visual world most of the time. Visual illusions like neon color spreading are also very useful probes of how the brain forms boundaries and surfaces by calling attention to properties of these processes during circumstances when they break down.

6.3. Recognizing without seeing: Complex cells imply that all boundaries are invisible

Before output signals from simple cells reach the two competitive stages, they overcome a different kind of uncertainty. Each simple cell can respond to either a light-dark or a dark-light oriented contrast within its receptive field, but not to both. If simple cells that are sensitive to just one contrast polarity, say light-dark, input to the competitive stages, then the brain would often create boundaries with big gaps in them. This would occur, for example, in response to objects that lie in front of textured backgrounds whose relative contrasts with respect to the background reverse along the object's perimeter (Fig. 5, bottom row). Brightnesses and colors could spread out of these boundary gaps as well, again obliterating critical scenic information. This perceptual catastrophe is averted by *complex cells* at which the outputs of like-oriented dark-light and light-dark simple cells are added at each position (Fig. 4(a); Hubel & Wiesel, 1968). Complex cells can then respond to contrasts of both polarities at every position along the bounding contour of an object in front of a textured background.

Pooling inputs from opposite contrast polarities at complex cells implies that boundaries cannot represent visual qualia. They cannot discriminate between dark-light and light-dark contrasts, or red-green and green-red contrasts, or blue-yellow and yellow-blue contrasts, because they pool together inputs from simple cells that are sensitive to all of these differences (Thorell, De Valois, & Albrecht, 1984) to form the best possible boundaries. In other words, boundaries are *insensitive to direction-of-contrast*. Although boundaries can vary in strength or distinctiveness as they receive inputs from variable numbers and strengths of inducers, they do not code for visible brightnesses or colors.

Said in a more striking way: *All boundaries are invisible*. Boundaries may be consciously recognized, even when they are invis-

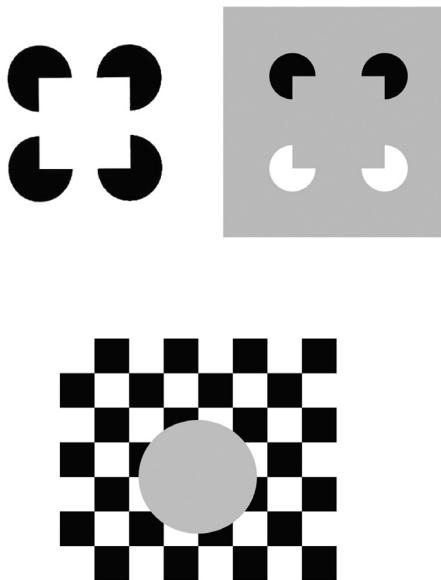


Fig. 5. The image in the upper row (left) is a Kanizsa square. The illusory square looks brighter than its background and in front of four partially occluded disks whose unoccluded parts have the shape of pac man figures. See Grossberg (2014) for an explanation of how the apparent brightness and depth of the emergent square covary. The image in the upper row (right) is a reverse-contrast Kanizsa square. The illusory square can be recognized, but many people do not see it because the filled-in gray colors inside and outside the square are approximately equal. This is due to the fact that there are two white pac men and two black pac men on a gray background. The white pac men cause darker feature contours within the illusory square, whereas the black pac men cause brighter feature contours within the illusory square. When these darker and brighter feature contours fill in within the square, they tend to cancel out. The same thing happens outside the square. The net effect is a similar gray color both inside and outside the square. The square thus can be recognized, but not seen. In the lower row, a gray circular disk is seen in front of a black-and-white checkerboard. The relative contrast along the disk boundary reverses as its perimeter is traversed. The text explains why this does not cause a percept in which gray color flows spuriously across the scene.

ble, as are the boundaries formed by the reverse-contrast Kanizsa square image in Fig. 5 (top row right).

If boundaries are invisible, then how do we see anything? *Consciously perceived qualia are predicted to be surface percepts*. Such surface percepts include the different surface brightnesses or colors that may occur on two sides of a boundary after surface filling-in occurs, and that thereby render it visible, as illustrated by the percept generated by a Kanizsa square stimulus (Fig. 5, top row left).

These results raise the basic question: Why do we consciously see? One possible answer is that we see objects to recognize them. However, the reverse-contrast Kanizsa square percept contradicts this hypothesis by showing that invisible boundaries can be consciously recognized (Fig. 5, top row right). In fact, there are hundreds of examples of recognition without seeing. Why, then, has evolution gone to the trouble of letting us see visual qualia? Our proposed answer to this question will be given in Sections 10.2 and 10.3 in terms of the surface percepts that are formed in V4 before triggering surface-shroud resonances and, from them, commands that can be used to control effective looking and reaching. Additional information about how boundary and surface processing support invariant object category learning and recognition are needed to understand how these events are proposed to happen, and will be described in Section 12.

6.4. Boundary completion and bipole grouping cells

The complex cells input to the two competitive stages (Fig. 4(a)). The spatial competition stage is sensitive to the length of lines, a

sensitivity that helps to create end gaps and end cuts. Its cells are often called *hypercomplex cells* (Hubel & Wiesel, 1968). However, many boundaries would still remain incomplete if boundary processing stopped with hypercomplex cells. For example, the two Kanizsa squares in Fig. 5 (top row) would just be seen and recognized as four pac-man figures.

Here one might reasonably ask: Why bother completing boundaries, indeed *illusory* boundaries, between pairs of collinear pac-man edges?

There are several important functional reasons for doing this. One reason is that the retina has large holes in it where light is not processed by photoreceptive cells. These holes include the blind spot and retinal veins, which prevent the processing of connected objects that are registered by the retina at their positions. Boundary completion generates boundaries of these objects across the blind spot and retinal veins. This same boundary completion process also generates the boundaries of Kanizsa squares.

Boundary completion cells cooperate across space with other boundary cells to complete a boundary between them whenever these cells are near enough to one another, are aligned across space in an approximately collinear arrangement, and have orientational tuning that is also approximately aligned (Fig. 4(b)). These boundary completion cells are often called *bipole cells* (Cohen & Grossberg, 1984; Grossberg, 1984a; Grossberg & Mingolla, 1985) because they complete boundaries *inwardly* in an *oriented* manner between pairs (bipoles!) of boundary inducers. Predicted bipole grouping properties (Fig. 6, top row, left column) have been supported by psychophysical data (e.g., Field, Hayes, & Hess, 1993; Kellman & Shipley, 1991), and neurophysiological data from cells in cortical area V2 (e.g., Peterhans & von der Heydt, 1989; von der Heydt, Peterhans, & Baumgartner, 1984), among others. Variants of bipole properties have also been used by other authors to model boundary grouping (Fig. 6).

When a bipole grouping starts to form, it is often initially fuzzy across space (Fig. 7, left column). If perfect alignment of inducers were required before grouping could start, then there would be a vanishingly small probability that boundary completion could begin. Instead, bipole cell receptive fields are coarse enough (Fig. 6) to enable multiple nearly collinear and nearly orientationally aligned inducers to start the grouping process. This coarseness can be understood as the embodiment within bipole receptive fields of perceptual experiences with nearly collinear and aligned visual stimuli during cortical development (Grossberg & Swaminathan, 2004; Grossberg & Williamson, 2001). However, if all perceptual groupings remained fuzzy, visual perception would be significantly degraded. In fact, most groupings become spatially sharp (Fig. 7, right column). Fig. 8 illustrates some percepts wherein locally preferred perpendicular end cuts may (Fig. 8, top row), or may not (Fig. 8, bottom row), determine the final grouping.

Given that groupings start out fuzzy in order to enable grouping to begin, why are so many completed boundaries sharp? Feedback within the boundary system can rapidly choose a final grouping that is maximally consistent with the spatial organization of the positional and orientational evidence in all of its inducers, as illustrated by Fig. 8.

6.5. Neon color spreading: Boundary-surface interactions and hierarchical resolution of uncertainty

Given this background, we can now better understand how neon color spreading (Fig. 3) occurs when basic properties of boundary completion and surface filling-in interact together, thereby providing a vivid example of Complementary Computing. Completed boundaries input topographically to surface representations where they are both generators of, and barriers to, surface filling-in (Grossberg, 1994, 1997; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000). These boundary-to-surface signals

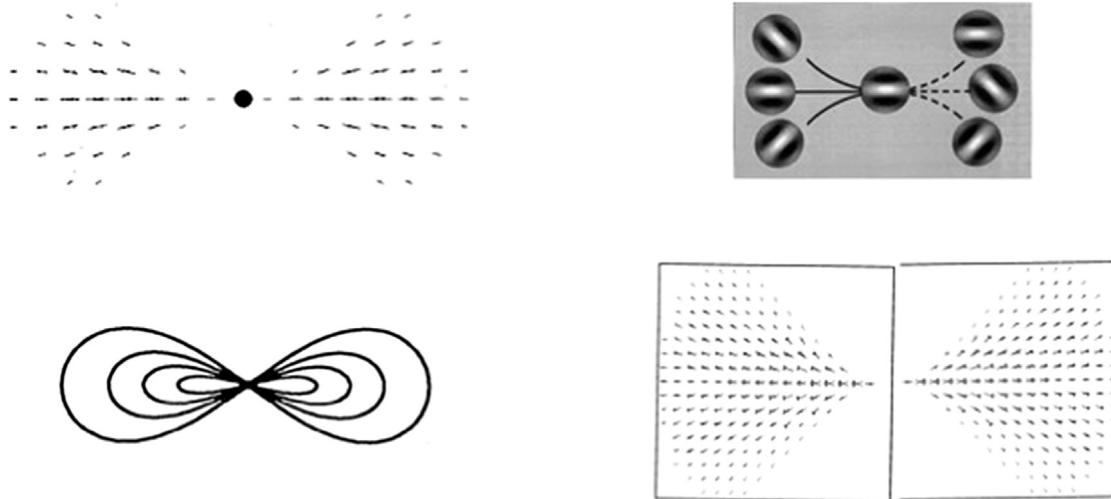


Fig. 6. The upper row (left) image shows the receptive field of the bipole cell that was simulated by Grossberg and Mingolla (1985). The black dot represents the position of the bipole cell body. The other markings represent the receptive field of the cell. The length and orientation of each line represents the relative strength of the bipole cell receptive field at that position and with that orientational preference. Colinear, like-oriented inputs are favored, but nearby orientations and positions can also influence bipole cell firing. The upper row (right) image depicts the “association field” that was described by Field et al. (1993) as a result of their psychophysical experiments. As in the bipole cell receptive field, collinear inputs are favored, but nearby orientations and positions can also have an effect. The bottom row (left) shows the version of bipole grouping modeled by Heitger and von der Heydt (1993), and the bottom row (right) shows the version modeled by Williams and Jacobs (1997).

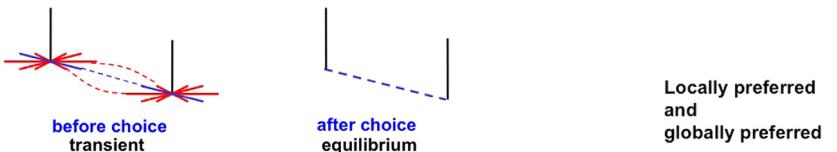


Fig. 7. (Left panel) The bipole cell receptive field enables multiple nearby orientations and positions to initiate grouping. (Right panel) Despite this initially coarse grouping, the final grouping is often sharp in both its positional and orientational selectivity..

are predicted to occur from boundary representations within the interstripes of cortical area V2 to surface representations within Filling-In Domains, or FIDOs, of the thin stripes of cortical area V2 (Fig. 9). Each FIDO also receives bottom-up topographic brightness or color signals. For example, when blue color inputs in response to Fig. 3 activate the corresponding FIDO, blue color can spread outward in an unoriented manner across this FIDO. In particular, blue color can flow out of the end gaps in the broken boundaries where the black and blue lines touch. The filling-in of blue color across space continues until the color hits the square illusory boundary, which prevents its further spread.

In addition to its outward and unoriented spread, surface filling-in is also *sensitive to direct-of-contrast*, because we can consciously see its effects. Neon color spreading hereby illustrates three pairs of computationally complementary properties of boundary completion and surface filling-in (Fig. 2): oriented vs. unoriented; inward vs. outward; insensitive to direction-of-contrast vs. sensitive to direction-of-contrast. A good boundary completion process thus cannot also be a good surface filling-in process, and conversely. However, without interactions between both processes that overcome their complementary deficiencies, vision would not be possible.

Neon color spreading is also the effect of three hierarchical resolutions of uncertainty acting together. It is only after all of these uncertainties are resolved that the brain can generate complete and stable enough visual representations of the world with which to control adaptive behaviors, and to mark these representations with conscious awareness. Our exposition will continue to explain how and why such representations are proposed to occur in cortical area V4.

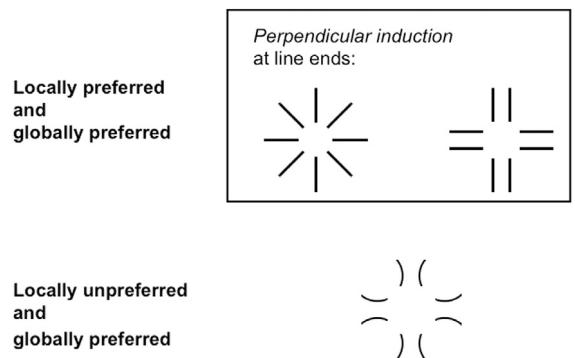


Fig. 8. (Upper row) In some groupings, the locally preferred orientations, in this case perpendicular end cuts to the inducing lines, group into an emergent global grouping. (Lower row) In other groupings, the locally preferred orientations are suppressed in the emergent global grouping. These examples illustrate that both kinds of grouping begin with a fuzzy band of grouping orientations from which the cooperative-competitive dynamics of the bipole grouping process choose the cells whose positions and orientations have the most support from their inducers.

To summarize what has already been described: The first hierarchical resolution of uncertainty uses spatial and orientational competition to overcome boundary-computing uncertainties caused by simple cells. The second hierarchical resolution of uncertainty uses bipole grouping feedback to generate positionally sharp groupings that overcome uncertainties in initial positionally fuzzy groupings. The third hierarchical resolution of uncertainty concerns why surface filling-in occurs. This final hierarchical resolution will now be described.

The surface system “discounts the illuminant”, or compensates for variable illumination, at an early processing stage. If this did not happen, then the brain could erroneously process changes in illumination as changes in perceived object shapes and colors. If object shapes seemed to plastically deform with every illumination change, then the brain could not learn to recognize a stable object percept. The discounting process inhibits luminance and color inputs at many positions across a scene’s surface. The process spares signals, called *feature contours*, near positions where color or luminance changes sufficiently rapidly across space. These are typically positions where material properties of scenic objects change, not just their illumination (Fig. 10, left column, top row). At

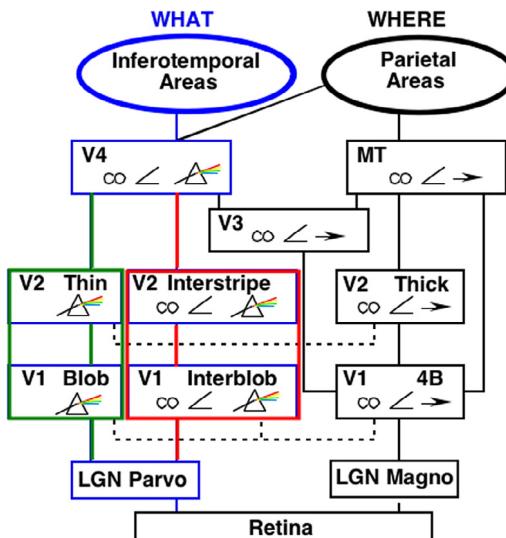


Fig. 9. Simplified schematic of the anatomy of three processing streams in the visual cortex. The LGN-blob-(thin stripe)-V4 stream (in green) fills-in visual surfaces, whereas the LGN-interblob-interstripe-V4 stream (in red) completes visual boundaries. LGN = Lateral Geniculate Nucleus; V1 = striate visual cortex; V2, V3, V4, MT = prestriate cortical areas. The motion stream goes through V1 and MT to the parietal areas. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
Source: Adapted with permission from DeYoe and Van Essen (1988).

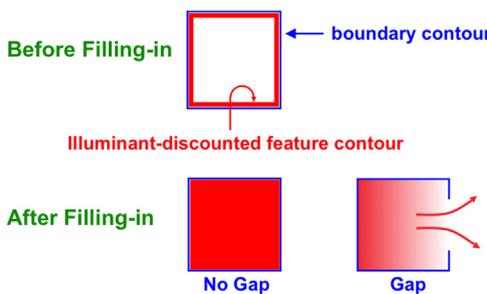


Fig. 10. (top row) A closed boundary contour (in blue) surrounds a pattern of illuminant-discounted feature contour activities (in red) before filling-in occurs. (bottom row, left column) After filling-in within the closed boundary, the filled-in activity fills the rectangular closed boundary, thereby generating large contrasts at the bounding positions of the surface. These contrasts will generate surface contour output signals to the next processing stage. (bottom row, right column) If there is a large hole, or gap, in a boundary, then color can flow out of it and equalize the filled-in surface activity on both sides of the boundary. When this occurs, surface contour signals cannot be generated at these boundary positions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

these positions, feature contours can compute material properties of a surface, such as its reflectances, or the proportions of reflected light within each wavelength. These reflectances are insensitive to illumination changes. At a later FIDO processing stage, surface filling-in spreads feature contour brightnesses and colors within closed boundary contours to reconstruct a surface representation of the scene in which the illuminant is significantly discounted (Fig. 10, left column, bottom row). The computation of illuminant-discounted feature contours, followed by surface filling-in of these feature contours within closed boundaries, is the third hierarchical resolution of uncertainty.

6.6. Why resonance?

The three pairs of complementary computational properties of boundaries and surfaces, along with the three hierarchical resolutions of uncertainty, illustrate that there is a great deal

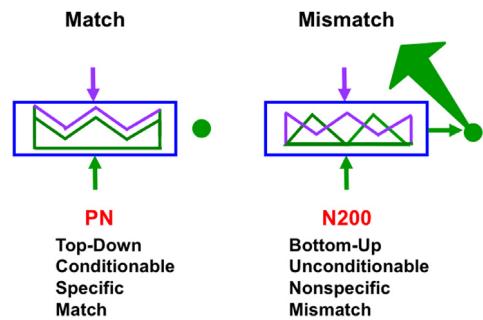


Fig. 11. Complementary properties of the attentional and orienting systems. (left panel) The read-out of an expectation in the attentional system is a top-down, conditionable, and specific signal that activates its target cells when there is a match with the bottom-up input. These are properties of the Processing Negativity, or PN, event-related potential. (right panel) When there is a mismatch, a bottom-up, unconditionable, and nonspecific arousal signal resets the currently active category and initiates a memory search for a better-matching category, or the learning of a new recognition category with which to represent truly novel information. These are properties of the N200 event-related potential.

of uncertainty in the early stages of visual processing by the brain. Only after all three hierarchical resolutions of uncertainty are complete, and after boundaries are completed and surfaces filled-in, has the brain constructed a contextually informative and temporally stable enough representation of scenic objects on which to base adaptive behaviors.

If this is indeed the case, then why do not the earlier stages undermine behavior? The proposed answer is that *brain resonance, and with it conscious awareness, is triggered at the processing stage that represents boundary and surface representations, after they are complete and stable enough to control visually-based behaviors like attentive looking and reaching*. Available experimental evidence suggests that this processing stage occurs in prestriate visual cortical area V4 (see Section 10.2). The fact that invisible boundaries can be recognized and that all conscious qualia are surface percepts suggests, moreover, that boundaries and surfaces each contribute to resonances that support recognition and seeing. These resonances link visual cortex with inferotemporal (IT) cortex and posterior parietal cortex (PPC). In particular, visual resonances focus attention upon these predictive boundary and surface representations while gain-amplifying and synchronizing them, without destroying their representational patterns.

To say more about why a surface-shroud resonance is proposed to be triggered at cortical area V4, and what a surface-shroud resonance is, requires the following summary of ART mechanisms.

7. An example of complementary computing: Attention vs. orienting during category learning

7.1. Object attention by a top-down, modulatory on-center, off-surround network

Complementary computing also occurs within ART circuits. Indeed, the process of attentive matching by top-down expectations, is computationally complementary to the process of memory search or hypothesis testing in response to mismatched inputs (Fig. 11). Attentive matching obeys the ART Matching Rule using top-down, modulatory on-center, off-surround circuits that read out the learned expectations to be matched against bottom-up input patterns (Fig. 12). This matching process has the following properties:

When a bottom-up input pattern is received at a processing stage, it can activate its target cells, if nothing else is happening. When a top-down expectation pattern is received at this stage, it can provide excitatory modulatory, or priming, signals to cells in its on-center, and driving inhibitory signals to cells in its off-surround,

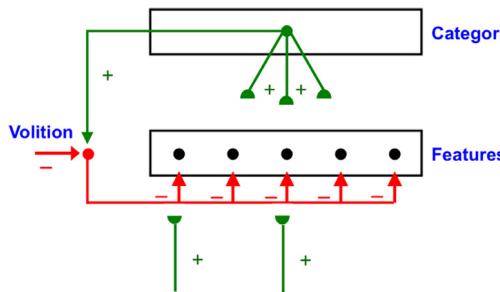


Fig. 12. ART Matching Rule. Bottom-up inputs can activate their target featural cells, other things being equal. A top-down expectation, by itself, can only modulate, prime, or sensitize cells in its excitatory on-center (green pathways with hemicircular adaptive synapses) because of the wider off-surround (red pathways) that tends to balance the top-down excitation (“one-against-one”) within the on-center, while causing driving inhibition in the off-surround. When bottom-up inputs and a top-down expectation are both active, only cells where bottom-up excitation and the top-down excitatory prime converge in the on-center can fire (“two-against-one”), while other featural cells are inhibited. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

if nothing else is happening. The on-center is modulatory because the off-surround also inhibits the on-center cells (Fig. 12), and these excitatory and inhibitory inputs are approximately balanced (“one-against-one”). When a bottom-up input pattern and a top-down expectation are both active, cells that receive both bottom-up excitatory inputs and top-down excitatory priming signals can fire (“two-against-one”), while other cells are inhibited. In this way, only cells can fire whose features are “expected” by the top-down expectation, and an attentional focus starts to form at these cells.

A top-down expectation is thus a *top-down, conditionable, and specific* event that activates its target cells during a *match* (Fig. 11, left panel). “Conditionable” means that the top-down pathways contain adaptive weights that can learn, or be conditioned, to encode a prototype of the recognition category that activates it. “Specific” means that each top-down expectation reads out its learned prototype pattern.

7.2. ART hypothesis testing and learning of a predictive recognition category

The ART hypothesis testing and learning cycle (see Fig. 13) explains why the above properties of attentional matching are complementary to those of orienting search: Suppose that an active recognition category reads out a top-down learned expectation that obeys the ART Matching Rule (Fig. 12). When a good enough match occurs between a bottom-up input pattern and a learned top-down expectation, the initially attended features can reactivate the category via the bottom-up adaptive filter, and the activated category can reactivate the attended features via its top-down expectation (Fig. 1). This self-reinforcing excitatory feedback cycle leads to a sustained *feature-category resonance* (Table 2), which can support attentional focusing and conscious recognition. Resonance can trigger fast learning of the attended, or matched, feature pattern, hence the name *adaptive resonance*. One psychophysiological marker of such a resonant match is the processing negativity, or PN, event-related potential (Fig. 11; Banquet & Grossberg, 1987; Grossberg, 1978a, 1984b; Näätänen, 1982).

For a top-down expectation of a new recognition category to match the feature pattern that activates it, all its top-down adaptive weights initially have large values, so that it can match *any* feature pattern. These adaptive weights are *pruned* as learning proceeds, leading to learning of an attended critical feature pattern. It has been mathematically proved that such *match learning* can solve the stability-plasticity dilemma by creating stable categories

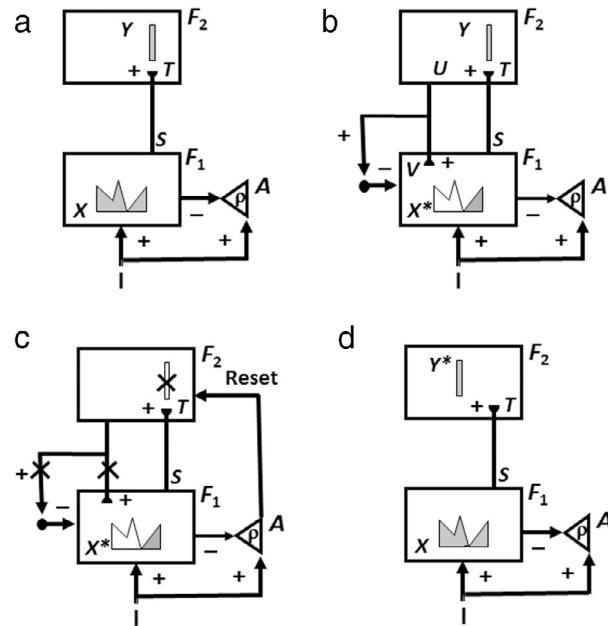


Fig. 13. How ART searches for and learns a new recognition category using cycles of match-induced resonance and mismatch-induced reset. (a) Input pattern I is instated across feature detectors at level F_1 as an activity pattern X , at the same time that it generates excitatory signals to the orienting system A with a gain ρ that is called the *vigilance* parameter. Activity pattern X generates inhibitory signals to the orienting system A as it generates a bottom-up input pattern S to the category level F_2 . A dynamic balance within A between excitatory inputs from I and inhibitory inputs from S keeps S quiet. The bottom-up signals in S are multiplied by learned adaptive weights to form the input pattern T to F_2 . The inputs T are contrast-enhanced and normalized within F_2 by recurrent on-center off-surround signals that obey the membrane equations of neurophysiology, otherwise called shunting interactions. This competition leads to selection and activation of a small number of cells within F_2 that receive the largest inputs. The chosen cells represent the category Y that codes for the feature pattern at F_1 . In this figure, a winner-take-all category is shown. (b) The category activity Y generates top-down signals U that are multiplied by adaptive weights to form a prototype, or critical feature pattern, V that encodes the expectation that the active F_2 category has learned for what feature pattern to expect at F_1 . This top-down expectation input V delivers an excitatory modulatory signal at F_1 cells in its on-center, at the same time that it inhibits F_1 cells in its off-surround (ART Matching Rule). If V mismatches I at F_1 , then a new STM activity pattern X^* (the gray pattern) is selected at cells where the patterns match well enough. In other words, X^* is active at I features that are confirmed by V . Mismatched features (white area) are inhibited. When X changes to X^* , total inhibition decreases from F_1 to A . (c) If inhibition decreases sufficiently, A releases a nonspecific arousal burst to F_2 ; that is, “novel events are arousing”. The vigilance parameter ρ determines how bad a match will be tolerated before a burst of nonspecific arousal is triggered. This arousal burst triggers a memory search for a better-matching category, as follows: Arousal resets F_2 by inhibiting Y . (d) After Y is inhibited, X is reinstated and Y stays inhibited as X activates a different category, that is represented by a different activity pattern Y^* , at F_2 . Search continues until a better matching, or novel, category is selected. When search ends, an attentive resonance triggers learning of the attended data in adaptive weights within both the bottom-up and top-down pathways. As learning stabilizes, inputs I can activate their globally best-matching categories directly through the adaptive filter, without activating the orienting system.

Source: Adapted with permission from Carpenter and Grossberg (1993).

in response to arbitrary sequences of events presented in any order; e.g., Carpenter and Grossberg (1987, 1991). Category learning goes on within the *attentional system* of ART which, for visual categories, includes brain regions such as prestriate visual cortex, inferotemporal cortex, and prefrontal cortex (Table 3, row 3).

If learning can occur only if there is a sufficiently good match, then how is anything new ever learned? Here is where complementary properties of attentional matching and orienting search are crucial: A sufficiently bad mismatch between an active top-down expectation and a bottom-up input, say because the input is unfamiliar, can drive a memory search. Such a mismatch within the attentional system activates the complementary

orienting system, which is sensitive to unexpected and unfamiliar events (Fig. 13(c)). The ART orienting system includes the nonspecific thalamus and the hippocampal system. Carpenter and Grossberg (1993) and Grossberg and Versace (2008) summarize supportive data.

More recently, Brincat and Miller (2015) have reported neurophysiological data that support the distinction between category learning within the attentional system that includes prefrontal cortex, and the orienting system that includes the hippocampus. They collected data from prefrontal cortex (PFC) and hippocampus (HPC) in monkeys learning object-pair associations (p. 576): “PFC spiking activity reflected learning in parallel with behavioral performance, while HPC neurons reflected feedback about whether trial-and-error guesses were correct or incorrect...Rapid object associative learning may occur in PFC, while HPC may guide neocortical plasticity by signaling success or failure via oscillatory synchrony in different frequency bands”. These results contribute to a long history of experiments that have implicated the hippocampus in mismatch processing, including the processing of novel events (e.g., Deadwyler, West, & Lynch, 1979; Deadwyler, West, & Robinson, 1981; Otto & Eichenbaum, 1992; Solokov, 1968; Vinogradova, 1975).

7.3. Complementary PN and N200 ERPs during attention and memory search

In contrast to the *top-down, conditionable, specific, and match* properties that occur during an attentive match, an orienting system mismatch is *bottom-up, unconditionable, nonspecific, and mismatch* (Fig. 11, right panel): A mismatch occurs when *bottom-up* activation of the orienting system cannot be adequately inhibited by the *bottom-up* inhibition from the matched pattern (Fig. 13(c)). These signals to the orienting system are *unconditionable*, or not subject to learning. Mismatch-activated output from the orienting system *nonspecifically* arouses all the category cells because the orienting system cannot determine which categories read out the expectation that led to mismatch. Any category may be responsible, and may thus need to be reset by arousal. Finally, the orienting system is activated by a sufficiently big *mismatch*. These are properties of the N200 event-related potential, or ERP (Fig. 11; Näätänen, Simpson, & Loveless, 1982; Sams, Paavilainen, Alho, & an, 1985). More generally, during an ART memory search, sequences of mismatch, arousal, and reset events occur that exhibit properties of P120, N200, and P300 ERPs, respectively (Banquet & Grossberg, 1987).

As Fig. 11 summarizes, these four sets of properties of the attentional system are complementary to those of the orienting system (*top-down vs. bottom-up, conditionable vs. unconditionable, specific vs. nonspecific, match vs. mismatch*), with the PN and N200 ERPs illustrating these complementary properties. The orienting system can detect that an error has occurred, but does know what category prediction caused it. The attentional system knows what categories are active, but not if these categories adequately represent current inputs. By interacting, these systems can determine what the error is and discover a new category to correct it.

7.4. Combining paradigms to disclose finer details about category learning

There are now multiple neurobiological experimental markers of how category learning occurs that, when combined, can lead to more mechanistically targeted experiments. Among others, Banquet and Grossberg (1987) provide ERP markers during memory search; Brincat and Miller (2015) provide oscillatory neurophysiological markers of the interplay between prefrontal cortex and hippocampus during learning and mismatch; Otto and Eichenbaum (1992) provide neurophysiological data during hippocampal

mismatch processing; and Spitzer, Desimone, and Moran (1988) provide neurophysiological data from cortical area V4 during the learning of easy vs. difficult discriminations, a process that is regulated within ART by a vigilance parameter (e.g., Fig. 13; Carpenter & Grossberg, 1987; Grossberg, 2013a, 2013b).

Spitzer et al. (1988) report “in the difficult condition, the animals adopted a stricter internal criterion for discriminating matching from non-matching stimuli...The animal's internal representations of the stimuli were better separated...increased effort appeared to cause enhancement of the responses and sharpened selectivity for attended stimuli...”, all properties of ART vigilance control, where higher vigilance is needed to make more difficult discriminations.

Combining these various paradigms suggests experimental tests of cortico-hippocampal interactions while discrimination difficulty is varied. More difficult discriminations, at least under proper circumstances, should lead to higher vigilance, more mismatch events, and thus more of the hippocampal novelty responses found by Brincat and Miller (2015) and Otto and Eichenbaum (1992). During these mismatch events, ERPs should progress in the temporal order that was described by Banquet and Grossberg (1987).

Grossberg and Versace (2008), Palma, Grossberg, and Versace (2012), and Palma, Versace, Grossbergand (2012) have furthermore proposed how mismatch-activated acetylcholine release may regulate vigilance in laminar neocortical circuits that are described by spiking neurons during category learning. The anatomical and pharmacological markers of vigilance control that they identify can add additional experimental manipulations to test the neural mechanisms of category learning when both easy and difficult discriminations are employed.

A related set of experiments concerns measuring more carefully what happens during both attentive recognition vs. mismatch reset in response to sequences of familiar vs. unfamiliar cues. Here, the following surprising discovery in Grossberg and Versace (2008) may provide a useful marker. Grossberg and Versace (2008) showed that, in their laminar cortical model of category learning with spiking neurons, a good enough top-down attentive *match* with a *bottom-up* feature pattern yields faster gamma oscillations, whereas a big enough *mismatch* yields slower beta oscillations. Both types of oscillations are emergent properties of network dynamics. Grossberg (2013a, Section 38) reviews the fact that this type of match-mismatch gamma-beta dichotomy has been reported in neurophysiological experiments about differential effects of attention on different layers in visual cortex (Buffalo, Fries, Landman, Buschman, & Desimone, 2011), spatial attention shifts in frontal eye fields (Buschman & Miller, 2009), and novel place cell learning in hippocampus (Berke, Hetrick, Breck, & Green, 2008).

Four implications of the prediction of gamma-beta dichotomy during match-mismatch dynamics are: they should occur in bursts; these bursts should alternate through time; the gamma bursts should be the ones that encode “information”; and the gamma bursts should exhibit higher spiking rates because they embody a resonant match state. All of these properties have been reported in experiments on working memory whose recordings were taken from lateral prefrontal cortex and the frontal eye fields (Lundqvist et al., 2016). The authors interpret their results in terms of working memory, and discuss the gamma oscillations in terms of “encoding/decoding events” and “re-activation of sensory information”, whereas beta oscillations are viewed as a “default state interrupted by encoding and decoding” (p. 152). This “default” interpretation does not explain why beta bursts are “interrupted by encoding and decoding” or why the gamma oscillations embody “encoding/decoding events”. The ART hypothesis in terms of dichotomous states of attentive match and mismatch reset does propose an explanation of this property.

In addition, the Lundqvist et al. (2016) article describes modeling ideas in which there is no temporal order represented in working memory, although temporal order information is essential for proper functioning of a working memory. It may be useful to test if and how these neurophysiological data fit into a scheme of resonantly categorizing sequences of item chunks before they are stored in a working memory that encodes temporal order information, as described in Section 16.

7.5. Resonance between distributed features and compressed categories solves symbol grounding

Another kind of complementary ignorance is overcome through resonance within the attentional system, and thereby solves the *symbol grounding problem* (Harnad, 1990): When a feature pattern is activated at level F_1 (Fig. 13(a)), its individual features have no meaning. They become meaningful only as part of the spatial pattern of features to which they belong. A recognition category at level F_2 can selectively respond to this feature pattern, but does not know what these features are. When levels F_1 and F_2 resonate due to mutual positive feedback, the attended features at F_1 are coherently bound together by the top-down expectation that is read out by the active category at F_2 (Fig. 1). The attended pattern of features at F_1 , in turn, maintains activation of the recognition category at F_2 via the bottom-up adaptive filter. The resulting *feature-category resonance* hereby becomes a *synchronous bound state* that carries information about both what category is active and what features are represented by it, and thus embodies sufficient information to support conscious recognition of a visual object or scene (Table 2).

7.6. Feature-category resonances support object attention, learning, and recognition

The ART search and learning cycle that is summarized in Fig. 13 clarifies how view-specific and position-specific categories are learned. This process is necessary to understand how surface-shroud resonances occur, but it is not sufficient to do so. It provides a foundation for understanding how view-, position-, and size-invariant object categories are learned and recognized (Section 12). The analysis of invariant object category learning and recognition leads directly to the concept of surface-shroud resonance.

Before turning to this analysis, it should be noted that evidence in support of the ART category learning and search cycle that is summarized in Fig. 13 has been supported by modeling studies of many different kinds of data. These studies have explained, and successfully predicted, data about such varied topics as texture segregation (Bhatt et al., 2007); visual scene recognition (Grossberg & Huang, 2009); learning of view-, position-, and size-invariant inferotemporal recognition categories, and recoding of inferotemporal categories during target swapping (Cao, Grossberg, & Markowitz, 2011); gamma and beta oscillations during top-down matches and mismatches, respectively (Grossberg, 2013a; Grossberg & Versace, 2008); learning of concrete or abstract recognition categories under variable vigilance (Amis, Carpenter, Ersøy, & Grossberg, 2009; Grossberg, Markowitz, & Cao, 2011); breakdown of vigilance control during mental disorders such as autism (Grossberg & Seidman, 2006); and fast visual attention switching during the recognition of letter or number targets during rapid serial visual presentations (Grossberg & Stone, 1986a).

As will be seen in Section 16, similar model mechanisms, specialized for learning about sequences of events, support explanations of data about contextually-cued search of visual scenes (Huang & Grossberg, 2010), recognition and recall of visually presented words (Grossberg & Stone, 1986b); and recognition and recall of auditorily presented words (Boardman, Grossberg, Myers, &

Cohen, 1999; Grossberg, Boardman, & Cohen, 1997; Grossberg & Kazerounian, 2011, 2016; Grossberg & Myers, 2000; Kazerounian & Grossberg, 2014).

The mathematical law for how top-down object attention obeys the ART Matching Rule (Fig. 12) in these modeling studies (e.g., Bhatt et al., 2007) is also supported by modeling studies by other authors, notably Reynolds and Heeger (2009) in their “normalization model of attention”.

These view-specific and position-specific ART object categories are proposed to occur in posterior inferotemporal cortex (ITp). The additional machinery of the ARTSCAN model (see Section 12) is needed to show how learning can link several such ART categories in ITp to an invariant object recognition category in anterior inferotemporal cortex (ITa) while eye movements freely scan a scene. The ART dynamics in Fig. 13 also consider only one form of object attention (Posner, 1980), the *prototype attention* within the What cortical stream that focuses attention upon the critical feature pattern of a view- and position-specific category prototype (Fig. 1). In order to explain how invariant category learning occurs with the help of surface-shroud resonances, ARTSCAN proposes how object attention interacts with *spatial attention* (Duncan, 1984) in the Where/How cortical stream. This kind of spatial attention is called *surface attention* to distinguish it from *boundary attention*, which is another kind of spatial attention. Each type of attention supports conscious perception or recognition of different aspects of experience. See Sections 10 and 12.

8. Linking brain to behavior: All conscious states are resonant states, but not conversely

The ART prediction that “all conscious states are resonant states” provides a linking hypothesis between brain mechanisms and psychological functions that enables our explanations of consciousness to be experimentally tested. It is the dynamical state that “lights up” to enable a conscious experience to occur, and not just be the “whir of information processing”. It is not, however, predicted that “all resonant states are conscious states”. Resonant states that are not accessible to consciousness include parietal–prefrontal resonances that trigger the selective opening of basal ganglia gates to enable the read-out of contextually-appropriate thoughts and actions (see Section 18.2; Brown, Bullock, & Grossberg, 2004; Buschman & Miller, 2007; Grossberg, 2016b) and entorhinal–hippocampal resonances that dynamically stabilize the learning of entorhinal grid cells and hippocampal place cells during spatial navigation (see Section 19.8; Grossberg & Pilly, 2014; Kentros, Agniotri, Streater, Hawkins, & Kandel, 2004; Morris & Frey, 1997; Pilly & Grossberg, 2012). These resonances do not include feature detectors that are activated by external senses, such as those that support vision or audition, or internal senses, such as those that support emotion. Because of their inaccessibility to consciousness, these resonances also do not have familiar words, like seeing, hearing, and feeling, with which to describe them using daily language.

9. The varieties of brain resonances and the conscious experiences that they support

ART predicts what resonances support vision, hearing, and feeling, and how resonances supporting knowing are synchronously linked to these qualia-supporting resonances (Table 2). For example, *surface-shroud resonances* are predicted to support conscious percepts of visual qualia. *Feature-category resonances* are predicted to support conscious recognition of visual objects and scenes. Both kinds of resonances may synchronize during conscious seeing and recognition. *Stream-shroud resonances* are predicted to support conscious percepts of auditory qualia. *Spectral-pitch-and-timbre resonances* are predicted to support conscious recognition

of sources in auditory streams. Stream-shroud and spectral-pitch-and-timbre resonances may synchronize during conscious hearing and recognition of auditory streams. *Item-list resonances* are predicted to support recognition of speech and language. They may synchronize with stream-shroud and spectral-pitch-and-timbre resonances during conscious hearing of speech and language, and build upon the selection of auditory sources by spectral-pitch-and-timbre resonances in order to recognize the acoustical signals that are grouped together within these streams. *Cognitive-emotional resonances* are predicted to support conscious percepts of feelings, as well as recognition of the source of these feelings. Cognitive-emotional resonances can also synchronize with resonances that support conscious qualia and knowledge about them. All of these resonances have distinct anatomical substrates.

It cannot be overemphasized that these resonances are not just correlates of consciousness. Rather, they embody the subjective properties of individual conscious experiences. They also enable effective behaviors to be elicited (Table 1). The following sections describe what these resonances are, indicate to the degree possible in the light of available data where they exist in the brain, and itemize some of their functional roles in the control of adaptive behaviors.

10. Surface-shroud resonances for conscious seeing of visual qualia

10.1. Surface-shroud resonances are generated between V4 and PPC

The previous descriptions of visual boundaries and surfaces set the stage for understanding surface-shroud resonances and their role in both invariant object category learning and in supporting the conscious awareness of visual experiences. Several perceptual and cognitive scientists have reported that spatial attention can fit itself to the shape of an attended object. The term *attentional shroud* for such a form-fitting distribution of spatial attention was introduced by Tyler and Kontsevich (1995).

An attentional shroud can form when bottom-up topographic signals from an active surface representation in V4 activate PPC cells, which then compete for spatial attention while also sending top-down topographic excitatory signals back to V4 (Fig. 14). Taken together, these signals form a *recurrent on-center off-surround* network that is capable of contrast-enhancing and normalizing its activities (Grossberg, 1973, 1980). This recurrent exchange of excitatory signals, combined with competitive inhibitory signals, helps to choose a winning focus of spatial attention in PPC that configures itself to the shape of the attended object surface in V4. The resulting *surface-shroud resonance* between a surface representation in V4 and spatial attention in PPC corresponds to paying sustained spatial attention to consciously visible surface qualia. Once triggered, a surface-shroud resonance can propagate top-down to earlier cortical areas such as V2 and V1, where it can attentively amplify and select consistent cell activations while inhibiting inconsistent ones, and bottom-up to higher cortical areas such as prefrontal cortex, or PFC.

10.2. V2 and V4: Recognizing occluded objects, seeing unoccluded surfaces, and transparency

Why is a surface-shroud resonance assumed to be triggered between V4 and PPC? The FACADE and 3D LAMINART models propose that cortical areas V2 and V4 resolve a basic design tension between recognition and seeing whose resolution prevents all occluding objects from looking transparent during 3D vision and figure-ground separation (Grossberg, 1994, 1997). Since this prediction was made, subsequent modeling articles have used these model neural mechanisms to successfully explain, simulate,

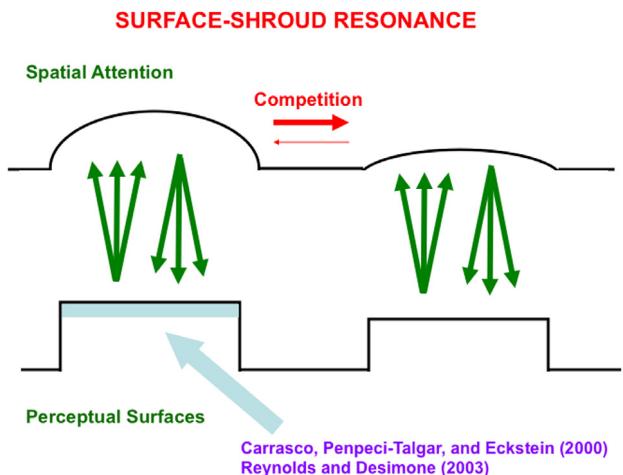


Fig. 14. A cross-section of a simple filled-in surface (e.g., in cortical area V4) is shown in which a more contrastive bar is to the left of a less contrastive bar. Each position in the surface sends topographic bottom-up excitatory signals to the spatial attention region (e.g., PPC). Each activated spatial attention cell sends topographic top-down excitatory signals to the surface, as well as broad off-surround inhibitory signals to other spatial attention cells, thereby activating a recurrent on-center off-surround network whose cells obey shunting laws. This recurrent network generates a *surface-shroud resonance* that contrast-enhances the more active spatial attention cells, while inhibiting the less active ones, thereby creating a form-sensitive distribution of spatial attention, or *attentional shroud*, that focuses spatial attention upon the more contrastive surface, while also increasing its effective contrast.



Fig. 15. (Left column) Three abutting rectangles cause a compelling percept of a vertical bar that partially occludes a horizontal bar. (Right column) The occluded region of the horizontal bar is amodally recognized without being consciously seen. If all such completed occluded regions could be seen, then all occluders would look transparent. Interactions between cortical areas V2 and V4 are predicted to prevent this from happening. See text for details about how this is proposed to happen.

and predict many interdisciplinary data about the visual percepts that occur during 3D vision and figure-ground perception (e.g., Cao & Grossberg, 2005, 2012; Fang & Grossberg, 2009; Grossberg, 1994, 1997; Grossberg, Kuhlmann, & Mingolla, 2007; Grossberg & McLoughlin, 1997; Grossberg & Swaminathan, 2004; Grossberg & Yazdanbakhsh, 2005; Kelly & Grossberg, 2000; Leveille, Versace, & Grossberg, 2010).

In particular, V4 is predicted to be the cortical region where figure-ground-separated 3D surface representations of the *unoccluded* regions of *opaque* object regions are completed, and thereupon support both seeing and recognition of these regions. These unoccluded object surface regions are the parts of a scene that are typically seen as we explore the world. The same neural mechanisms also explain how V4 also supports seeing of 3D *transparent* surfaces. Cortical area V2, in contrast, is proposed to complete object boundaries and surfaces of *occluded* object regions that may be amodally recognized, but not seen.

The percept that is generated by Fig. 15 (left column) illustrates this issue. Although this image is composed of three abutting rectangles, it generates a compelling percept of a horizontal bar that is partially occluded by a vertical bar. The partially occluded portion of the horizontal bar (Fig. 15, right column, dashed line) is amodally recognized, but it is not consciously seen. The model proposes how this vertical rectangle is separated from the partially-occluded horizontal rectangle, which can then be completed “behind” the horizontal rectangle in V2. Then direct

pathways from V2 to higher cortical areas such as inferotemporal (IT) cortex, and back, are used to recognize this completed perceptual representation as part of a feature-category resonance (Table 2), despite the fact that the occluded part of this rectangle is not consciously seen. Such recognition without seeing is said to be *amodal*.

If the completed boundary and surface behind the horizontal rectangle could also be seen, then the horizontal rectangle would look transparent, because both the horizontal and vertical rectangles could be seen at the same spatial positions, albeit at different depths. V2 and V4 are predicted to work together to ensure that not all objects look transparent. As a result of this teamwork between V2 and V4, partially occluded objects can be recognized without forcing all occluders to look transparent. The hypothesis that V4 represents 3D surfaces whose objects have been separated from one another in depth is consistent with several different types of neurobiological experiments (e.g., Chelazzi, Miller, Duncan, & Desimone, 2001; Desimone & Schein, 1987; Lueck et al., 1989; Ogawa & Komatsu, 2004; Reynolds, Pasternak, & Desimone, 2000; Schiller & Lee, 1991; Zeki, 1983).

A surface-shroud resonance is thus assumed to be triggered between V4 and PPC because V4 is predicted to be the earliest cortical stage after which the three hierarchical resolutions of uncertainty have already taken place, and boundaries have been completed and surfaces filled in, leading to figure-ground-separated 3D surface representations of unoccluded object regions. The surface-shroud resonance between V4 and PPC initiates the process whereby these surface representations may become consciously visible.

There are many informative examples of how V2 is proposed to complete the boundaries and surfaces of partially occluded objects while separating these objects in depth. Consider, for example, the Kanizsa stratification images in Fig. 16 (top row). The image in Fig. 16 (top row, left column) has ambiguous white regions where the cross and rectangle overlap, but these regions are completed to generate an unambiguous percept of either a cross in front of a partially occluded square, or a square in front of a partially occluded cross, with the cross perceived most of the time as these percepts bistably switch through time. The vertical lines that are drawn in Fig. 16 (top row, right column) force the corresponding vertical bars of the square to always be seen in front. The vertical bar of the cross nonetheless still appears in front most of the time, even though this creates an image of a cross that is bent in depth, thereby contradicting the perceptual interpretation that a Bayesian account would predict.

In the percepts that are induced by these stratification images, again the occluded parts of the objects can be recognized but not seen, even though these *same positions in space* can be seen as part of either occluding object when the percepts flip bistably in time. Kelly and Grossberg (2000) explain stratification percepts, including simulations of their conscious 3D surface percept properties. The same model mechanisms also explain key data about when transparent, rather than opaque, surface percepts are seen, including how percepts of unique transparency, bistable transparency, or no transparency at all, are generated in response to the images in Fig. 16 (bottom row, three leftmost images), which differ only in terms of the relative contrasts of their respective regions, since all of these regions have the same geometry (Fig. 16, bottom row, right column). Grossberg and Yazdanbakhsh (2005) provide model explanations and simulations of these transparency percepts, among others, including simulations of the consciously seen surface percepts, much as the model has simulated 3D surface percepts in response to many other stimuli, including stereogram images (Fang & Grossberg, 2009; see Fig. 23).

In addition to explaining many perceptual data about 3D vision and figure-ground perception, the FACADE and 3D LAMINART models have anticipated many neurophysiological data about

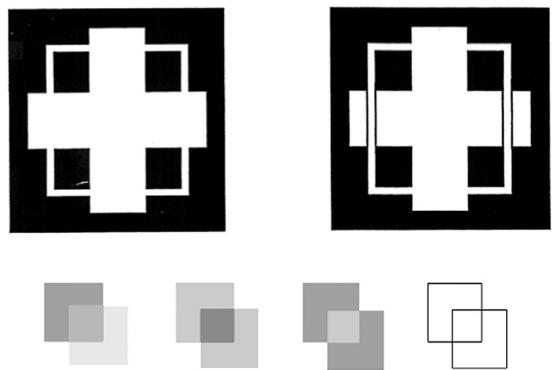


Fig. 16. The two figures in the top row illustrate Kanizsa stratification percepts. In the percept that is caused by the left figure, the white cross appears in front of the square border most of the time. The white in positions where the cross occludes the square appears to belong to the cross, and appears in front of the square, which is amodally completed behind it. On occasion, the percept flips with the square appearing in front of the cross. Then the white area that previously belonged to the cross appears to belong to the square, with the cross amodally completed behind it. In the percept that is caused by the right figure, even when the extra black vertical lines force the vertical square bar to always appear in front of the cross, the horizontal branches of the square are usually amodally recognized behind the vertical bars of the cross, leading to a percept of a cross and a square that are bent in depth. This latter result is incompatible with a Bayesian statistics account of what the percept should look like based upon the high probability of experiencing flat crosses and squares in the world. These percepts are explained in Grossberg (1997) and simulated in Kelly and Grossberg (2000). The three left-most figures in the bottom row generate, reading from left to right, a percept of unique transparency, bistable transparency, and no transparency, just by changing the relative contrasts of the different regions of the inducing figures, which share the same geometry, as shown in the rightmost figure of this row. These percepts are simulated in Grossberg and Yazdanbakhsh (2005) using the same kinds of neural circuits that explain the percepts in the first row.

figure-ground properties of V2 cells, and have recently explained (see Grossberg, 2016a) all the key data properties reported in a remarkable series of neurophysiological experiments on V2 cells from the von der Heydt lab (O'Herron & von der Heydt, 2009; Qiu, Sugihara, & von der Heydt, 2007; Qiu & von der Heydt, 2005; von der Heydt, Zhou, & Friedman, 2000; Zhang & von der Heydt, 2010; Zhou, Friedman, & von der Heydt, 2000). Thus the model hypothesis that V2 and V4 carry out different roles in recognition and seeing is supported by both psychological and neurobiological data.

10.3. Conscious seeing, orienting, and reaching

If occluding and partially occluded, objects can be recognized using boundary and surface representations that are completed in V2, then what functions are enabled by conscious 3D percepts of the unoccluded parts of opaque surfaces that are proposed to occur in V4? A synthesis of the FACADE and ARTSCAN models proposes that a surface-shroud resonance with V4 provides an *extra degree of freedom*, namely conscious visibility, with which to distinguish directly reachable surfaces from non-reachable ones. The need for such an extra degree of freedom seems to have arisen because of the urgent need during evolution for more powerful object recognition capabilities, notably the ability to complete the boundaries and surfaces of partially occluded objects behind their occluders so that they can be recognized. Animals who could not recognize such partially occluded objects would be at a severe survival disadvantage compared to those who could. Fortunately, basic properties of bipole cells for perceptual grouping, and simple feedback interactions between the boundary and surface streams, go a long way to accomplish figure-ground separation. (Grossberg, 1994, 2016a).

If the completed parts of these partially occluded objects could also be seen, then great confusion could occur in the planning of reaching behaviors, since *all occluders would look transparent*, and

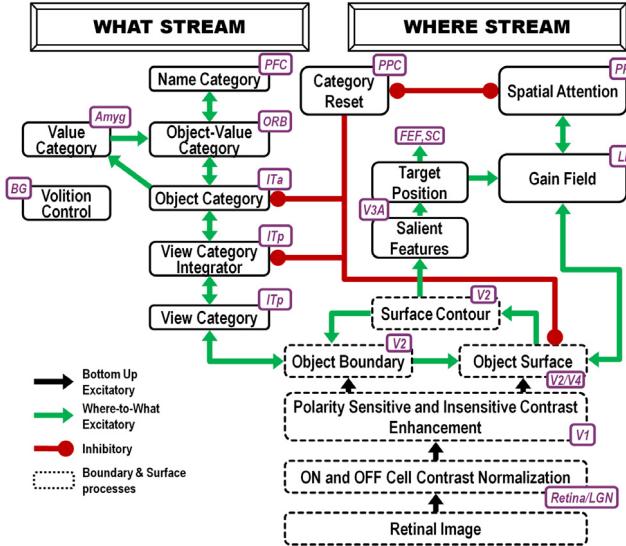


Fig. 17. ARTSCAN Search model macrocircuit (learning and consciousness). A surface-shroud resonance between an object surface representation within cortical area V4 and a form-fitting attentional shroud in a spatial attention region of posterior parietal cortex (PPC) is predicted to initiate a conscious visual percept of surface qualia. This circuit can learn invariant object categories in cortical area ITa using Where-to-What modulation of category learning. See text for details. Source: Reprinted with permission from Chang et al. (2014).

it would seem natural to reach directly through occluding objects to the occluded objects behind them. In brief, then, there is a *design tension during evolution between the requirements of recognition and reaching*. Conscious visibility enables the unoccluded parts of many surfaces to appear opaque, and thus good targets for reaching, without eliminating the ability of the visual cortex to correctly represent surfaces that are, in fact, transparent.

Thus, the triggering of a surface-shroud resonance between V4 and PPC not only selects surface representations that are sufficiently complete and stable to support adaptive behaviors, but also provides a surface visibility signal to mark the opaque surface regions, as opposed to the completed regions of occluded surfaces right behind them, to which orienting eye movements and reaching arm movements can be successfully directed.

Crick and Koch (1995) also proposed that visual awareness may be related to planning of voluntary movements, but without any analysis of how 3D vision occurs. They also made the strong claim that “What does enter awareness, we believe, is some form of the neural activity in certain higher visual areas, because they project directly to prefrontal areas” that help to plan movements. This claim captures part of the current claim, but also ignores the main issue of where the brain completes 3D surface representations that separate unoccluded surface regions from one another in depth. Only after this stage can these representations be brought into consciousness by a surface-shroud resonance, also not specified in Crick and Koch (1995) or in their subsequent analyses.

It needs also to be acknowledged that reaching behind occluders is possible, and does not require conscious seeing of occluded object features. However, this skill builds upon a great deal of experience with the circular reactions that support direct reaching behaviors during development (see Section 14). Further evidence for a role of consciously visible 3D surface representations in movement planning is discussed as part of the explanations of clinical data about visual neglect in Section 10.5.

10.4. A link between object learning, recognition, and seeing

The concept of a surface-shroud resonance was discovered during the development of the family of ARTSCAN models (Fazl,

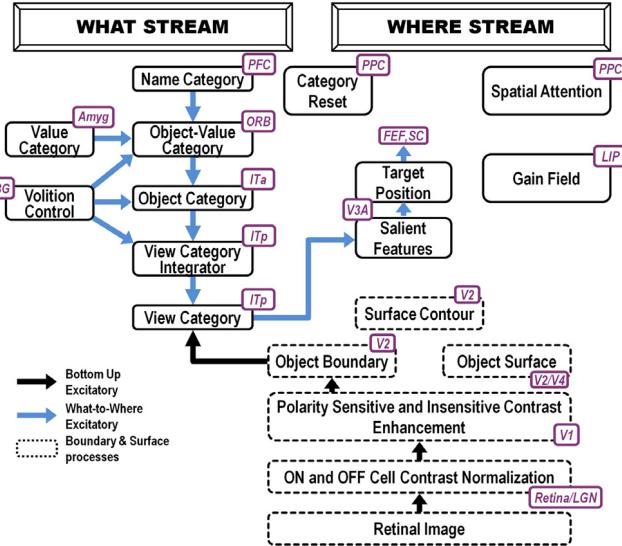


Fig. 18. ARTSCAN Search model macrocircuit (Where's Waldo search for a valued goal object). How the model can search for a valued goal object using What-to-Where interactions. See text for details and Figs. 38 and 39. Source: Reprinted with permission from Chang et al. (2014).

Grossberg, & Mingolla, 2009; Grossberg, 2009), whose culmination is currently the 3D ARTSCAN Search model (Figs. 17 and 18; Chang, Grossberg, & Cao, 2014; Grossberg, Srinivasan, & Yazdabakhsh, 2014). The 3D ARTSCAN Search model proposes explanations of data about visual consciousness that go beyond the role of a surface-shroud resonance in regulating learning of the invariant recognition categories that enable us to know about the objects that are consciously seen during the corresponding surface-shroud resonance. In particular, this model helps to explain how consciously visible surface percepts are computed in retinotopic coordinates, even though the perceptual stability that is experienced during eye movements is achieved by computing attentional shrouds in head-centered coordinates that are invariant under eye movements; see Section 12.

The family of ARTSCAN models explains how invariant object categories are learned as our eyes scan a scene. It thus contributes to the area that is called “active vision” in the computer vision community. To this end, the model proposes solutions to basic problems that confront a human or other primate as it learns to recognize objects in a 3D scene while scanning the scene with eye movements. One such problem is that, as our eyes scan a scene, two successive eye movements may focus on different parts of the same object or on different objects. How does the brain avoid learning to erroneously classify views of different objects together, without an external teacher? One cannot say that the brain does this by knowing that some views belong together whereas others do not, because this can happen even before the brain has a concept of what the object is. Indeed, such scanning eye movements may be used to learn the object concept in the first place.

The ARTSCAN models show how this may be accomplished by coordinating spatial attention and eye movements in the Where cortical processing stream with object attention and invariant object category learning in the What cortical stream (Mishkin et al., 1983). This coordination is only possible because a surface-shroud resonance can sustain spatial attention upon an object surface while its invariant object category is being learned. And one cannot understand what an object surface is without the kind of overview that is given above of how such a surface forms.

The Where stream is also called the How stream because of its important role in controlling actions in space (Goodale & Milner, 1992). This proposed link of space to action gains new significance in terms of my proposal of how the conscious qualia of

unoccluded surfaces help to determine reachable target positions in space. The cART model hereby links consciousness to action. To emphasize this link, the Where stream will henceforth be called the Where/How stream.

The prediction that a surface-shroud resonance supports percepts of conscious visual qualia answers a long-standing question that arose from two earlier predictions: (1) ART predicted that “all conscious states are resonant states” (Grossberg, 1980), and (2) the FACADE and 3D LAMINART models predicted that “all conscious visual qualia are surface percepts” (Grossberg, 1994). These two predictions led to the question: What type of resonance supports surface percepts of conscious qualia?

10.5. Explaining data about visual neglect: Coordinates, competition, grouping, and action

After a delay of 20 years, Grossberg (2013a) finally had enough evidence to propose that surface-shroud resonances play this role. The hypothesis that a surface-shroud resonance has its generators between V4 and PPC helps to explain various data about visual consciousness, such as data about visual neglect (e.g., Bellmann, Meuli, & Clarke, 2001; Driver & Mattingley, 1998; Marshall, 2001; Mesulam, 1999), perceptual crowding (e.g., Bouma, 1970, 1973; Foley, Grossberg, & Mingolla, 2012; Green & Bavelier, 2007; He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Levi, 2008), change blindness and motion-induced blindness (Grindley & Townsend, 1965; Pashler, 1988; Phillips, 1974; Ramachandran & Gregory, 1991; Rensink, O'Regan, & Clark, 1997; Simons & Rensink, 2005), all of which will now be discussed. After providing these explanations of clinical and perceptual data, the text will return in Sections 11–13 to the question of how a surface-shroud resonance maintains spatial attention on an object of interest as it controls the eye movements that scan the object surface in order to learn an invariant object category. As this process takes place, the brain can consciously see the object at the same time that it knows how to search for and recognize it in a cluttered scene. In particular, this summary will explain how we consciously see the object in *retinotopic* coordinates, even while the perceptual stability that is maintained during scanning eye movements is achieved by computing the attentional shroud in head-centered, or *spatial* coordinates that are invariant under eye movements.

The clinical literature on visual neglect is extensive, complex, and subtle, with uncertainties and debates still ongoing (e.g., Behrmann, Geng, & Shomstein, 2004; Husain & Nachev, 2007; Karnath & Rorden, 2012; Nachev & Husain, 2006). Despite these uncertainties, there is convergent evidence that a particular region of PPC, the inferior parietal lobule (IPL), plays an important role in visual neglect. Karnath, Ferber, and Himmelbach (2001) had concluded that the superior temporal cortex (STC), not the IPL, is important in neglect, but other investigators raised doubts about this strong claim: “the disparity between this new proposal and the results of other studies has attracted considerable debate and has been attributed to differences in clinical selection criteria, quality of imaging and the methods of ensuring comparisons between homologous structures across subjects...” (Husain & Nachev, 2007, p. 34). Other studies (e.g., Mort et al., 2003) provided compelling evidence that the IPL “is the crucial anatomical correlate of neglect” (Behrmann et al. (2004, p. 215)), leading Karnath and Rorden (2012) to describe a perisylvian network for spatial neglect that includes IPL and STC.

One reason for the subtlety of data about visual neglect is, as we shall see in this section, that properties of surface-shroud resonances, including the competitive attentional mechanisms that allow shrouds to be selected and the coordinates in which they are computed, interact with all the visual processes that go into creating visible surface representations, including properties

of boundary completion and figure-ground separation, as well as with category learning and recognition processes, and the motor planning and execution processes that attended surfaces feed.

Despite these complexities, a main fact about visual neglect is immediately clear from the most basic properties of a surface-shroud resonance: The very fact that visual neglect occurs at all when an appropriate region of parietal cortex such as IPL is lesioned is immediately clarified by the proposed role of a surface-shroud resonance in visual consciousness: Assuming that IPL is part of the PPC anatomy that supports surface-shroud resonances, a lesion of IPL will prevent the corresponding visual surface representations in V4 from reaching consciousness. Despite the apparent simplicity of this assertion, this hypothesis can also explain subtle clinical properties of visual neglect. The classical article of Driver and Mattingley (1998) reviews visual neglect properties in individuals who have experienced IPL lesions, particularly in the right hemisphere. The text below takes as explanatory targets properties emphasized in that article, which include the following:

Head-centered shroud coexists with retinotopic surface qualia. A neglect patient who appeared to be blind in the left visual field when fixating straight ahead, or to her left, could detect events in her left visual field when she fixated to the right (Kooistra & Heilman, 1989). This shows that neglect is not computed entirely in retinotopic coordinates. Rather, these and related data can be explained by the fact that the shroud is computed in head-centered coordinates that are invariant under eye movements, even though consciously seen surface qualia are computed in retinotopic coordinates that move with the eyes (see Sections 12.3 and 12.5–12.7).

Competition for spatial attention across parietal cortex. Some visual neglect patients show a deficit in response to an isolated stimulus on the affected side, but others do not. In the latter patients, whose lesions are presumed to be less severe, a deficit in awareness only emerges when stimuli are presented simultaneously on both sides; e.g., Posner, Walker, Friedrich, and Rafal (1984). This type of effect can be explained by the spatial competition across the parietal cortex that leads to the choice of a surface-shroud resonance (see Fig. 14 and Section 10.1).

Preserved figure-ground segmentation during neglect. Several kinds of stimuli illustrate that perceptual groupings can restore awareness in a hemifield that would otherwise be neglected. In particular, as noted in the previous paragraph, when two visual stimuli are briefly presented, the more contralateral stimulus can be extinguished from awareness while the ipsilateral stimulus captures attention entirely. However, such extinction can be eliminated if the two events get grouped as a single object, even if the link between the two stimuli is amodally completed behind an occluder (Mattingley, Davis, & Driver, 1997). This effect can be explained by how boundaries can get amodally completed (e.g., Fig. 5, top row, right column) and by how boundary and surface attention can flow behind an occluder. For example, bipole grouping mechanisms can generate amodal boundary completions of occluded object parts in response to the images in Figs. 15 and 16. Spatial attention can flow along such a boundary grouping (“boundary attention”), whether real or illusory, as reported in neurophysiological experiments by Roelfsema, Lamme, and Spekreijse (1998) and Wanning, Stanisor, and Roelfsema (2011), and simulated using the LAMINART model in Grossberg and Raizada (2000). In addition, surface attention can flow from one part of a surface to another via filling-in (Fig. 10). When these boundary and surface properties act, a shroud that forms in response to the ipsilateral stimulus can flow between the ipsilateral and contralateral stimulus, thereby rendering the latter stimulus visible as well. Such a flow of spatial attention also helps to explain properties of social cognition, such as joint attention between a teacher and a student (Grossberg & Vladusich, 2010) and how it may break down during autism (Grossberg & Seidman, 2006).

Unconscious processing of neglected object identity. Unconscious processing of neglected stimuli can occur in parietal patients to a much larger degree than in patients who have visual cortical lesions yet still exhibit blindsight. The implicit knowledge of parietal patients includes object attributes of neglected stimuli such as their color, shape, identity, and meaning (Mattingley, Bradshaw, & Bradshaw, 1995; McGlinchey-Berroth, Milberg, Verfaellie, Alexander, & Kilduff, 1993). These properties can be explained by the fact that feature-category resonances in the What cortical stream (cf. Table 2 and Fig. 1) can be triggered by visual cortical representations in order to control object learning, recognition, and prediction, and these resonances are not eliminated by a parietal lesion in the Where cortical stream. In other words, there are distinct resonances for conscious seeing and knowing. It is easy to overlook this fact because these resonances are often triggered together, so that we know what familiar objects are when we see them (see Section 13). When the surface-shroud resonance for seeing is eliminated by a parietal lesion, the feature-category resonance for knowing can still be activated by visual stimuli.

A link between visual neglect and motor planning deficits: Seeing to reach. Patients who experience inferior parietal lobe lesions that lead to visual neglect often also experience abnormal motor biases. These include a reluctance to respond to the left, even with the ipsilesional right hand (Heilman, Bowers, Coslett, Whelan, & Watson, 1985), and slowness in initiating leftward movements of the right hand towards visual targets in the left hemispace (Mattingley, Husain, Rorden, Kennard, & Driver, 1998). These data support the hypothesis that surface-shroud resonances between V4 and PPC create visibility properties whereby PPC can contribute to guiding actions to reachable surfaces (see Section 10.3).

The converse problem can also occur: Patients who exhibit visual agnosias and cannot recognize basic properties of object shape can, as in the case of patient DF, nonetheless carry out effective reaches to these objects (Goodale & Milner, 1992; Goodale, Milner, Jakobson, & Carey, 1991).

IPL lesions lead to deficits in sustained visual attention. Lesions of the right IPL that cause visual neglect also impair the ability to maintain visual attention over sustained temporal intervals (Rueckert & Grafman, 1998), whether for visual or auditory attention (Robertson et al., 1997). This impairment can be explained by the fact that a surface-shroud resonance maintains spatial attention on an object surface (see Fig. 14 and Section 12), so that when a suitable parietal lesion occurs, this resonance collapses, and along with it the ability to sustain spatial attention.

10.6. Explaining data about visual crowding: A unified view of crowding, neglect, and search

An explanation of data about crowding requires more background about how ART, notably ARTSCAN, supports learning of recognition categories. An outline of this explanation is immediately given to set the stage for further clarification in the subsequent paragraphs. As in the case of visual neglect, explaining the full range of crowding data also requires understanding of visual processes such as boundary completion and figure-ground separation.

ARTSCAN's explanation of visual crowding (Foley et al., 2012), and with it an explanation of the Koch and Tsuchiya (2007) discussion of how “subjects can attend to a location for many seconds and yet fail to see one or more attributes of an object at that location” (see Section 2), follows from its analysis of how a surface-shroud resonance regulates both seeing and recognition of objects. In particular, this analysis proposes how the spatial attention in the Where/How cortical stream that supports conscious seeing via a surface-shroud resonance, also regulates object recognition in the What cortical stream via a feature-category resonance (Table 2).

In the simplest crowding paradigm, an object, such as a letter, is visible and recognizable when presented by itself, but it is not recognizable when the letter is presented concurrently with similar flanking letters (Levi, 2008; Toet & Levi, 1992). The distance between the target letter and the flanking letters at which the target letter becomes unrecognizable is a function of the eccentricity of the target and the flankers, and their relative positions, with larger letter sizes and spacing between the letters needed to see and recognize a target letter as the eccentricity of the letters from the fovea increases (Bouma, 1970, 1973; Levi, 2008).

Given the proposed link within ARTSCAN between conscious perception and recognition, the most basic percepts of visual crowding have a simple explanation; namely, if several objects share a single surface-shroud resonance – that is, they are all covered by a single spatially-extended shroud – then they cannot be individually recognized.

Under what conditions can multiple objects all be covered by a single shroud? Cortical area V1 embodies a *cortical magnification factor* whereby spatial representations of retinal inputs get coarser as they are moved from the foveal region to the periphery (Daniel & Whitteridge, 1961; Fischer, 1973; Polimeni, Balasubramanian, & Schwartz, 2006; Schwartz, 1984; Tootell, Silverman, Switkes, & DeValois, 1982). A group of spatially close objects that have their own shrouds when viewed on the fovea may thus share a single shroud when they are viewed in the periphery. Since surface-shroud resonances create a link between conscious perception and recognition, objects that share a single shroud cannot be individually seen or recognized.

Foley et al. (2012) have supported this qualitative explanation of visual crowding by using the distributed ARTSCAN, or dARTSCAN, model to simulate how objects that have their own shrouds when viewed by the fovea may be enveloped by a single shroud when they are moved to the retinal periphery. In addition, if the objects and their spacing are magnified by the cortical magnification factor as they are moved to the periphery, then they retain their own shrouds and can thus still be individually seen and recognized. Foley et al. (2012) used the same model to simulate other challenging psychophysical data that involve manipulations of spatial attention, including the reaction times that are caused by intra-object attention shifts, inter-object attention shifts, and shifts between visible objects and covertly cued locations (Brown & Denney, 2007; Egly, Driver, & Rafal, 1994); individual differences in reaction times for invalid cues in the two-object cueing paradigm (Roggeveen, Pilz, Bennett, & Sekuler, 2009); and individual differences in detection rate of peripheral targets in useful-field-of-view tasks, including data from video game players (Green & Bavelier, 2003, 2007).

Recent psychophysical experiments have noted how factors like perceptual grouping and figure-ground separation can influence crowding percepts, just as they can influence the visual neglect properties that were reviewed in Section 10.5. It seems that data about crowding can also be explained by the same combinations of mechanisms. For example, Herzog, Sayim, Chicherov, and Manassi (2015, p. 1) assert “that the spatial configuration across the entire visual field determines crowding. Only when one understands how all elements of a visual scene group with each other, can one determine crowding strength. We put forward the hypothesis that appearance (i.e., how stimuli look) is a good predictor for crowding, because both crowding and appearance reflect the output of recurrent processing...” Implicit in this assertion is that boundary grouping and surface filling-in properties can influence, and be influenced by, the allocation of spatial attention, which in turn can influence conscious visual “appearance” and thus, through Where-to-What interactions, recognition.

Recent psychophysical experiments have also described circumstances where adding additional flankers can lead to a release from crowding; e.g., Manassi, Lonchampt, Clarke, and Herzog (2016). Such results emphasize that surface-shroud resonances

have always been only been part of the story. Once psychophysical manipulations influence boundary grouping and surface-filling in properties that can cause more or less crowding, the utility of even using the word crowding begins to weaken. What one is then studying are how boundary, surface, spatial attention, and object attention processes interact to regulate seeing and recognizing. This is a huge issue that goes way beyond classical concerns of crowding experiments.

Crowding data are not the only ones that require such a perspective. The same is true, as one of many examples, of visual search. That is the main point of the article by Grossberg, Mingolla, and Ross (1994) whose title “A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations” emphasized the role of these four kinds of processes. This article developed the Spatial Object Search, or SOS, model to explain and simulate many classical and recent search data whose properties are clarified by interactions of these processes.

In summary, visual neglect, crowding (or the lack thereof), and visual search data probe different aspects of a similar set of interacting processes. Further experimental and modeling studies of crowding and visual search, and their interactions, from this perspective are much to be desired and, as illustrated by the modeling simulations of Fazl et al. (2009) and Foley et al. (2012), need to include surface-shroud resonances as one part of a unifying theory.

10.7. Explaining data about change blindness and motion-induced blindness

Change blindness experiments provide important examples of the crucial role of spatial attention in supporting percepts of conscious qualia. In one striking classical example, alternating displays of an original and a modified scene are separated in time by brief blank fields (Rensink et al., 1997). Even large changes that are persistently repeated between the alternating scenes – such as the removal and replacement of an airplane wing – may not be noticed for a long time, even though these changes are readily seen when spatial attention is directly focused at the scenic positions where these changes occur.

One mechanism that contributes to an explanation of change blindness is the competition for spatial attention that occurs across the parietal cortex, and that helps to select winning shrouds (Fig. 14). The same property also played a role in explaining the deficit in awareness of visual neglect patients when stimuli are presented simultaneously on both sides of a display (Section 10.5). This competition uses a network of cells that obey the membrane equations of neurophysiology, otherwise called shunting dynamics. When such cells compete in an on-center off-surround network, they tend to conserve the total activity across the network, so that activating more cells may make each cell less active, other things being equal. This property is called *contrast normalization* (Grossberg, 1973, 1980; Heeger, 1992). Due to contrast normalization, when attention focuses upon one position, activity decreases at other positions.

A second important factor is that the alternating scenes are separated by blank intervals. Thus, each time a scene turns on or off, it activates cells across the entire scene that are sensitive to transient changes in each scenic input. Such transient on-cells and off-cells play a critical role in the 3D FORMOTION neural model of motion perception (e.g., Baloch & Grossberg, 1997; Baloch, Grossberg, Mingolla, & Nogueira, 1999; Berzhanskaya, Grossberg, & Mingolla, 2007; Chey, Grossberg, & Mingolla, 1997; Grossberg, Mingolla, & Viswanathan, 2001). During the change blindness manipulation, these transient signals flood each just presented or withdrawn scene.

The third important factor is that transient scenic changes automatically attract spatial attention (e.g., Corbetta & Shulman,

2002; Yantis & Jonides, 1990), a property that has also been explained within the 3D FORMOTION model. When these three properties (contrast normalization, transient cell activation, attention automatically drawn to transients), work together in response to change blindness stimuli, spatial attention is spread thinly across the entire scene, so that only the most persistent scenic properties – the ones that occur in both scenes – may have enough activity to resonate sufficiently to be consciously seen. As soon as spatial attention is volitionally focused on the region of change, however, a strong surface-shroud resonance can form around that region. Foley et al. (2012) also model why the remainder of a scene does not go totally dark when such a surface-shroud resonance of focused spatial attention forms.

Motion-induced blindness (Grindley & Townsend, 1965; Ramachandran & Gregory, 1991) is closely related to change blindness. During motion-induced blindness, salient objects may fluctuate into and out of conscious awareness when superimposed onto certain global motion patterns. For example, peripherally-viewed yellow dots may intermittently fluctuate in and out of consciousness when a rectangular blue-and-black rectangular grid that intersects and extends beyond them rotates around a flashing central fixation point. As noted by Bonneh, Cooperman, and Sagi (2001, p. 798), motion-induced blindness may be influenced by “perceptual grouping effects, object rivalry, and visual field anisotropy...Disappearance might reflect a disruption of attentional processing...uncovering the dynamics of competition between object representations within the human visual system”.

These percepts also depend upon all the factors that have been mentioned to explain data about visual neglect and crowding. For example, boundary grouping helps to determine whether individual objects fluctuate into and out of awareness independently, or together when they are parts of a single object. The latter kind of situation can occur, for example, when two disks are connected into a dumbbell by a single-pixel line. Mitroff and Scholl (2005) showed in this case “that object representations can be formed and updated without awareness” by making changes in displays when they were out of awareness (e.g., connecting or disconnecting the disks with a line) and demonstrating how these changes could influence whether the objects reappeared simultaneously or separately. A boundary grouping can occur automatically and pre-attentively (e.g., Fig. 4) and, once formed, enable spatial attention to flow along the entire grouping as a conscious percept emerges through a surface-shroud resonance (e.g., Figs. 14 and 19), in particular between the connected disks.

Motion-induced blindness also illustrates the role of transient detectors that continually refresh the moving grid and thereby support its visibility, whereas the peripheral stationary dots do not activate their transient detectors to the same degree when the observer fixates on a central flashing fixation point. This difference can strengthen the boundaries of the moving grid relative to those of the stationary dots, and therefore favor the competitive advantage of the grid during boundary competition.

In response to any moving target, the visual boundary system also triggers an inhibitory rebound after a boundary moves by in order to prevent uncontrolled boundary persistence, and consequent perceptual smearing, or streaking, in its wake (Wallis & Arnold, 2009). Francis, Grossberg, and Mingolla (1994) have quantitatively simulated many psychophysical data about visual persistence in which such an inhibitory rebound mechanism plays an important role. Significantly, activity-dependent transient responses play a role in creating these rebounds. If this transiently acting inhibition occurs when a grid boundary moves across a peripheral dot boundary, it can momentarily inhibit the dot boundary, and thus the filling-in of surface color and brightness within the dot boundary, thereby preventing it from competing for spatial attention and visibility. In the case where the stationary dots and the moving grid are yellow and blue, respectively,

opponent color competition may additionally occur when they intersect in space, thereby directly suppressing the surface color of the stationary dot.

The boundary mechanisms that influence motion-induced blindness are presumed to occur in cortical area V2, among other areas, and their predicted dynamics could benefit from additional experiments. Additional tests would be most welcome to further test the form-motion, also called *formotion*, interactions between the form stream, notably cortical area V2, and the motion stream, notably cortical area MT. The form and motion streams have been predicted to be computationally complementary (Table 3; Section 4). A series of neural modeling studies have developed the 3D FORMOTION model (e.g., Baloch & Grossberg, 1997; Berzhanskaya et al., 2007; Francis & Grossberg, 1996b; Grossberg, 1991, 2014), to explain and predict psychophysical and neurobiological data about how these streams work, and interact. These complementary laws enable the orientation-sensitive form stream to generate effective representations of 3D object boundaries and surfaces, and the direction-sensitive motion stream to generate effective representations of object motion direction and speed. The streams interact to generate effective representations of moving-form-in-depth.

Such sharing of featural properties due to interactions across cortical streams has been known for some time. For example, the effects of reversibly cooling V2 on recorded cell properties in MT that were reported by Ponce, Lomber, and Born (2008) were anticipated by earlier 3D FORMOTION model predictions (Grossberg, 1991). Even more recent simultaneous recording methods have demonstrated widespread sharing of cue selectivities across cortical streams (Siegel, Buschman, & Miller, 2015). These methods may allow additional tests of predicted interactions between the form and motion streams, as well as other computationally complementary cortical streams, to be made.

10.8. Explaining many data with the same model mechanisms

The prediction that surface-shroud resonances support conscious percepts of visual qualia is consistent with many additional FACADE and 3D LAMINART model simulations of the parametric properties of consciously seen surfaces during 3D vision and figure-ground perception, including percepts of random dot stereograms (Fang & Grossberg, 2009); Necker cubes (Grossberg & Swaminathan, 2004); unique and bistable transparency (Grossberg & Yazdanbakhsh, 2005); Panum's limiting case, dichoptic masking, the Venetian blind illusion, and da Vinci stereopsis (Cao & Grossberg, 2005; Grossberg & Howe, 2003; Grossberg & McLoughlin, 1997); texture segregation (Bhatt et al., 2007; Grossberg & Pessoa, 1998); 3D shape-from-texture (Grossberg et al., 2007); visual persistence (Francis & Grossberg, 1996a; Francis et al., 1994); Bregman–Kanizsa figure-ground separation, Kanizsa stratification, and the Munker–White, Benary cross, and checkerboard percepts (Kelly & Grossberg, 2000); the McCullough effect (Grossberg, Hwang, & Mingolla, 2002); the watercolor illusion (Pinna & Grossberg, 2005); and illumination-discounted and anchored brightness and color percepts (Grossberg & Hong, 2006; Grossberg & Todorovic, 1988; Hong & Grossberg, 2004). These explanations and simulations of many different kinds of surface percepts, and their parametric properties in psychophysical data, provide accumulating evidence for the correctness of the prediction that surface-shroud resonances, interacting with boundary and surface perceptual processes, support percepts of conscious visual qualia.

10.9. Shrouds have complex internal structure that influences properties of visual search

These modeling results clarify that 3D surface representations are typically distributed over multiple depth-selective and color-selective FIDOs. The shrouds with which they resonate reflect these

structural distinctions, which influence how easily different combinations of surface features can be searched (e.g., Grossberg et al., 1994; He & Nakayama, 1992; Humphreys, Quinlan, & Riddoch, 1989; Nakayama & Silverman, 1986; Treisman & Gelade, 1980; Treisman & Gormican, 1988; Wolfe, 1994; Wolfe, Cave, & Franzel, 1989). These complexities of internal shroud structure in the parietal cortex (e.g., Silver & Kastner, 2009; Swisher, Halko, Merabet, McMains, & Somers, 2007) will not be further discussed herein.

11. From view- and position-specific categories to view- and position-invariant categories

11.1. Conscious visual qualia are computed in retinotopic coordinates

Let us now selectively review the implications for consciousness of the 3D ARTSCAN Search model, notably how including scanning eye movements into discussions of visual awareness, learning, and recognition allows a deeper analysis of (1) how the brain learns invariant object recognition categories that can individually respond to multiple views, positions, and sizes of an object in a 3D scene (Cao et al., 2011; Fazl et al., 2009; Foley et al., 2012; Grossberg, 2009; Grossberg et al., 2011), and (2) how searches for valued objects in such a scene are accomplished (Chang et al., 2014; Grossberg et al., 2014). Along the way, I will offer explanations of data that have not been explained before.

Let us begin with the observation that the visual qualia that we consciously see in a scene shift in a direction opposite to our eye movements. In other words, *conscious visual qualia are computed in retinotopic coordinates*. Understanding how this happens sheds light on data showing that percepts of visual qualia may be influenced by computations in cortical areas V1 and V2. For example, studies of visual perception and visual imagery in humans using PET and rTMS show activation of V1 during both kinds of percepts, and interference of both by applying rTMS to V1 (Kosslyn et al., 1999). How can these facts be reconciled with the hypothesis that a surface-shroud resonance is triggered between V4 and PPC? The simplest explanation is that V4 cannot operate if it does not receive inputs from V1. This fact does not, however, fully explain how V1 and V2 contribute to conscious percepts when they are both intact. The following analysis leads to additional insights.

11.2. Visual imagery: Basal ganglia volition converts top-down modulation to driving inputs

Before turning to this discussion, a related question can quickly be answered: How is visual imagery possible if top-down expectations are modulatory in their on-center, as the ART Matching Rule proposes (Section 7.1)? Would not conscious visual imagery require supra-threshold, indeed resonant, cell activation to occur? How does a modulatory top-down expectation cause a resonant activation? Fig. 12 includes an answer to this question. Volitional signals can alter the balance between excitation and inhibition in the on-center of the top-down expectation by weakening inhibition relative to excitation. The on-center of the expectation can then fire targeted cells and lead to resonance. These volitional signals are controlled by the basal ganglia. This effect of the basal ganglia on whether cortical feedback is modulatory or driving is just one of many examples of basal ganglia volitional control that neural models have explained. See Grossberg (2016b) for a review.

12. Learning invariant object categories using Where-to-What stream interactions

12.1. From complementary consistency to figure-ground separation: Surface contours

Let us now consider how an invariant object category is learned during free scanning of a scene with eye movements, and how that clarifies data about conscious seeing and recognition.

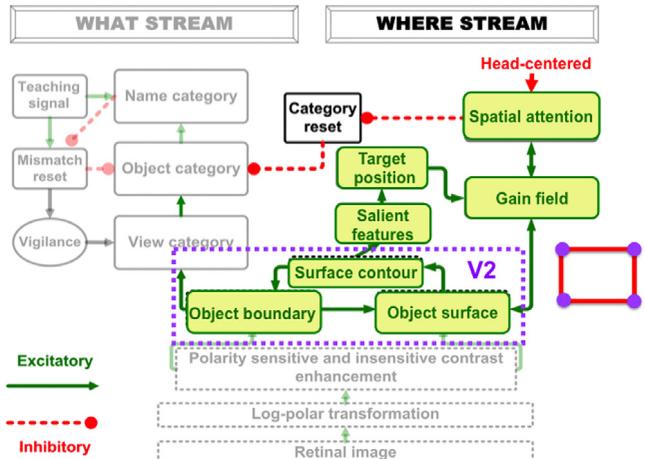


Fig. 19. How ARTSCAN maintains spatial attention in the Where cortical stream upon an object surface via a surface-shroud resonance, while the What cortical stream learns view-specific categories that get associated with view-invariant object categories. While the attentional shroud focuses attention upon the object surface, it also inhibits parietal reset cells. When spatial attention shifts to another object, this inhibition shuts off, thereby disinhibiting the reset cells and enabling them to send a transient inhibitory burst to the invariant category representation, which is then also inhibited. A shift of spatial attention and learning of an invariant object category for the newly attended surface can then proceed. Feedback interactions between the object's boundaries and surfaces include surface contour signals that assure perceptual consistency while also initiating figure-ground separation within cortical area V2 (see Fig. 20). Surface contour signals are largest at high curvature boundary positions (see red surface contour rectangular boundary and purple high curvature positions). These high curvature positions mark locations of salient features, and can be converted into eye movement target positions that enable the eyes to explore the object surface while it is attended. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In order to be able to focus spatial attention upon one object rather than another, the brain must first separate the surface representations of different objects from each other via figure-ground separation. Fig. 19 summarizes one step in this process, which is closely linked to the way in which visual boundaries and surfaces are computed. As noted in Section 6, visual boundaries and surfaces are computed using computationally complementary laws. How, then, is a consistent percept generated from these complementary processes? I call the process whereby the brain converts complementary boundary and surface computations into a consistent percept *complementary consistency*. How this is predicted to occur was first proposed in Grossberg (1994, 1997), where it was also shown that, remarkably, the process that assures perceptual consistency also initiates figure-ground separation.

The process that achieves complementary consistency is proposed to occur using feedback signals between boundaries in the V2 interstripes and surfaces in the V2 thin stripes (Fig. 9). As noted in Figs. 10 and 19, object boundaries act as barriers to the filling-in of object surfaces. In addition, successfully filled-in surfaces generate topographic feedback signals to the boundaries that generated them. These feedback signals, which are called *surface contours* (Fig. 19), are generated by contrast-sensitive on-center off-surround networks that act across space and within each depth. Because of their contrast-sensitivity, these networks generate output signals only at positions where they detect a rapid change in contrast across space. Such rapid contrast changes occur only at the contours of successfully filled-in surfaces, which are the surfaces that are surrounded by closed boundaries (Fig. 10, bottom row, left column). Surface contours are not generated at positions where open boundaries occur (Fig. 10, bottom row, right column) because the surface filling-in that is caused by feature contours within regions with open boundaries can spread to both sides of their boundaries, and thus do not generate large contrasts at boundary positions.

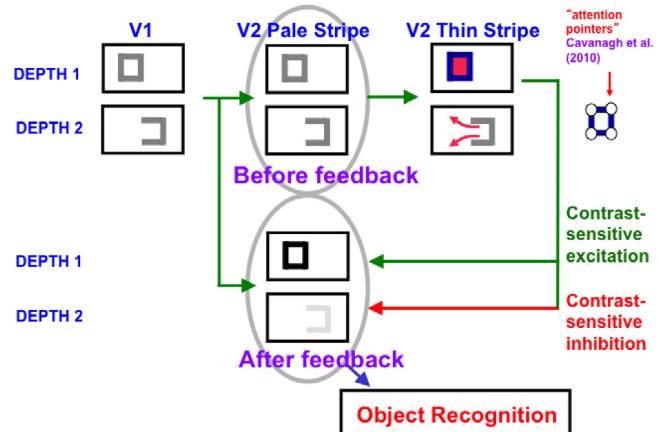


Fig. 20. How surface contour signals initiate figure-ground separation. The filled-in surface color or brightness of a region that is surrounded by a closed boundary (at Depth 1 of V2 thin stripes) can generate surface contour signals at the positions of that boundary. These surface contour signals strengthen the boundaries that induced the surface (at Depth 1 of V2 pale stripes), while inhibiting the spurious boundary signals at the same positions, but further depths (at Depth 2 of V2 pale stripes). With these spurious boundaries eliminated, partially occluded objects (not shown) can be amodally completed at the further depth plane (Depth 2) via collinear bipole boundary completion. A percept of two figures, one partially occluding another, can thereby be generated, as in response to the image in Fig. 15 (left column). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The example shown in Fig. 20 illustrates the simple case where a rectangular boundary is viewed in depth, but only its left vertical boundary can be seen by both eyes, as often happens in examples of daVinci stereopsis when distant surfaces are partially occluded by nearer surfaces to different extents in the two eyes (Nakayama & Shimojo, 1990). Cao and Grossberg (2005, 2012), Grossberg and Howe (2003), and Grossberg and McLoughlin (1997) have explained and simulated a variety of daVinci stereopsis percepts. Here only the simple example in Fig. 20 will be discussed in order to illustrate its relevance for figure-ground separation and invariant object category learning and recognition.

In Fig. 20, there is an open boundary at Depth 2 that is created by two horizontal boundaries and one vertical boundary, where the vertical boundary abuts the horizontal boundaries at their rightmost ends. The vertical boundary is caused by viewing this part of the daVinci stereopsis image with only one eye. Such monocularly-viewed boundaries do not represent any particular depth, so are projected along the line of sight to all depth-selective boundary representations. That is why the right vertical boundary is projected to both Depth 2 and Depth 2. The horizontal boundaries also do not generate strong depth signals, so are also projected along the line of sight to all depth-selective boundary representations, notably Depth 1 and Depth 2. The second vertical boundary, to the left of the horizontal boundaries, is generated at Depth 1 by binocular viewing of the daVinci image. It is caused by the binocular disparity of viewing the image at a particular depth.

The surface contour signals that are generated at Depth 1 send excitatory signals to boundaries at their own depth and positions, and inhibitory signals to further depths at the corresponding positions (Fig. 20). As a result, the *closed* boundary is strengthened, whereas open boundaries at the same positions and further depths are inhibited. This feedback assures complementary consistency by selecting and strengthening only the “correct” boundaries that generate surfaces which are worthy of further processing, and can project to object recognition regions in inferotemporal cortex and beyond (Fig. 20).

Inhibition of spurious boundaries at further depths enables amodal completion to occur at Depth 2 behind the occluding rectangle between any collinear uninhibited boundaries that are gen-

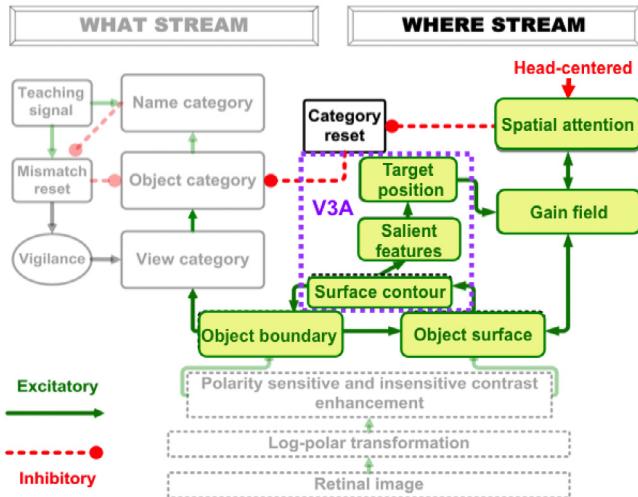


Fig. 21. The transformation from positions of strongly activated surface contours into target positions for eye movements that explore an attended object surface is predicted to be carried out in cortical area V3A. See text for details.

erated by the image or scene, as it does in response to the images in Figs. 15 and 16. Figure-ground separation is hereby initiated.

12.2. The main problem: Why inhibiting view categories does not inhibit invariant categories

The above considerations imply that the process which enables the brain to focus spatial attention on different surfaces must occur *after* their figure-ground separation occurs in V2. With this background, we can now turn to how the brain learns view-invariant categories as the eyes scan a scene.

Suppose that the first view of a novel object leads to rapid learning of a view-specific category in posterior inferotemporal cortex (ITp). Suppose that this view category also activates some cells in anterior inferotemporal cortex (ITa) that will become a view-invariant object category via associative learning with multiple view-specific categories (Figs. 17 and 19). Each view-specific category is reset to enable the next view-specific category to be learned when the object view changes sufficiently. When the first view-specific category is reset, the input is removed that it generated from ITp to ITa, and that initially activated the emerging view-invariant category. How does the brain prevent the invariant object category from also being reset when this input from ITp ends? How does the brain maintain activity of these ITa cells while they are associated with multiple view-specific categories of a single object, and thereby become a view-invariant category with which to represent this object?

12.3. Attentional shroud inhibits reset of invariant object category during object search

ARTSCAN predicts that an object's *pre-attentively* formed surface representation in V4 activates an attentional shroud, even before the brain learns to recognize the surface as representing a particular object (Fig. 19). This shroud persists via a surface-shroud resonance during active scanning of an object. The shroud protects the view-invariant category from getting reset, even while view-specific categories are reset, as the eyes explore an object. The shroud does this by inhibiting the reset mechanism while spatial attention is focused on the object, by using the inhibitory connections from the Spatial Attention to Category Reset processing stages in Figs. 21 and 22 (red dotted pathway from Spatial Attention to Category Reset). The reset mechanism is found

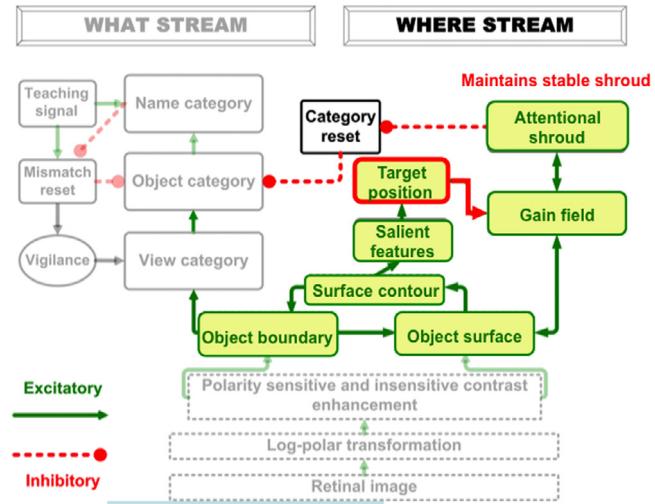


Fig. 22. Target position commands can generate saccadic eye movements to the corresponding positions in space via brain regions such as superior colliculus and frontal eye fields. While these eye movement commands are being converted into eye movements, the target position commands can also much more rapidly activate predictive remapping signals that update gain fields which transform retinotopic surface representations into head-centered attentional shrouds. The shrouds then remain invariant under the eye movements, and can maintain inhibition on the parietal reset cells while the eyes explore the object surface, thereby allowing multiple view-specific categories to be learned and associated with an emerging view-invariant object category.

in the parietal cortex, as is the attentional shroud, as noted in Section 12.4.

When spatial attention shifts from an object, its shroud collapses, thereby disinhibiting the Category Reset stage in PPC (see Fig. 22, red dotted pathway from Category Reset to Object Category stage). A transient burst of inhibition is then released that resets the active invariant object category in ITa. The collapse of the shroud enables the eyes to move to and attend another surface, whereupon new view-specific and view-invariant object categories can be learned because the previously active invariant category has also been inhibited. The cycle can then repeat itself.

12.4. Human and monkey data support shroud reset properties: Explanations and predictions

Chiu and Yantis (2009) used rapid event-related MRI in humans to provide strong evidence for the ARTSCAN prediction of how a surface-shroud resonance in the Where/How stream protects an emerging view-invariant category from being prematurely reset in the What stream. These authors found that a shift of spatial attention evokes a transient domain-independent signal in the medial superior parietal lobule that corresponds to a shift in categorization rules. In ARTSCAN, collapse of an attentional shroud (spatial attention shift) disinhibits the parietal reset mechanism (transient signal) that leads to inhibition of the active invariant object category and instatement of a new one (shift in categorization rules). The transient signal is “domain-independent” because the parietal reset mechanism can be inhibited by spatial attention that focuses upon any object surface, and can reset any active invariant category when it is disinhibited. This experiment provides a useful marker for experimentally testing additional properties of the ARTSCAN model and its variants.

One such additional test can be derived from an ARTSCAN extension that proposes how view-, position-, and size-invariant object categories are learned, not just view-invariant categories. In particular, Cao et al. (2011) developed the positional ARTSCAN, or pARTSCAN, extension of the ARTSCAN model to explain how these additional object category invariances can be learned. They

have used this extended model to simulate neurophysiological data of Li and DiCarlo (2008; see also Li & DiCarlo, 2010) which show that features from different objects can be merged through learning within a single invariant IT category when monkeys are presented with an object that is swapped with another object during an eye movement to foveate the original object. Why does not such a learned merging of features across objects lead to catastrophic forgetting of all learned invariant recognition categories? pARTSCAN simulates the swapping data by showing how the swapping procedure occurs without activating the reset mechanism by instating the swap before the animal can shift its spatial attention.

The Li and DiCarlo (2008) and Chiu and Yantis (2009) experimental paradigms may be combined to further test the model prediction of how spatial attention may modulate learning of invariant object categories. The prediction is that there should not be a transient parietal burst under the usual target swapping conditions, hence features from different objects could merge through learning. As the delay between the initial target and the swap is increased, a reset should occur at a sufficiently long inter-stimulus interval. When that happens, the learning of merged categories should either be prevented or significantly attenuated.

Many other paradoxical data have been explained by these concepts. By providing a mechanistic account to explain each of these data, each explanation constitutes a prediction that can benefit from additional experimental tests. These explanations include how spatial attention can increase the perceived brightness of a surface (see Fig. 14; Carrasco, Penpeci-Talgar, & Eckstein, 2000; Reynolds & Desimone, 2003), how the eyes can prefer to move within the same object for awhile (see Section 12.5; Theeuwes, Mathot, & Kingstone, 2010), how predictive remapping of eye position occurs (see Section 12.6; Duhamel, Colby, & Goldberg, 1992; Gottlieb, Kusunoki, & Goldberg, 1998; Melcher, 2007), and what sort of category invariance can be learned, given the limitations that are imposed by the cortical magnification factor (Grossberg et al., 2011; Zoccolan, Kouno, Poggio, & DiCarlo, 2007). Data about predictive remapping are particularly relevant to understanding how we consciously see visual qualia that are retinotopically coded.

12.5. A surface-shroud resonance enables the eyes to explore multiple object views

Surface contour signals can generate target position commands whereby eye movements can foveate salient features of an object surface while a surface-shroud resonance maintains spatial attention upon that surface (Fig. 19). This happens because surface contours are computed using a contrast-sensitive on-center off-surround network whose activities are largest at high curvature positions along the object boundary (see the rectangular surface contour with high-curvature-enhanced, activation corners in Figs. 19 and 20). These high curvature positions mark the locations of salient features. They can therefore be chosen, one at a time, by a second competitive network that chooses the most active position as a target position for the next eye movement. In this way, the eyes can explore salient features while the object is attended, and different view-specific categories can be learned during this exploratory process. These target positions hereby control shifts of spatial attention across an attended object, and have properties that Cavanagh, Hunt, Afraz, and Rolfs (2010) have called *attention pointers* (Fig. 20).

Where does this transformation from visual computations to motor commands occur in the brain? This is proposed to occur in cortical area V3A (Fig. 21). As noted by Caplovitz and Tse (2007, p. 1179): “neurons within V3A...process continuously moving contour curvature as a trackable feature...not to solve the ‘ventral problem’ of determining object shape but in order to solve the ‘dorsal problem’ of what is going where”.

12.6. Predictive remapping: Gain fields maintain shroud stability

When an eye movement occurs on an object surface, why does not the shroud shift so much that it disinhibits category reset cells which can then inhibit the emerging invariant category? ARTSCAN proposes that this is accomplished by computing shrouds in head-centered coordinates that do not move when the eyes move. This transformation of a retinotopic surface representation into a head-centered shroud is proposed to occur by using *gain fields* (Fig. 22). Gain fields are populations of cells that are activated by outflow eye movement target position signals and used to transform the retinotopic coordinates of the attentional shroud (Figs. 21 and 22). The (target position)-to-(gain field) signals that are used to update a head-centered shroud are assumed to occur very quickly, before an eye movement is complete, to preserve the shroud’s head-centered representation.

This rapid coordinate change is achieved by *predictive remapping*. Predictive remapping has been used to interpret neurophysiological data about how parietal representations are updated by intended eye movements (Duhamel et al., 1992; Gottlieb et al., 1998; Mathot & Theeuwes, 2010; Melcher, 2007, 2008, 2009; Saygin & Sereno, 2008; Sommer & Wurtz, 2006; Tolia et al., 2001; Umeno & Goldberg, 1997) by using gain fields (Andersen, Essick, & Siegel, 1985, 1987; Andersen & Mountcastle, 1983; Deneve & Pouget, 2003; Fazl et al., 2009; Gancarz & Grossberg, 1999; Grossberg & Kuperstein, 1986; Pouget, Dayan, & Zemel, 2003); see LIP in Fig. 17. The ARTSCAN model links the process of predictive mapping to a host of related processes whereby spatial attention is maintained upon an object surface, eye movements to salient features within the object are controlled, and view-specific and invariant object categories are learned. Data explanations by the model provide multiple experimental markers in several brain regions. Further tests of these predictions with psychophysical, multiple microelectrode, and functional neuroimaging methods should disclose a wealth of important new information about these fundamental processes, and how they interact.

12.7. Both retinotopic and spatial coordinates are needed during active vision

How perceptual stability is maintained as the eyes actively scan a scene leads to the realization that interacting combinations of retinotopic and head-centered, or spatial, representations can be expected to be found in the neocortex. In particular, conscious visual percepts shift with the positions that are foveated by the eyes, and are thus computed in retinotopic coordinates. Shrouds, in contrast, are computed in spatial coordinates to maintain sustained spatial attention upon an object surface, while enabling an invariant object category to be learned. fMRI data of Burr and Morrone (2011, p. 504) illustrate this subtlety: “We firstly report recent evidence from imaging studies in humans showing that many brain regions are tuned in spatiotopic [head-centered] coordinates, but only for items that are actively attended”. The current theory provides a clear mechanistic explanation of this otherwise potentially confusing assertion.

12.8. Perceptual stability: Binocular fusion during eye movements

Predictive remapping with gain fields is not only used to maintain shroud stability. It also directly impacts how a 3D scene is consciously seen as it is scanned with eye movements. In particular, it seems obvious that our brains seamlessly maintain binocular fusion as eye movements occur, thereby supporting stable percepts through time of the world in depth.

This apparent simplicity disguises complex underlying machinery, as can be realized by noting that each eye movement causes

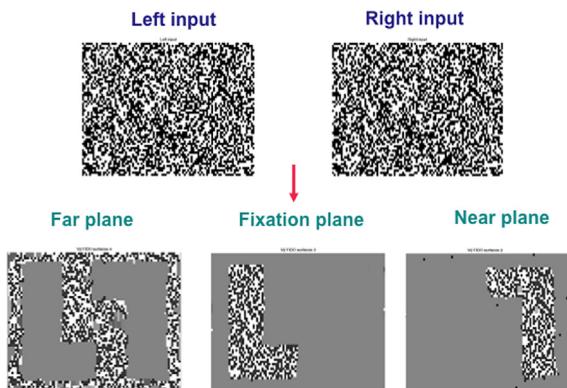


Fig. 23. (Top row) A binocular stereogram may take some time to be binocularly fused. Once fused, however, the eyes can move across the fused images without breaking fusion. The text explains how this may happen. (Bottom row) Computer simulation of the 3D surface representation that is generated by the 3D LAMINART model in response to the stereogram stimulus in the top row.
Source: Adapted with permission from Fang and Grossberg (2009).

each of our foveas to process a different set of scenic features on each retina. There could be temporal breaks in binocular fusion after each eye movement if fused boundaries were computed only in retinotopic coordinates. Such breaks are not, however, consciously perceived. How does the brain convert intermittent fusions into a stable 3D percept that persists across eye movements?

Although perceptual stability during eye movements is so familiar that it does not initially seem to require explanation, this is no longer the case when we binocularly fuse a random dot stereogram (Fig. 23, top row) or a Magic Eye autostereogram, in which the number of salient fusible features is less than in a typical natural scene. It may take tens or hundreds of milliseconds to *initially* binocularly fuse the left and right eye images in these stereograms. Fig. 23 (bottom row) shows the 3D surface percept that is generated by the 3D LAMINART model in response to the binocular stereogram in Fig. 23 (top row; Fang & Grossberg, 2009). After fusion occurs, if we scan the fused image with eye movements, fusion is maintained even though all the retinotopic matches are different after each eye movement. Why does not each eye movement require tens or hundreds of milliseconds to binocularly fuse each new set of retinotopic matches?

The 3D ARTSCAN model (Grossberg et al., 2014) predicts how this happens by augmenting the ARTSCAN machinery in Figs. 19–22 with additional processes of predictive remapping. In addition to the predictive remapping that enables an attentional shroud to modulate learning of invariant object categories, predictive remapping circuitry with additional gain fields enables head-centered binocular boundary representations to maintain their stability during eye movements, and to control the filling-in of conscious qualia in retinotopic binocular surface representations (Fig. 24). The 3D ARTSCAN model simulates why and how breaks in binocular fusion are not seen in retinotopic binocular surface percepts because their head-centered binocular boundary representations are predictively updated before each eye movement is complete. Fig. 24 summarizes three of the gain fields, at different stages of processing, that support such seamless percepts. This proposal has been supported by simulations of predictive updating of 3D percepts in response to CalTech 101 natural images, and by simulations of reaction time costs during spatial attention shifts under different perceptual conditions.

12.9. Two types of perceptual stability cooperate during active conscious vision

We can now reconcile the claim that a surface-shroud resonance has its generators in V4 and PPC (Section 10), even though

various critical perceptual processes occur in V2 and V1 (Section 11.1). A surface-shroud resonance is triggered when the filled-in 3D surface representations in V4 are complete and stable enough to support the resonance's excitatory feedback interactions. These resonant signals can then also propagate down to V2 and V1, where they can select and focus attention upon signals at these levels that support the ongoing resonances.

The main new fact is that binocular boundary signals in V2 also contribute to perceptual stability: They are transformed from retinotopic into head-centered coordinates, using gain fields, in order to retain the stability of their binocularly-fused boundaries during eye movements, and then are converted back from head-centered to retinotopic coordinates, again using gain fields, in order to maintain the stability of consciously seen binocular surface representations that are computed in retinotopic coordinates (Fig. 24). These retinotopic surface representations include feature-selective cells that can support percepts of conscious qualia, whether in response to bottom-up inputs from the world, or top-down signals that cause percepts of visual imagery.

In summary, the simulated interactions that are summarized in Fig. 24 include the minimal number of processing stages and gain-field-mediated coordinate transformations that are needed to maintain perceptual stability of consciously seen surface qualia as eyes scan a scene. Let us interpret in Fig. 24 the monocular left and right eye surface representations as occurring in V1 and V2, the binocular surface representation that drives a surface-shroud resonance as occurring in V4, the shroud itself as occurring in PPC, and the invariant binocular boundaries as occurring in V2. Then Fig. 24 illustrates how the requirements of perceptual stability include V4 and PPC, as well as lower cortical levels such as V1 and V2, all connected via resonant feedback loops. This predicted circuitry can guide many new experiments about how predictive remapping maintains the stability of 3D percepts as the eyes scan a scene.

13. Seeing and knowing: Synchronous surface-shroud and feature-category resonances

Often when we consciously see a familiar object, we also *know* what it is. ART proposes that these two kinds of awareness are due to different kinds of resonances (Fig. 25), with knowing what is seen supported by feature-category resonances that include What stream regions such as IT, and conscious seeing supported by surface-shroud resonances that include Where/How stream regions such as PPC. How, then, do we know what a familiar object is when we see it? ART proposes that this happens because both resonances interact with shared visual cortical areas, such as V4, and can thus synchronize with each other, often with gamma oscillations (Fries, 2009; Grossberg & Versace, 2008).

14. What and Where/How streams for audition

The concept of a surface-shroud resonance was published before the current article was written. However, the way in which the current article unifies the discussion of the visual processes that feed into it and why, and its implications for consciousness and the control of eye movements and reaching, are new. Also new are explanations of various visual databases that this synthesis enables. The summary of how multiple brain regions interact to learn, recognize, and search for valued visual objects also implies multiple testable neurobiological predictions about how these processes work.

Various of the discussions herein of auditory and speech perception and production processes also represent a new synthesis of separate modeling streams. Other explanations are new, such as the predicted homolog between visual circular reactions for motor-equivalent reaching and auditory circular reactions for

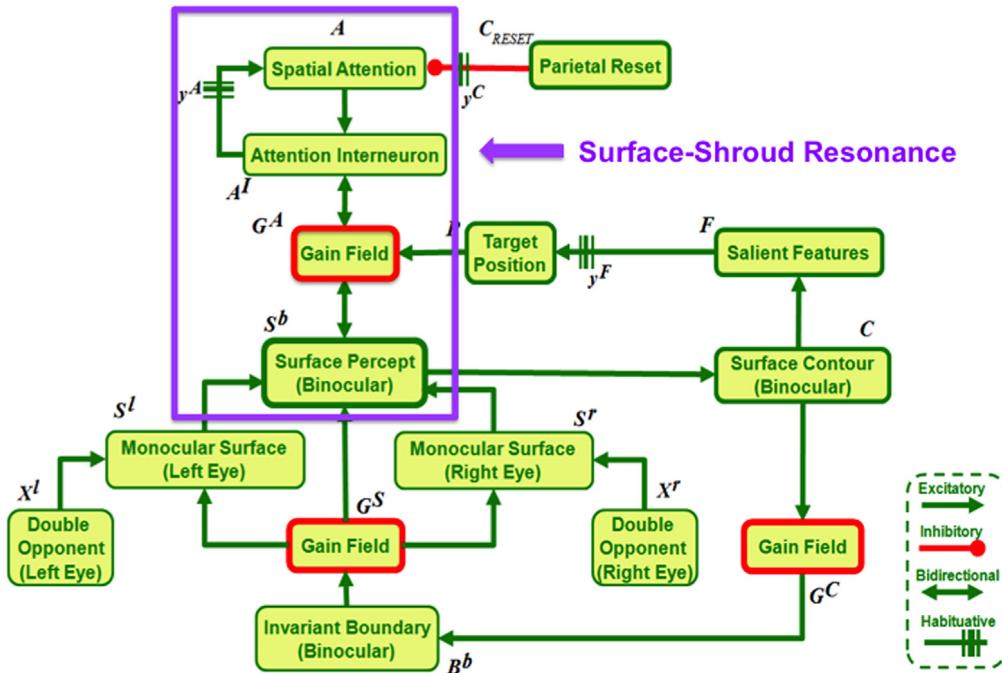


Fig. 24. Circuitry within the 3D ARTSCAN model that maintains the perceptual stability of fused binocular surfaces during eye movements. The circuit uses predictive remapping by gain fields at several processing stages to maintain the stability of binocular boundaries in head-centered coordinates during eye movements, while also supporting rapid filling-in of the consciously seen binocular surface representations in retinotopic coordinates. A surface-shroud resonance between the binocular surface percept in retinotopic coordinates and the spatial attentional shroud in head-centered coordinates is also mediated by gain fields.
Source: Adapted with permission from Grossberg et al. (2014).

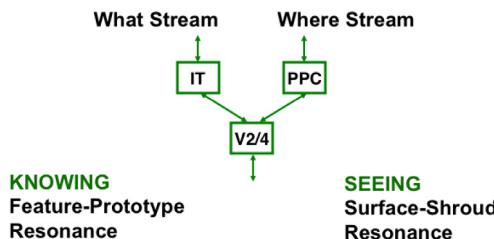


Fig. 25. Seeing and knowing. A surface-shroud resonance that supports conscious seeing and a feature-category resonance that supports conscious knowing, or recognition, can occur simultaneously and be supported by a synchronous resonance that bridges the What and Where cortical streams.

motor-equivalent speaking (Sections 14.2–14.4), and the concept of a stream-shroud resonance that supports conscious hearing (Section 15). So too are concepts that build upon these insights about how auditory consciousness may work, and explanations of various data about topics ranging from auditory neglect to rhythmic performance of speech and songs.

14.1. Ventral sound-to-meaning vs. dorsal sound-to-action: Complementary invariants

The concepts and processes that have helped to understand aspects of conscious visual perception will now be useful towards explaining properties about auditory consciousness, because there are strong homologs between the cortical streams that process vision and audition. Indeed, What and Where/How streams for auditory processing have also been described (e.g., Rauschecker, 1998; Rauschecker & Tian, 2000), with a ventral pathway into anterior temporal cortex that carries out identification of complex auditory patterns or objects, and a dorsal pathway into posterior parietal cortex that carries out spatial processing (Fig. 26). Hickok and Poeppel (2007) have summarized this dichotomy by linking the ventral pathway with *sound-to-meaning* and the

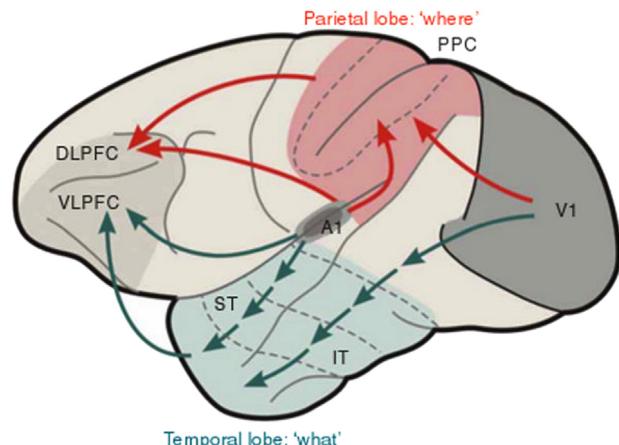


Fig. 26. What and Where processing of auditory information in non-human primates. V1 = primary visual cortex; A1 = primary auditory cortex; IT = inferotemporal cortex; ST = superior temporal cortex; PPC = posterior parietal cortex; VLPFC = ventrolateral prefrontal cortex; DLPFC = dorsolateral prefrontal cortex.
Source: Reprinted with permission from Rauschecker and Scott (2009).

dorsal pathway with *sound-to-action*. In other words, auditory as well as visual processes are organized into computationally complementary cortical processing streams (Section 4).

How complete is the homology that is suggested by the fact that both auditory and visual processing use temporal and parietal cortices? If the homology is strong, what types of auditory properties should we expect from it? In attempting to answer this question, one needs to acknowledge that both auditory and visual processing compute *invariants* to be effective. For example, conscious seeing uses retinotopic surface representations and head-centered shrouds to maintain perceptual stability during eye movements, thereby both stabilizing conscious percepts of visual qualia as the eyes move, and enabling visual recognition

categories to be learned that are significantly invariant under changes in object view, position, and size. What mechanisms maintain perceptual stability of auditory conscious percepts, and what kinds of invariants are learned by auditory recognition categories?

In the auditory domain, any mechanistically satisfying answer to this question would address, at least in part, issues such as: How can a young child babble sounds in one set of acoustic frequencies and use them to learn how to imitate sounds from teachers who speak using different frequencies? How can codes for speech and language meanings be coded in a speaker-invariant and rate-invariant way in order to prevent a combinatorial explosion of stored memories, much as learned visual object recognition codes are invariant under view, position, and size, despite the fact that these speaker and rate invariants are not present in acoustic signals? How can speaker identity be learned as well, so that different voices can be consciously heard and recognized? How do these distinct speaker-invariant and speaker-sensitive representations interact to understand the meaning and identity of an identifiable acoustic source? How can rate-invariant codes for speech be performed and experienced at variable rates that are under volitional control? How can invariant codes be performed as the lyrics of songs with different rhythms and pitches?

The sections below propose answers to these questions that illustrate complementary processes and ART resonances operating at multiple levels of the auditory system, just as they do in the visual system, to support conscious auditory experiences while stabilizing learning of auditory representations along the way. The text outlines an emerging unified neural system that combines modeling discoveries which have been developed to explain distinct databases about audition, speech, and language. In order to achieve this synthesis, new ideas about the mechanistic neural substrates of conscious auditory experiences are proposed, and explanations of many additional data are provided for the first time to support these ideas. A great deal of additional experimental and theoretical work can profitably be done to further develop this theoretical foundation.

14.2. Circular reactions in vision and audition: Babbling to reach and to speak

A helpful way to understand auditory-visual homologs is to consider the perception-action *circular reactions* that occur during auditory and visual development (Piaget, 1945, 1951, 1952). During a visual circular reaction, babies endogenously babble hand/arm movements and reactively foveate their moving hands, thereby learning a map from hand position to eye position and from eye position to hand position. After map learning occurs, looking at a target position can trigger a command, that is under volitional control, for the hand/arm to reach the target. During an auditory circular reaction, babies endogenously babble sounds and learn a map between the auditory feedback from these sounds and the commands that produced them. After a sufficient amount of map learning occurs, a child can use it to imitate sounds from adult speakers, and thereby begin to learn and perform speech and language, again under volitional control.

14.3. Motor-equivalent reaching and speech production: Tool use and coarticulation

Sections 10.3 and 12.5 explained how and why conscious visual surface representations are well-positioned to control scanning eye movements and arm reaches. To what extent do similar conclusions hold for the relationship between auditory consciousness and speech production? In particular, are reaching and speech production controlled by homologous neural systems? We now summarize evidence that the answer is Yes.

During a visual circular reaction, the brain learns to associate hand positions with eye positions and eye positions with hand positions. Visual cues enter the brain through the retina, but the eyes move in the head and the hand/arm system moves in the body. This learning process thus includes learned coordinate transformations from retinotopic visual coordinates to head-centered and body-centered motor coordinates, so that visually seen targets on the retina can command reaches by the hand/arm from the body. Although the mappings that are learned via the circular reaction are derived from hand-arm movements that are triggered by endogenous brain activity during babbling, they can later be used to generate volitional looking and reaching movements, under basal ganglia control, towards valued goal objects.

In particular, when a desired goal object is visually attended, the learned mapping can activate a desired hand/arm *target position vector*, or T , with which to reach it (Fig. 27, left panel). An outflow representation of the current hand/arm position, called the *present position vector*, or P , is subtracted from the target position to compute a *difference vector*, or D (Georgopoulos, Kalaska, Caminiti, & Massey, 1982; Georgopoulos, Schwartz, & Kettner, 1986) that codes the direction and distance that the hand/arm needs to move to reach the target. A volitional *GO signal*, or G , from the basal ganglia (Prescott, Gurney, & Redgrave, 2003) gates, or multiplies, the difference vector to determine which arm will move and the speed with which it will approach the target position. The Vector Integration to Endpoint, or VITE, model (Fig. 27, left panel; Bullock & Grossberg, 1988) modeled these processes to clarify how the Three S's of reaching are carried out: the flexible choice of motor Synergies, and their Synchronous performance at variable Speeds. VITE model computations are supported by anatomical and neurophysiological data about the brain regions that carry them out, and by computer simulations of challenging psychophysical data about reaching behaviors, including changes in velocity profile asymmetry at higher speeds, isotonic arm movement properties before and after de-afferentation, Fitt's law, the 2/3 Power Law of Lacquinti, and velocity amplification during target switching.

The VITE model was just a first step in clarifying the neural mechanisms of arm movement trajectory control. One elaboration of VITE, called the Vector Associative Map, or VAM model (Gaudiano & Grossberg, 1991, 1992) used a circular reaction to show how the T -to- D and P -to- D signals could be rendered dimensionally consistent by mismatch learning that uses D as an error-based teaching signal. Another refinement of VITE showed how arm movements can compensate for variable loads and obstacles, and interpreted the hand/arm trajectory formation stages in terms of identified cells in motor and parietal cortex, whose temporal dynamics during reaching behaviors were simulated (Bullock, Cisek, & Grossberg, 1998; Cisek, Grossberg, & Bullock, 1998).

A refinement that sheds the most light on auditory-visual homologs of reaching and speaking circuits is called the DIRECT model (Bullock, Grossberg, & Guenther, 1993), which also learns through a circular reaction. Learning in the DIRECT model (Fig. 27, right panel) clarifies how a crucial *motor-equivalence* property of arm movement control is achieved; namely, during movement planning, either arm, or even the nose, could be moved to the goal object, depending on which movement system receives a GO signal. This fact implies that trajectory planning does not just combine the position of the target on the retinas with the motorically-computed positions of the eyes in the head, and the head in the body, to compute a body-centered representation of the target position that can command a movement of an arm. Instead, these visual and motor signals are first combined to learn a representation of space around the actor. The position of the target in these *spatial coordinates* can then be downloaded into motor commands for moving the different limbs under volitional control. This spatial representation is computed in

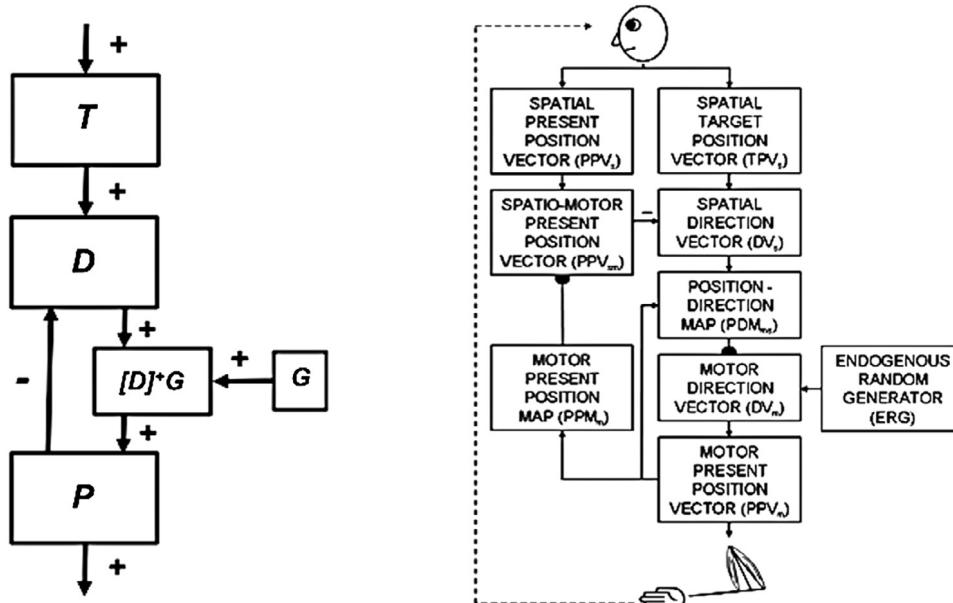


Fig. 27. (Left panel) VITE model circuit for reaching. A present position vector (P) is subtracted from a target position vector (T) to compute a difference vector (D) that represents the distance and direction in which the arm must move. The rectified difference vector ($[D]$) is multiplied by a volitional GO signal (G) before the velocity vector $[DG]$ is integrated by P until P equals T , hence the model name Vector Integration to Endpoint, or VITE [Adapted with permission from Bullock and Grossberg (1988).] (Right panel) DIRECT model circuit. An endogenous random generator (ERG) provides the “energy” during a critical period of motor babbling to activate a motor direction vector (DV_m) that moves the hand/arm via the motor present position vector (PPV_m). As the hand/arm moves, the eyes reactively track the position of the moving hand, giving rise to the visually-activated spatial target position vector (TPV_s) and spatial present position vector (PPV_s). These vectors are used to compute the spatial difference vector (DV_s). It is this spatial transformation, along with the mapping from spatial directions into motor directions, that gives the model its motor-equivalence reaching capabilities. To compute these transformations, the PPV_s first activates the spatio-motor present position vector (PPV_{sm}), which is then subtracted from the TPV_s . As a result, the PPV_s signal that reaches the TPV_s is slightly delayed, thereby enabling the DV_s computation to occur. The PPV_{sm} stage is one of two stages in the model where spatial and motor representations are combined. The subscripts “ s ” and “ m ” denote spatial and motor, respectively. A circular reaction is learned from spatial-to-motor and motor-to-spatial representations at the two adaptive pathways in the model, which are denoted by hemispherical synapses. In particular, the spatial direction vector (DV_s) is adaptively mapped into the motor direction vector (DV_m), thereby carrying out the transformation from visual *Direction Into joint Rotation* that gives the DIRECT model its name. [Reprinted with permission from Bullock et al. (1993).]

parietal cortex. To enable this downloading process, DIRECT, which stands for Direction To Rotation Effector Control Transform, learns to map visual motion *directions* into joint *rotations* during reaching movements, as embodied in the transformation from Spatial Direction Vector to Motor Direction Vector in Fig. 27 (right panel).

Movements other than arm reaching movements also depend upon spatial representations. For example, neurophysiological data (e.g., Andersen et al., 1985, 1987; Bisley, Krishna, & Goldberg, 2004) and neural models (e.g., Gancarz & Grossberg, 1999; Grossberg & Kuperstein, 1986; Silver et al., 2011) about saccadic eye movement control both support the hypothesis that head-centered spatial representations in the parietal cortex help to select movement targets. This is again a subtle issue, as illustrated by the above discussions (see Figs. 21 and 22) of how both retinotopic and spatial representations are needed to move the eyes during learning of invariant recognition categories.

Remarkably, after the DIRECT model learns these representations and transformations, its motor-equivalence properties enable it to manipulate a tool in space. Without measuring tool length or angle with respect to the hand, the model can move the tool’s endpoint to touch the target’s position correctly under visual guidance *on its first try*, without additional learning. In other words, the spatial affordance for tool use, a critical foundation of human societies, follows directly from the brain’s ability to learn a circular reaction for motor-equivalent reaching. We will see below that the corresponding affordance for speaking enables coarticulation of speech sounds to occur.

Because of this visual affordance, a monkey who picks up a stick could direct its endpoint to desired positions in space. If during exploratory movements using the stick, the monkey happened to put the end of the stick into an anthill, and then removed the stick with ants attached, the monkey could then learn this skill

as a way to efficiently eat ants. This skill could not, however, efficiently be learned by other monkeys through imitation of the first monkey’s behavior until social cognitive abilities like *joint attention* developed. Grossberg and Vladusich (2010) have described a role for surface-shroud resonances in the learning of joint attention and imitative behaviors.

This background helps to propose solutions to problems that are relevant to both auditory and visual consciousness: As Fig. 22 illustrates, spatial attention in the visual system helps to control eye and arm movements, while also supporting visual consciousness through surface-shroud resonances. Does their auditory homolog help to support auditory consciousness, keeping in mind the caveat that auditory information enters the brain in head-centered coordinates, and not in retinotopic coordinates? Does auditory spatial attention support conscious auditory states in brain networks that input to motor representations for speech production, as would be expected if Piagetian circular reactions for vision and speech are embodied using homologous brain circuits?

14.4. From eating to speaking

During the development of the DIRECT model by Bullock et al. (1993), an evolutionary rationale was noted for why both the hand/arm and speech articulator systems may use similar, indeed homologous, neural circuits; namely, eating preceded speech during human evolution (MacNeilage, 1998), and skillful eating requires movements that coordinate hand/arm and mouth/throat articulators, including motor-equivalent solutions for reaching and chewing. If the behaviors of reaching and eating are part of an integrated sensory-motor system, then these component processes may be expected to use similar circuit mechanisms.

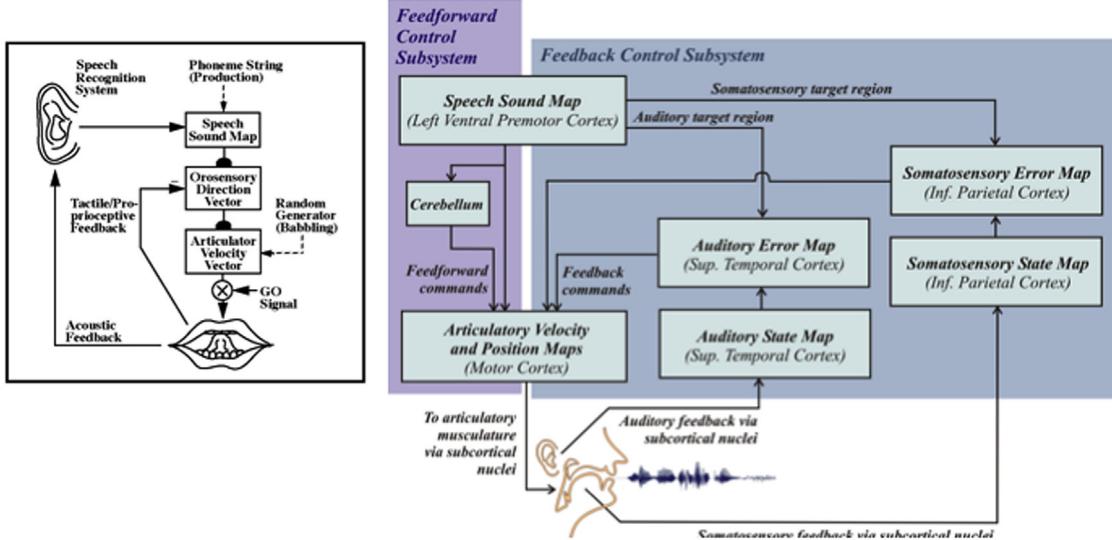


Fig. 28. (Left panel) Original DIVA model circuit. Note the similarity of this circuit to the DIRECT model circuit of Fig. 27 (Right panel) with reach movements replaced by speech articulator movements. [Reprinted with permission from Guenther (1995)] (Right panel) A later version of DIVA in which the processing stages are elaborated and compared with known anatomical brain regions. [Reprinted with permission from Guenther et al. (2006).]

With this insight in mind, Frank Guenther and his colleagues adapted the DIRECT model to develop the DIVA model for motor-equivalent speech production using circuits homologous to DIRECT reaching circuits (Fig. 28, left panel). DIVA stands for Directions Into Velocities of Articulators, which is the analog for speaking of the direction-into-rotation transform that DIRECT models for reaching. DIVA showed how babbled speech articulator movements may be used to learn volitionally controlled speech sounds that are capable of motor-equivalent articulator movements such as those that occur during coarticulation (Guenther, 1995; Guenther, Hampson, & Johnson, 1998).

As with the VITE and DIRECT models, the DIVA model has been progressively developed, leading to an anatomical interpretation of model computational stages in terms of identified anatomical regions and their functions (Fig. 28, right panel; e.g., Guenther, Ghosh, & Tourville, 2006). With this background as motivation, homologs between auditory and visual brain processes underlying consciousness can be discussed, including neural mechanisms and psychological competences that go beyond the explanatory range of DIRECT and DIVA.

14.5. Auditory neglect and speech production deficits

The following data support the hypothesis that visual and auditory consciousness arise from similar circuit designs, modulo their different learned invariants. As in the case of visual neglect (Section 10.5), auditory neglect can be produced by lesions of the IPL, particularly in the right hemisphere (e.g., Bellmann et al., 2001; Clarke & At, 2013; Driver & Mattingley, 1998; Gutschalk & Dykstra, 2015; Marshall, 2001). Deficits in perception of auditory space can be severe, including a spatial bias to the ipsilesional side in sound localization, and marked left ear perceptual extinction during dichotic listening tasks with pairs of simultaneous disyllabic words. This deficit includes a systematic bias in auditory space representation with respect to the body. It also often includes impairment in maintaining sustained attention (Robertson et al., 1997), as reflected by impaired auditory target detection during continuous stimulation. This link between auditory space representation and sustained attention is reminiscent of how surface-shroud resonances sustain spatial attention in the visual system.

In addition, the DIVA model anatomy of Fig. 28 (right panel) implies that damage to the inferior parietal cortex could disrupt

or destroy its Somatosensory State Map and/or Somatosensory Error map, and thereby lead to problems in learning correct speech productions of articulatory velocity and position maps in the Motor Cortex. Thus, just as damage to visual representations in the inferior parietal cortex can undermine surface-shroud resonances that can lead to problems in eye movement control and reaching, damage to auditory representations in the inferior parietal cortex can lead to problems in speech production.

Discussions of lesions in auditory spatial representations and how they may lead to neglect and deficits in sustained attention and production will continue in the next section as we continue to explicate the complementary global organization of auditory processing into What and Where/How cortical streams in human and monkey brains. These results will show that ART resonances exist at multiple levels of the auditory system, where they enable fast learning and stable memory of several different types of auditory recognition categories.

15. Stream-shroud resonances for conscious hearing of visual qualia

15.1. Auditory scene analysis: Tracking sound sources through noise and overlapping frequencies

Auditory resonances occur in cortical regions that support conscious percepts of auditory sounds as parts of auditory streams. The process that forms auditory streams is called *auditory scene analysis* (Bregman, 1990). Auditory scene analysis enables the brain to separate multiple sound sources, such as voices or instruments, into trackable auditory streams, even when these sources may contain harmonics that overlap and are degraded by environmental noise.

Auditory scene analysis hereby helps to solve the so-called *cocktail party problem*, which occurs when listening to a friend at a noisy cocktail party or other social occasion. By separating overlapping frequencies that are due to different acoustic sources into distinct streams, higher auditory processes, such as those involved in speech and language, can learn to recognize the meaning of these streams. This separation process also endows these streams with the coherence that is needed to bridge across spectrally ambiguous durations, as when noise occludes a speaker's acoustic signals.

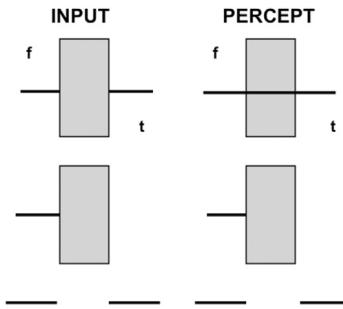


Fig. 29. Auditory continuity illusion. (Top row) When a tone occurs before and after broad-band noise (left column), it may be consciously perceived as continuing through the noise (right column) if the noise does not last too long. (Middle row) When the second tone is removed, this does not occur. (Third row) When the noise is removed, this again does not occur. These properties are consistent with the amount of time that it takes a resonance to develop, and with how the ART Matching Rule uses top-down attentional matching to select consistent feature patterns as the resonance develops.

The formation of multiple auditory streams may be compared with the process of 3D figure-ground separation during vision. As noted in Section 10, visual figure-ground separation seems to be completed in cortical area V4, from which its 3D surface representations can resonate in surface-shroud resonances that can reach conscious awareness. The analog of a surface-shroud resonance in the auditory system is proposed to be a *stream-shroud resonance* (Table 2). Before discussing stream-shroud resonances, it is necessary to say more about how ART mechanisms form auditory streams.

15.2. Auditory continuity illusion and spectral-pitch resonances

The *auditory continuity illusion* (Bregman, 1990) illustrates ART properties during auditory streaming. This illusion occurs, for example, when a steady tone occurs both before and after a burst of broadband noise (Fig. 29, top row, left column). Under appropriate temporal and amplitude conditions, a percept is consciously heard in which the tone continues through the noise (Fig. 29, top row, right column), and the residual noise sounds like it occurs in a different stream. If a second tone does not follow the noise burst (Fig. 29, middle row, left column), then the tone is not heard continuing through the noise (Fig. 29, middle row, right column).

Comparing these two cases in Fig. 29 shows that, until the second tone occurs, they are identical. Somehow the second tone operates “backwards-in-time” to create a conscious percept of the first tone continuing through the noise. Without the second tone, this “backwards-in-time” effect does not occur.

Noise is critical for this percept to occur. If two tones are played before and after a period of silence, with no noise between them (Fig. 29, bottom row, left column), then silence is heard (Fig. 29, bottom row, right column).

These are properties of the ART Matching Rule (Fig. 12) acting through time at the level of auditory scene analysis, and modeled by the ARTSTREAM model (Fig. 30; Grossberg, Govindarajan, Wyse, & Cohen, 2004). The tone continues through the noise when the noise is followed by a subsequent tone, but not otherwise, because of how such a resonance develops through time between a feature level that represents the tone's auditory frequency spectrum (the spectral stream layer in Fig. 30), and a category level that selectively responds to the pitch of this spectrum (the pitch stream layer in Fig. 30). It takes awhile for the following events to occur: First, a pitch category is activated in response to the tone input at the frequency spectrum level. Second, the activated pitch category feeds back a top-down expectation that amplifies the frequency of the tone input, while inhibiting other frequencies, using the ART Matching Rule. Third, this bottom-up/top-down interaction cycles until the mutually reinforcing signals exceed a resonance

threshold. Then a *spectral-pitch resonance* is generated that can support a conscious percept of the auditory stream.

How does a second tone create the percept of a tone continuing through the noise? Once the spectral pitch resonance is underway, it takes much less time for the second tone to keep the resonant activity above threshold, thereby enabling its top-down expectation to continue selecting the tone from the noise, and maintaining a percept of tone continuing through noise. For this “backwards-in-time” effect to occur, the noise duration cannot be too long, or else the second tone will not arrive until after the noise interval is consciously perceived. This duration can be used to estimate the amount of time needed to trigger a spectral-pitch resonance.

The fact that a tone followed by noise, with no subsequent tone, sounds just like a tone followed by noise is due to the fact that there is no backwards-in-time boost of the resonance due to the occurrence of a second tone.

The fact that the tone is not completed over silence is explained by the fact that top-down expectations that obey the ART Matching Rule are modulatory and cannot create a conscious experience when there is no bottom-up signal.

One implication of this explanation is that conscious percepts can lag in time behind their generative stimuli. Many examples of such lags exist, including classical data of Libet, Gleason, Wright, and Pearl (1983) on the cortical readiness potential.

15.3. ARTSTREAM model and stable learning of pitch, timbre, and voice categories

The ART Matching Rule helps to dynamically stabilize learned memories. What is learned during an auditory scene analysis process like the auditory continuity illusion? The ARTSTREAM model of auditory streaming (Fig. 30), including its SPINET model front end (the peripheral processing layers plus a single spectral stream layer and the corresponding pitch summation layer in Fig. 30; Cohen, Grossberg, & Wyse, 1995), proposes that spectral-pitch resonances occur between logarithmic spectral representations of sound frequency patterns and recognition categories that code for pitch. Each pitch category can categorize the sound frequency harmonics corresponding to the pitch of a single acoustic source. Multiple streams (within the parallel copies 1, 2, 3, . . . of the spectral stream layer in Fig. 30) can each activate their own pitch category (within the parallel copies 1, 2, 3, . . . of the pitch summation layer in Fig. 30).

Pitch categories must be learned. ART proposes that this occurs using spectral-pitch resonances that enable the learned pitch categories to solve the stability-plasticity dilemma and thereby dynamically stabilize themselves via the ART Matching Rule.

Not only pitch categories can be learned by this resonant process. Any acoustic frequency patterns that occur with sufficient statistical consistency through time can be learned. Pitch categories compress relatively simple, harmonically-related spectral frequencies within a given time interval. They are a type of *item category*, or *item chunk* (see Section 16). Timbres have more complex spatio-temporal spectral distributions that can also include frequency sweeps, different onset and offset times, and different relative amplitudes in different frequency bands (Grey, 1977; McAdams, 2013). A timbre may be categorized as a *list category*, or *list chunk*, of sequences of sound-selective item chunks that are temporarily stored in a *working memory* (see Section 16). The characteristic sounds of voices and instruments (e.g., Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993) may be stored and categorized in this way, although their different spectral properties may cause their adaptive filters to selectively activate different cortical areas (e.g., Belin et al., 2000; Fecteau, Armony, Joanette, & Belin, 2004). Such a working memory is called a *spectral working memory* to emphasize that it supports the recognition of spectral characteristics of acoustic sources.

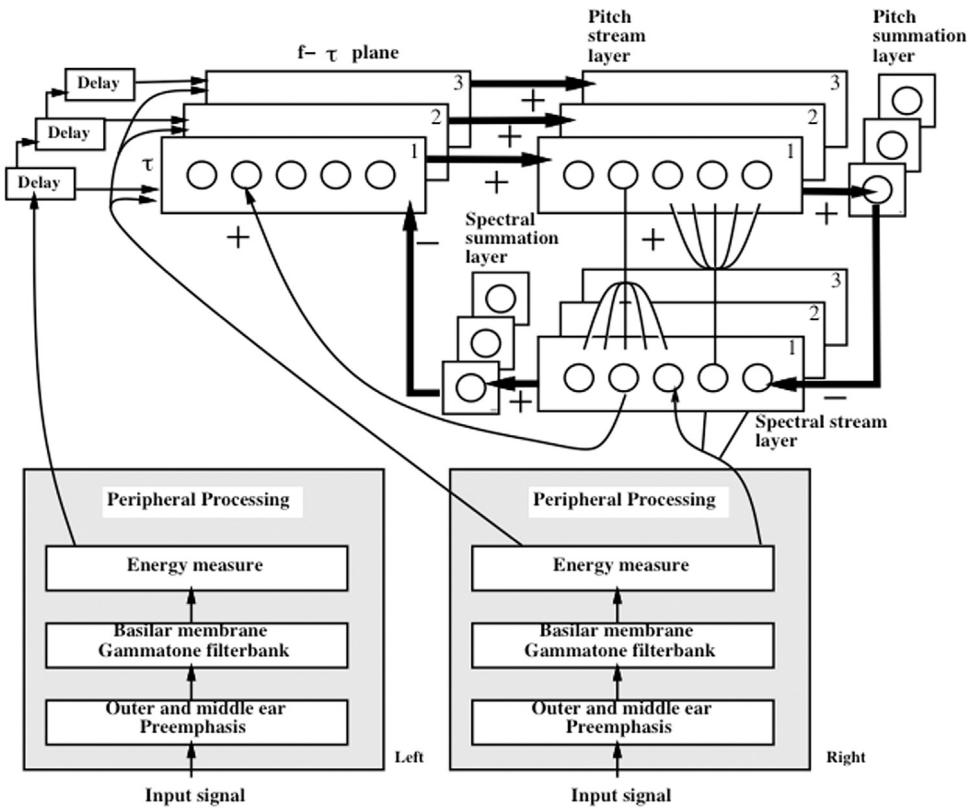


Fig. 30. ARTSTREAM model macrocircuit. This circuit can create multiple auditory streams in response to multiple acoustic sources. After preprocessing by the SPINET model circuitry (left and right eye regions at the figure bottom; Cohen et al., 1995), a log frequency representation of the sound activates the spectral stream layer. Each frequency is redundantly represented in a strip, or “hypercolumn”, of cells. Multiple auditory streams can develop across these strips due to asymmetric competition across the various copies of the frequency spectrum and pitch category layers. The spectral stream layer sends bottom-up signals via an adaptive filter to the pitch stream layer. This filter acts like a “harmonic sieve” that activates a pitch category which selectively responds to the harmonics of a sound at the spectral stream layer. The pitch category, in turn, reads out a top-down expectation which selects harmonics that are consistent with it. The top-down expectation embodies the ART Matching Rule: the top-down excitatory pathway (+) represents the on-center prototype of the expectation that is rendered modulatory by the top-down inhibitory nonspecific pathway (−) that also inhibits other frequencies. When the bottom-up and top-down pathways approximately match, a spectral-pitch resonance can develop between the frequency spectrum and pitch category layers. Such a resonance can support conscious recognition of the sound. The left and right eye inputs also compute ITD positional estimates (see text) that enable a resonance to multiplex both pitch and positional representations of an auditory source.
Source: Reprinted with permission from (Grossberg et al., 2004).

Multiple working memories and list chunking networks exist in the What and Where/How auditory cortex, some devoted to representing frequency-sensitive and rate-sensitive information like voices and instruments, using categories that learn to identify a source, and others devoted to representing processes that are frequency-invariant and rate-invariant, using categories that learn to recognize speech and language meanings. In the theory presented in Section 16 and beyond, all these networks obey similar design principles that are embodied by specializations of similar types of circuits, and each can categorize different types of acoustic information rapidly and stably through time.

When the totality of these working memories resonates with the logarithmic representations of sound frequency patterns, this resonance is called a *spectral-pitch-and-timbre resonance* to emphasize that acoustical categories other than pitch categories can drive it. A spectral-pitch-and-timbre resonance can support conscious recognition of a complex sound in an auditory stream (Table 2).

15.4. Stream-shroud resonances: Spatial attention and neglect in the auditory stream

A surface-shroud resonance has been proposed to support percepts of conscious visual qualia (Section 10). As noted in Section 15.1, an analogous process of *stream-shroud resonance* is proposed to support percepts of conscious auditory qualia. Data

were reviewed in Section 10.5 to support the hypothesis that a surface-shroud resonance enables sustained spatial attention to focus on a visual object and generate its conscious qualia, and that there is a close link between problems with sustained visual spatial attention and unilateral visual neglect. A stream-shroud resonance is proposed to play a similar role in sustaining auditory spatial attention and conscious auditory quality, and problems with sustained auditory attention, say due to a parietal lesion, can cause unilateral auditory neglect (e.g., Robertson et al., 1997).

An auditory shroud cannot represent *spatial* attention unless it can represent the attended position of a sound in space. The position of auditory sources is computed from interaural time differences (ITD) and interaural level differences (e.g., Bronkhorst & Plomp, 1988; Darwin & Hukin, 1999; Rayleigh, 1907). As a result, auditory spatial attention in humans does not appear to have a map structure (Kong et al., 2012), but rather seems to be embodied by an opponent process whose neurons are tuned to (e.g.) ITDs (Magezi & Krumbholz, 2010). The ARTSTREAM model in Fig. 30 proposes how both pitch and ITD spatial position information can cooperate during auditory scene analysis (cf. data in Bregman, 1990). In this way, an auditory stream can resonate in a spectral-pitch-and-timbre resonance for auditory recognition, while a stream-shroud resonance supports conscious hearing, and both resonances synchronize via their shared spectral representations, analogous to what happens during vision (Section 13), along the way explaining data like the Deutsch (1975) scale illusion. Further

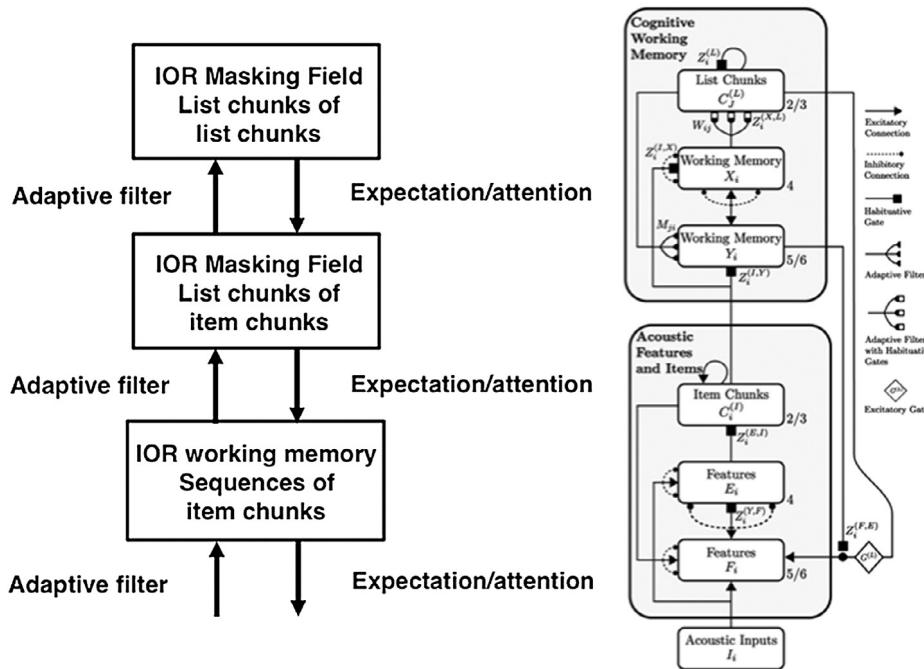


Fig. 31. (Left panel) Variant of the ARTWORD model macrocircuit. Interactions among three speech processing levels are capable of working memory storage, chunking, and performance of word sequences. Each level consists of an Item-Order-Rank working memory that can store sequences with repeated items. The second and third levels are, in addition, multiple-scale Masking Fields that enable selection and storage of sequences of variable length. All the levels are connected by ART bottom-up adaptive filters and top-down learned expectations and their attentional focusing capabilities. In addition, these bottom-up and top-down pathways experience activity-dependent habituation. The first level stores sequences of item chunks. Its inputs to the second level enable that level to store list chunks that can selectively categorize item chunk sequences. The inputs from the second level to the third level enable it to store list chunks of list chunks, in particular sequences of words that can include repeated words. (Right panel) cARTWORD model circuit. cARTWORD includes a hierarchy of two cortical processing levels that model different cortical regions. Each level is organized into laminar cortical circuits that share a similar laminar organization. Cells in both levels model layers 5/6, 4, and 2/3 with a similar distribution of inter-laminar connections. In both levels, deep layers (6 and 4) are responsible for processing and storing inputs via feedback signals between them. Superficial layers (2/3) respond to signals from layer 4 to categorize, or chunk, distributed patterns across these deeper layers into unitized representations. The first level processes acoustic features and item chunks. The second level stores sequences of acoustic items in an Item-and-Order working memory. The stored sequences send signals via a bottom-up adaptive filter to a Masking Field that chooses the list chunks that best represent the current stored sequence of item chunks. The multiple-scale, self-similar, and shunting on-center off-surround network of the Masking Field enables its list chunks to selectively represent sequences (e.g., words) of multiple lengths. Top-down connections exist both within and between levels. Intra-level connections enable item chunks in layer 2/3 of the first level to send top-down attentional matching signals to their distributed features in layer 5/6, and list chunks in layer 2/3 of the second level to send top-down signals to their working memory item chunks in layer 5/6. Both types of signals can modulate, but not fire, their target cells when acting alone. Inter-level top-down signals are the ones that can trigger resonance. They occur from list chunks in layer 2/3 of the second level to a basal ganglia gate (diamond), and from stored item chunks in layer 5/6 of the second level to the features in layer 5/6 of the first level. The basal ganglia gate opens when a list chunk in layer 2/3 of the second level is chosen in response to a sequence of item chunks in level 4 of the cognitive working memory. Once the gate opens, top-down feedback from the cognitive working memory in layer 5/6 of the second level can trigger a system-wide item-chunk resonance that supports conscious recognition of the sequence through time.

Source: Reprinted with permission from Grossberg and Kazerounian (2011).

experimental and theoretical analysis of how these two types of resonances interact is much to be desired.

16. Item-list resonances for recognition of speech and language

16.1. From streaming to speech: Item-and-Order working memories and list chunks

ART mechanisms also operate at higher cortical levels of the auditory system, where they help to clarify how speech and language are rapidly learned, stably remembered, flexibly performed, and consciously heard. These ART models have been developed as part of an emerging neural theory of speech learning, perception, recognition, and recall whose various increasingly developed models are called PHONET, ARTPHONE, ARTWORD (Fig. 31, left panel), and conscious ARTWORD, or cARTWORD, (Fig. 31, right panel). This theory proposes how sequences of acoustic *item chunks* are stored temporarily in working memory before being categorized by *list chunks*. The term *working memory* is here used to mean the temporary storage of sequences of events through time, not just persistent storage of a single event. The subsequent text summarizes enough information about these models to clarify how they interact with the stream-shroud circuits

that determine conscious hearing, and the spectral-pitch-and-timbre resonances that determine the streams upon which these higher-order cortical processes build.

Grossberg (1978a, 1978b) introduced a neural model of working memory upon which the more recent models listed above consistently built. This *Item-and-Order* working memory model posits that a temporal stream of inputs is stored through time as an evolving spatial pattern of activity across a network of content-addressable item representations. Inputs to this working memory are unitized item chunks of individual sounds. An item chunk is a learned recognition category that selectively responds to a spatial pattern of activation across a network of auditory feature detectors during a brief time interval.

Active item chunks input to the next processing stage, where sequences of them are stored in real time within a working memory. This temporal-to-spatial transformation converts the input sequence into a temporally evolving spatial pattern of activity, or *activity gradient*, across item-selective cells (Fig. 32). The *relative activities* of different cell populations code the temporal order in which the items will be recalled, with the largest activities recalled earliest; hence, the name *Item-and-Order* working memory for this class of models.

A *primacy gradient*, in which the first item stored has the largest activity, the second the next largest activity, and so on, can be

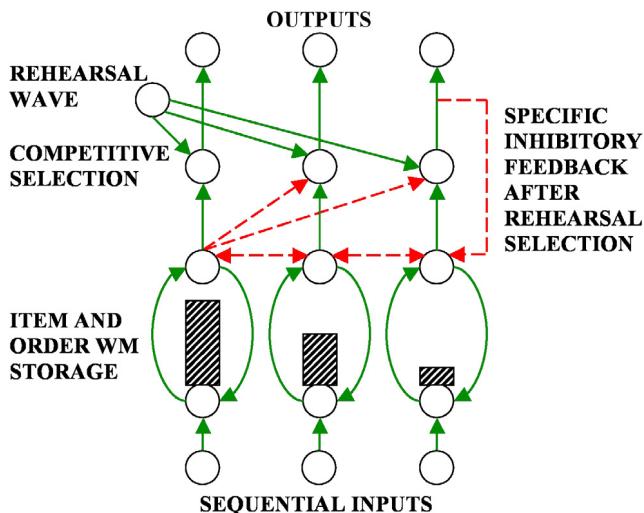


Fig. 32. A temporal sequence of inputs creates an evolving spatial pattern of activity across item chunks in an Item-and-Order working memory, where the height of each hatched rectangle is proportional to each cell's current activity. Relative activity level codes for item and order. A rehearsal wave allows item activations to compete before the maximally active item elicits an output signal and self-inhibits via feedback inhibition to prevent its perseverative performance. The process then repeats itself. Solid arrows denote excitatory connections. Dashed arrows denote inhibitory connections.

Source: Adapted from Grossberg (1978a).

recalled in the correct order. Recall occurs when a nonspecific rehearsal wave opens all the output gates of the network, thereby enabling the most active item chunk to be read-out first (Fig. 32). As it is read out, it self-inhibits its stored activity, thereby enabling the next item chunk to be read out, and so on until the entire stored list is performed.

A more recent name for this class of models is *competitive queuing* (Houghton, 1990). When an Item-and-Order working memory can store repeated items in a sequence, it is called an *Item–Order–Rank* working memory (Bradski, Carpenter, & Grossberg, 1994; Grossberg & Pearson, 2008; Silver et al., 2011). Item–Order–Rank working memories, including the role of primacy gradients in correct list storage and recall, are supported by psychophysical and neurophysiological data (e.g., Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Farrell & Lewandowsky, 2002; Grossberg & Pearson, 2008; Page & Norris, 1998; Silver et al., 2011).

These stored working memory sequences are unitized through learning into list chunk, or sequence category, representations at the next processing level. List chunks can be selectively activated by different item sequences during speech perception and word recognition (Boardman et al., 1999; Grossberg et al., 1997; Grossberg & Kazerounian, 2011, 2016; Grossberg & Myers, 2000; Grossberg & Stone, 1986b; Kazerounian & Grossberg, 2014).

16.2. Top-down attentive matching and item-list resonances

Top-down expectation signals from the list chunks are activated as the items are being stored in working memory through time. These top-down expectations satisfy the ART Matching Rule, and thereby select consistent item sequences, while inhibiting inconsistent ones, to create a temporally-evolving attentional focus. As the cycle between bottom-up activation of list chunks and top-down attentive matching of item sequences continues through time, an *item-list resonance* can develop whose list chunks represent the best prediction of what item chunk sequences are attended at any given time within the working memory. Such an item-list resonance is proposed to support conscious recognition of attended speech and language representations (Table 2) when it interacts with lower-level resonances, such as stream-shroud and spectral-pitch-and-timbre resonances.

16.3. Phonemic restoration: How the future can influence what is consciously heard in the past

Phonemic restoration (Warren & Sherman, 1974) illustrates the operation of ART mechanisms during speech perception in much the same way as the auditory continuity illusion represents them during auditory streaming. Suppose that broadband noise replaces the phonemes /v/ and /b/ in the words “delivery” and “deliberation”, respectively. Despite the identical initial portion of these words (“deli-”), if the broadband noise is immediately followed by “ery” or “eration”, listeners hear the /v/ or /b/ as being fully intact. However, if the noise is replaced by silence, then restoration does not occur.

These experimental properties of phonemic restoration match properties of top-down matching by the ART Matching Rule (Fig. 12). Thus, phonemic restoration properties illustrate attentive matching processes that enable speech and language to be learned quickly, without risking catastrophic forgetting. In the case of speech perception, the categories that are learned represent phonemes, syllables, and words, rather than the pitch and timbre categories that are learned during auditory streaming.

Phonemic restoration also illustrates temporal properties of a consciously heard and recognized resonance that emerges through time as items are temporarily stored in working memory (see Section 16.4 for a discussion of how working memories are designed). In particular, the noise in “deli-noise-[ery/eration]” is not consciously heard and recognized before the last portion of the word occurs because, if the resonance has not developed fully before the last portion of the word is presented, then this portion can influence the expectations that determine the conscious percept and meaning. As this resonance unfolds, its expectations convert the noise in “deli-noise-[ery/eration]” into a percept of [/v/-/b/], while suppressing spectral components of the noise that are not part of the expected consonant sound. A variation of this experiment in which a reduced set of spectral components is used in the noise leads to hearing a correspondingly degraded consonant sound (Samuel, 1981a, 1981b).

As in the case of the auditory continuity illusion, future events can influence past events without smearing over all the events that intervene. Again this is due to the fact that the top-down attentive matching process is modulatory. The opposite concern is also of importance: How can sharp word boundaries be perceived even if the sound spectrum that represents the words exhibits no silent intervals between them? ART proposes that “conscious speech is a resonant wave” and that “silence is a discontinuity in the rate at which the resonant wave evolves”. The top-down matching process coherently groups familiar linguistic units together through time, even if they are not separated by silent intervals in the input stream.

The auditory continuity illusion and phonemic restoration support the hypothesis that multiple levels of auditory processing use resonant dynamics to coordinate fast learning, self-stabilizing memory, and conscious awareness of different types of auditory information.

16.4. A shared design for linguistic, motor, and spatial working memories: LTM Invariance and Normalization

What properties recommend Item-and-Order working memories more than other possible alternatives, over and beyond their broad explanatory and predictive range? Grossberg (2003b, 2013b), Grossberg and Pearson (2008), and Grossberg and Kazerounian (2016) review the hypothesis, first proposed in Grossberg (1978a), that Item-and-Order working memories satisfy two postulates which ensure that speech and language can be learned in a stable way through time: the LTM Invariance Principle and the Normalization Rule. The LTM Invariance Principle guarantees, for

example, that the first time a novel word, such as MYSELF, is stored in working memory, it does not destabilize previously learned list chunks that code for its familiar subwords MY, ELF, and SELF. Without such a property, longer chunks (e.g., for MYSELF) could not be stored in working memory without risking the catastrophic forgetting of previously learned memories of shorter chunks (e.g., for MY, SELF, and ELF). Language, motor, and spatial sequential skills would then be impossible.

The LTM Invariance Principle is achieved by preserving the *relative activities*, or ratios, between previously stored working memory activities as new items are presented through time. Newly arriving inputs may, however, alter the *total activity* of each active cell across the working memory. How does preserving activity ratios in an Item–Order–Rank working memory stabilize previously learned categories? These activities send signals to the next processing stage, where the category cells are activated. The signals are multiplied by adaptive weights, or LTM traces, before the net signals activate their target categories (Figs. 1 and 13). The total input to a category thus multiplies a *pattern*, or vector, of activities times a *pattern*, or vector, of LTM traces. By preserving relative activities, the relative sizes of these total inputs to the category cells do not change through time, and thus nor do the corresponding LTM patterns that track these activities when learning occurs at their category cells.

The Normalization Rule insists that a working memory's maximum total activity tends to be independent of the number of stored items. Thus, if more items are stored, then each item tends to be stored with less activity. Normalization implies the *limited capacity* of working memory by redistributing, rather than simply adding, activity when new items are stored.

Both postulates are automatically satisfied by a ubiquitously occurring neural design, thereby clarifying how a working memory could arise through evolution. This design is a recurrent on-center off-surround network (Fig. 32) whose cells obey shunting dynamics. Such network occurs ubiquitously (cf. Section 10.7) because they enable the brain to process and store distributed patterns of inputs without being degraded by noise – when their inputs are small – or saturation—when their inputs are large, thereby solving the *noise-saturation dilemma* that is faced by every brain network (Grossberg, 1973, 1980).

As noted above, when item chunk sequences that are stored in such a recurrent network are read-out by a volitionally-controlled rehearsal wave, the most active item representation is performed first (Fig. 32). As each item is read out, it self-inhibits its working memory representation, thereby causing inhibition-of-return. Then the next most active item is read out, self-inhibits, and so on. Because the working memory is defined by a recurrent *shunting* network, it tends to contrast-normalize its stored activities and to conserve its total activity through time, even after its largest activities are read out. The tendency to conserve total activity automatically achieves the Normalization Rule. These properties of competitive normalization, rehearsal waves, and self-inhibitory feedback have all been supported by psychological and neurophysiological data, as well as by model simulations (e.g., Boardman & Bullock, 1991; Grossberg & Pearson, 2008; Silver et al., 2011).

The LTM Invariance Principle and Normalization Rule imply that there is an intimate connection between the process of storing sequences temporarily in working memory and the learning of list chunks. Agam, Galperin, Gold, and Sekuler (2007) reported data consistent with the formation of list chunks as movement sequences are practiced, thereby supporting the prediction that working memory networks are designed to interact closely with list chunking networks. Various psychophysical experiments have also successfully tested this prediction during speech perception (e.g., Auer & Luce, 2008; Goldinger & Azuma, 2003; Luce &

McLennan, 2008; McLennan & Luce, 2005; McLennan, Luce, & Charles-Luce, 2003; Vitevitch & Luce, 1999).

The LTM Invariance Principle and Normalization Rule also imply that all linguistic, spatial, and motor working memories share a similar network design, and thus should exhibit similar data properties, such as error distributions. Jones, Farrand, Stuart, and Morris (1995) reported similar performance characteristics to those of verbal working memory for a spatial serial recall task in which visual locations were remembered. Agam, Bullock, and Sekuler (2005) reported psychophysical evidence of Item-and-Order working memory properties in humans as they performed sequential copying movements, and Averbeck et al. (2002); Averbeck, Crowe, Chafee, and Georgopoulos (2003a, 2003b) reported neurophysiological evidence for such a working memory in monkeys during performance of sequential copying movements.

Because of this shared design, it becomes easier to understand how language in young children can begin to develop in a way that parallels the motor behaviors of adult teachers during mutual play (Bruner, 1975), or how sign language by hearing adults can coordinate signing with speaking (Neville et al., 2002). Both of these activities need much more model development to be fully understood.

16.5. Phonemic restoration in a laminar cortical model of speech perception

The cARTWORD model (Fig. 31, right panel; Grossberg & Kazerounian, 2011, 2016; Kazerounian & Grossberg, 2014) is the most recent example of an ART model of speech perception. It provides the only available neural model that simulates how, during phonemic restoration, future acoustic context can disambiguate noise-occluded speech sounds that occurred earlier, yet generate a completed percept that is consciously heard in the correct temporal order. cARTWORD clarifies how a hierarchy of coordinated resonances can generate this conscious percept. This is accomplished within *laminar* cortical circuits that, remarkably, are specializations of laminar cortical circuits that have also been used to model 3D vision and figure-ground perception in the 3D LAMINART model (e.g., Section 6).

The cARTWORD model proposes how active list chunks, which represent the most predictive representation of the current linguistic context, can open basal ganglia gates that release a multi-level resonance which includes acoustic features, item chunks, working memories, and list chunks. The unfolding through time of this multi-level resonance simulates the consciously recognized, phonemically restored, speech representations progressing in their correct temporal order from past to future. Future work on cARTWORD should focus on, among other improvements, including more realistic auditory features.

In summary, figure-ground separation in vision may be compared with streaming in audition and speech perception. If these visual and auditory processes have homologous brain designs, then the auditory analog of V4 in visual surface representation may be sought by future research in the anatomical substrate of auditory streaming. A spectral-pitch resonance has some of the required properties to support conscious percepts of auditory streams, but the resonances that support conscious hearing cannot be restricted only to the sustained sounds that support pitch percepts. As explained in Section 17, both transient and sustained feature detectors are needed to consciously hear, understand, and perform the sounds of the world.

17. Learning and storage of speaker-invariant and rate-invariant working memories

Multiple working memories and list chunking networks in the What and Where/How auditory cortices are used to encode

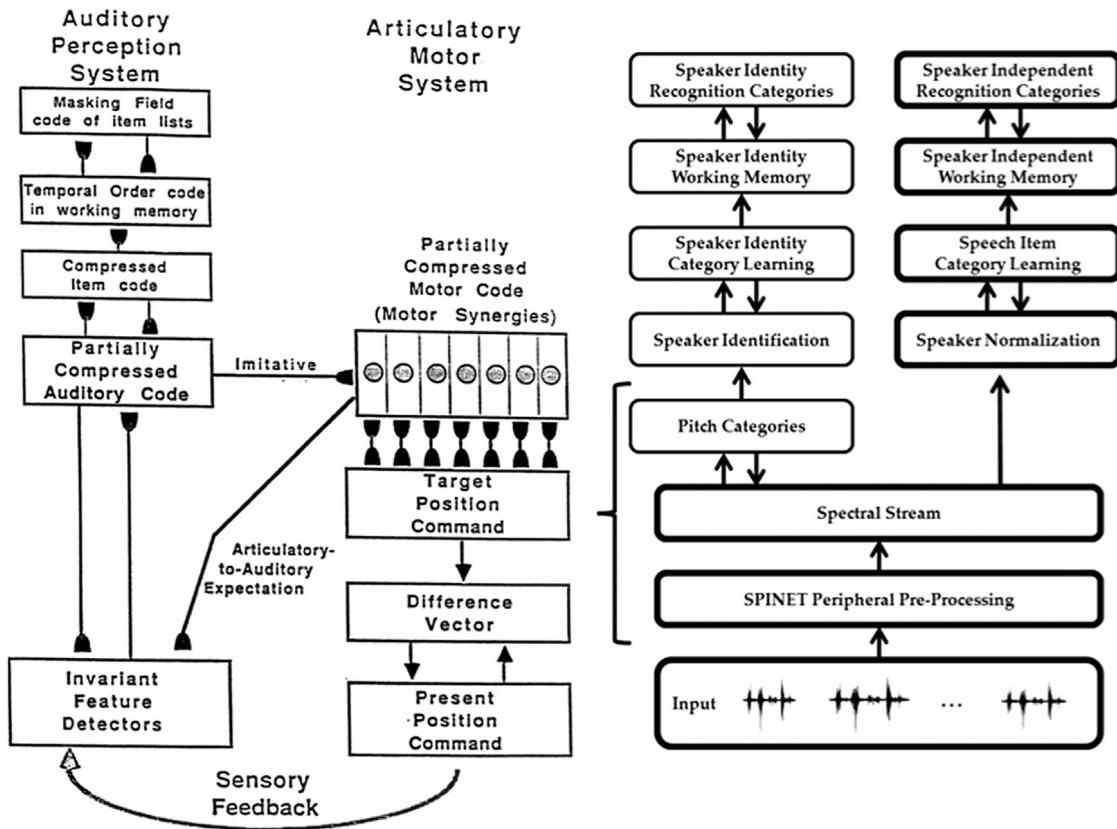


Fig. 33. (Left panel) A perception–production model macrocircuit in which an endogenous random generator (ERG) can cause motor synergies to generate babbled speech sounds. The auditory feedback from these self-generated sounds can be transformed into a partially compressed auditory code by an ART circuit before being associatively mapped into the motor synergy representations via an imitative map, at the same time that item chunk, working memory, and list chunk representations are stored and learned. When a different speaker is heard, these sounds are speaker-normalized, leading to activation of invariant feature detectors. The pattern of activation across these feature detectors can generate a novel combination of activation at motor synergies with which to imitate the novel sound. Additional processing stages (not shown) enable imitative speech errors to be corrected (say as in the DIVA model of Fig. 28). Volitional speech productions, including learned sequences of words (see Fig. 31, left panel) can be read out while top-down expectations are matched against the produced sounds, generating an item-list resonance that supports conscious recognition when the match is good enough (see Fig. 31, right panel). [Reproduced with permission from Cohen et al. (1988).] (Right panel) ARTSPEECH model macrocircuit. This circuit schematizes how speaker identification representations (left cortical stream) can be learned in parallel with speaker-normalized speech and language meaning representations (right cortical stream) using similar hierarchies of item chunk, working memory, and list chunk representations. The bracket designates processing stages of the ARTSTREAM model (Fig. 30). [Reprinted with permission from Ames and Grossberg (2008).]

sounds. Some of these working memories and list chunking networks embody frequency- and rate-sensitive processes like voice, instrument, and speaker identification, whereas others embody frequency- and rate-invariant processes, such as speech and language recognition and meaning. Section 16.4 noted that these working memory and list chunking networks obey similar design principles, notably the LTM Invariance Principle and Normalization Rule, that are realized by similar types of circuits, in order to rapidly learn and stably remember different kinds of acoustic information through time. Sections 15 and 16 illustrated, however, that different processing levels process different types of information. Below are summarized key pre-processing stages that are needed to more fully understand the types of conscious experiences of hearing and knowing that these sections outlined. In particular, how does the brain learn and understand the speech and language of multiple speakers as they speak at variable rates (speaker-independence and rate-invariance), while also being able to consciously hear and identify the voices of individual speakers (speaker-dependence and rate-sensitivity)?

17.1. Speaker normalization in the What stream: Circular reactions and imitative maps

Speech signals are speaker-dependent, but What stream representations of language meaning are speaker-independent. The

transformation from speaker-dependent to speaker-independent language representations enables speech to be learned and understood from different speakers. This process is called *speaker normalization*.

Speaker normalization allows language learning to get started. As noted in Section 14.2, a circular reaction enables a baby to learn a map between the auditory feedback from its own babbled sounds to the motor commands that caused them. For the baby to imitate sounds of adult teachers who speak in a different frequency range, both the baby's sounds and those of adult teachers need to be filtered by a network that normalizes them into a common frequency range. After speaker normalization is achieved, an *imitative map* can be learned from the baby's speaker-normalized babbled sounds to the motor commands that produced them. Subsequently, the speaker normalization processing stage enables the imitative map to filter heard adult sounds into novel speech productions using the learned connections from the normalized babbled sounds to their motor commands. These imitated sounds can then be refined by speech and language learning experiences throughout life.

Learning in the imitative map remains active for many years to enable an individual's changing voice through puberty and adulthood to update the map. This fact raises two design problems: What process dynamically stabilizes the map learning process, while also allowing the map to undergo controlled changes? What process ensures dimensional consistency of the auditory speech

representations and motor speech commands that are linked by the map, so that the map can, in fact, be learned?

The ART architecture in Fig. 33 (left panel) incorporates a solution to both problems. In it, babbled sounds are generated by a VITE circuit (Fig. 27, left panel) when it is activated by an Endogenous Random Generator, or ERG, during a speech circular reaction. These babbled sounds create auditory feedback that activate acoustic feature detectors. At the time that this circuit was published (Cohen, Grossberg, & Stork, 1988), it was not yet understood how speaker normalization might occur, so these acoustic features were simply said to be “invariant”;—that is, speaker-normalized—at the model’s Invariant Feature Detectors level. These speaker-normalized feature patterns are partially compressed into multiple auditory item chunks (at the model’s Partially Compressed Auditory Code level), which are then associated through learning via the imitative map with the still-active motor synergy commands that produced the babbled sounds (at the model’s Partially Compressed Motor Code level). This learned auditory-to-motor imitative map is rendered dimensionally consistent and temporally stable by motor-to-auditory learned expectations that use the ART Matching Rule to attentively select and learn auditory feature patterns that are consistent with the active motor commands. The motor-to-auditory selection process mechanistically explicates the “motor theory of speech perception” (Galantucci, Fowley, & Turvey, 2006; Liberman & Mattingly, 1985).

The subsequent processing stages for learning item chunks, storing sequences of item chunks in working memory, and learning list chunks of these stored sequences use the mechanisms that were outlined in Section 16 and further explicated below. These sequence processing stages occur after the speaker normalization stage, and thus enable language meanings that were learned from one teacher’s voice to be understood when uttered by another speaker.

17.2. From auditory streaming to speaker normalization: A shared circuit design

How does speaker normalization occur in the brain? How can evolution be smart enough to discover speaker normalization? Section 16.4 proposed that working memories are just specialized recurrent shunting on-center off-surround networks, which are ubiquitous because they solve the noise-saturation dilemma. Has the brain specialized another broadly used design to accomplish speaker normalization?

The Neural Normalization Network, or NormNet, model (Ames & Grossberg, 2008) proposes that speaker normalization specializes the same kinds of neural mechanisms that are used to form auditory streams. This proposal offers a parsimonious solution to the speaker normalization problem because it is only after acoustic sources, including speakers, have been separated from one another by streaming that they can be speaker-normalized, just as it is only after visual figure-ground separation occurs that invariant visual categories can be learned (Fig. 19). Speaker normalization specializes streaming mechanisms at the next processing stage.

Both the ARTSTREAM and NormNet models use neural circuits that are called *strip maps* and *spatially asymmetric competitive maps*. Both circuits also contribute to many other brain processes. The most familiar example of strip maps are ocular dominance columns in visual cortical area V1 that are sensitive to inputs from one eye at each position, but whose constituent cells respond selectively to differently oriented visual features at that position (Grossberg & Seitz, 2003; Hubel & Wiesel, 1968). Strip maps have also been used to model how place value numbers may be learned by human brains (Grossberg & Repin, 2003). Here, different portions of each numerically-specific strip represent different place values; e.g., 10s or 100s.

In strip maps for auditory streaming, different portions of each frequency-specific strip are devoted to different streams (within the parallel copies 1, 2, 3, ... of the spectral stream layer of Fig. 30). In strip maps for speaker normalization, different portions of each frequency-specific strip shift the frequency range of heard sounds by a different amount. Thus, speaker normalization circuits may have arisen during evolution from an elaboration and specialization of auditory streaming circuits, and streaming circuits, in turn, may have specialized broadly occurring brain designs.

This proposal clarifies how speaker normalization can transform auditory signals, right after they are separated into separate streams, for purposes of speech and language classification and meaning extraction, yet how the frequency content of the streams can be preserved for purposes of speaker identification in a separate processing stream, as illustrated in the ARTSPEECH architecture (Fig. 33, right panel) of Ames and Grossberg (2008). Phonetic processing within item-list resonances (Section 16) that are computed after speaker normalization occurs would not be expected to be as sensitive to frequency harmonics as are the spectral-pitch resonances that occur before speaker normalization occurs, and that are used to identify voices and speakers in a separate cortical stream (Fig. 33, right panel). Consistent with this expectation, harmonics are more important during auditory streaming than during phonetic perception (Remez, 2003; Remez, Pardo, Piorkowski, & Rubin, 2001; Remez, Rubin, Berns, Pardo, & Lang, 1994).

NormNet was tested by speaker-normalizing and learning steady-state vowel categories from the Peterson and Barney (1952) database with an accuracy similar to that of human listeners. Further development of ARTSPEECH for processing naturally occurring speech is much to be desired. The following kinds of additional processes are needed to accomplish this goal.

17.3. From speaker normalization to rate normalization

Learned codes for speech and language meaning need also to be *rate-invariant* to understand multiple speakers talking at variable rates. Many features of speech signals are, however, not preserved as speech rate changes, so rate-invariant representations must be actively constructed by the brain.

Human listeners can also consciously hear and understand particular voices speaking at variable rates, and human speakers can use their speaker- and rate-invariant representations of meaning to speak or sing in their own voices and at their own volitionally-controlled rates. How the brain accomplishes these various competences, and how they are consciously heard and recognized will be proposed below.

Three additional design features are needed to solve the rate-invariance problem:

(1) *Sustained and transient streams*. Parallel cortical streams exist for processing sustained (S) vs. transient (T) acoustic features to represent all the sounds that occur during audition, speech, and music.

(2) *Intra-word rate-invariance*. Gain control signals from the T stream to the S stream compensate for non-invariant changes within syllables, such as CV pairs, as speaking rate changes. This interaction creates more invariant S/T ratios across rates.

(3) *Inter-word rate-invariance*. A rate-dependent gain control adjusts the integration rates of working memory and list chunking networks as the overall speech rate changes across syllables. This process generates more rate-invariant storage of speech and language representations in working memories and list chunking networks.

Each process will be briefly reviewed in turn. Each provides more examples of how Item-List resonances support conscious speech percepts.

17.4. Transient and sustained processing streams in audition and speech

The brain's auditory system includes sustained cells that are sensitive to particular frequencies over time, and transient cells that are sensitive to FM sweeps across multiple frequencies at a particular rate and direction (e.g., Britt & Starr, 1976; Delgutte & Kiang, 1984a, 1984b, 1984c; Mendelson, Schreiner, Sutter, & Grasse, 1993; Møller, 1983; Pickles, 1988; Rhode & Smith, 1986a, 1986b; Sachs & Young, 1979; Tian & Rauschecker, 1984; Young & Sachs, 1979). Cohen and Grossberg (1997) modeled how coarticulated consonants and vowels may activate distinct, but parallel, transient (T) and sustained (S) channels, and how this selectivity helps to explain data about auditory nerve processing. The S channel can discriminate synchronous vocalic quality, while suppressing transient speech signals. The parallel T channel can discriminate onsets and offsets of fricatives and stop consonants, as well as detect vowel onsets and offsets. Rauschecker and Tian (2000) review cortical regions that include S and T cells, including the core and lateral belt regions, respectively.

Transient and sustained cell responses also occur within the visual system, as illustrated by surface-shroud resonances (S) and change blindness (T); see Section 10.

17.5. PHONET: Asymmetric T-to-S gain control creates within-syllable rate-invariance

A faster speech rate may more strongly activate transients in the speech signal but leaves less time to integrate sustained signals. Thus, faster rates can change the relative activities of T and S cells, with T cells being favored. Such rate-dependent changes in relative T and S activities could cause serious problems for learning speech codes because the bottom-up adaptive filter pathways tend to activate the same set of chunks when their relative activities across the network are preserved (Section 16.4).

A more rate-invariant T/S ratio is created by letting T and S inputs activate their own working memories, after which excitatory signals from the T working memory increase the integration rate of the S working memory, thereby compensating for the relative strengthening of T relative to S activation as speaking rate increases (Fig. 34, left column, upper row). Several experiments have reported such asymmetric vocalic context effects from T to S, but not conversely (Kunisaki & Fujisaki, 1977; Mann & Repp, 1980), and psychophysical experiments support the importance of consonant/vowel ratio as a cue for voicing in English (e.g., Port and Dalby (1982)).

Boardman et al. (1999) developed the PHONET model to quantify how T and S working memories can use asymmetric T-to-S gain control to create rate-invariant representations of individual speech syllables or words. PHONET was tested by quantitatively simulating how, using CV syllables such as /ba/ and /wa/, an increase in duration of vowel /a/ can switch the percept of the preceding consonant from /w/ to /b/ (Miller & Liberman, 1979). This switch does not occur if the relative durations of these sounds are preserved as their absolute durations are changed. A change in frequency extent (total frequency change), but not rate, can also influence the /b/-/w/ distinction (Schwab, Sawusch, & Nusbaum, 1981). PHONET fits 99.3% of the variance in both data sets with the same parameters.

17.6. ARTPHONE: Rate-sensitive gain control creates rate-invariant working memories

Rapidly-acting T-to-S gain control cannot compensate for speech properties that change with speech rate across multiple words. As speech rate increases, each feature pattern and its

item chunk has less time to be stored in working memory. The same problem afflicts the list chunk working memory. If working memory integration rates do not keep up with changes in speech rate, then the activity patterns that are stored in working memory may be altered. The ability to learn rate-invariant list chunks and their top-down expectations will be correspondingly degraded. An effective compensatory mechanism needs to speed up or slow down integration rates across all working memory and list chunking networks to mirror changes in speech rate.

Taken together, these two sorts of gain control, one acting within individual words and the other acting across sequences of words, begin to explain how a rate-invariant speech representation may be created internally by brain dynamics from variable-rate speech signals that do not themselves exhibit properties of invariance. How rate-dependent productions and their conscious percepts via auditory feedback can be created from these rate-invariant representations will be described in Section 18.

The ARTPHONE model (Fig. 34, left column, lower row; (Grossberg et al., 1997)), and its elaboration in the ARTWORD and cARTWORD models (Fig. 31), proposes how rate-dependent gain control can act across sequences of words. In particular, a rate-dependent gain control processing stage (variable g in Fig. 34, left column, lower row) time-averages the number of transient signals through time (from the input rate estimate), and thus depends in yet another way on the T stream. More transient signals per unit time occur at faster speech rates, and are integrated through time to create a larger gain control signal. Larger gain control signals speed up the integration rate of working memory and list chunking networks more than smaller signals to compensate for the changes in speech rate.

This kind of gain control helps to explain paradoxical properties of psychophysical data about conscious speech perception and recognition. In brief, the heuristic ideas are as follows: Listener percepts are approximately invariant with respect to mean silent interval between stop-consonant clusters. Durations of silent intervals are inversely related to the speech rate. The speech rate alters the working memory integration and list chunking network integration rates via T-sensitive gain control. The result is a tendency to achieve a rate-invariant percept. Paradoxically, this *rate-invariant* percept can produce *rate-dependent* category boundaries of heard speech.

A classical example of this phenomenon was reported by Repp (1980). Repp constructed VC-CV syllables from the syllables [ib], [ga], and [ba] to form [ib]-[ga] and [ib]-[ba]. The durations of the silent intervals between these syllables were varied. There were multiple conditions, but the main results were the following (Fig. 34, right column): S-shaped category boundary curves were measured for the probability of hearing two syllables [ib]-[ga] (with [ib]-[ga] more probable at longer silent intervals and [iga] more probable at shorter silent intervals) and [ib]-[ba] (with [ib]-[ba] more probable at long silent intervals and [iba] more probable at short silent intervals). Remarkably, the category boundary curve for [ib]-[ba] was shifted 125 ms to the right of that for [ib]-[ga]. Consequently, subjects could hear two consonants [ib]-[ga] at some of the same silent intervals as they heard only one consonant in the single syllable [iba].

Why did switching from the different consonant /g/ to the same consonant /b/ in the CV syllable shift the latter category boundary curve to the right by 125 ms, a huge duration compared to the time scale of individual neuronal responses? Why was silence heard between [ib] and [ga] at some of the same times that /b/ was continuously heard in response to [iba]? These properties follow naturally from ART mechanisms operating in time within working memories and their list chunking networks. In particular, the separate syllables [ib]-[ga] are heard at shorter silent intervals because /g/ mismatches, and thereby inhibits, /b/ at these times using the mechanism of mismatch reset that occurs in all ART

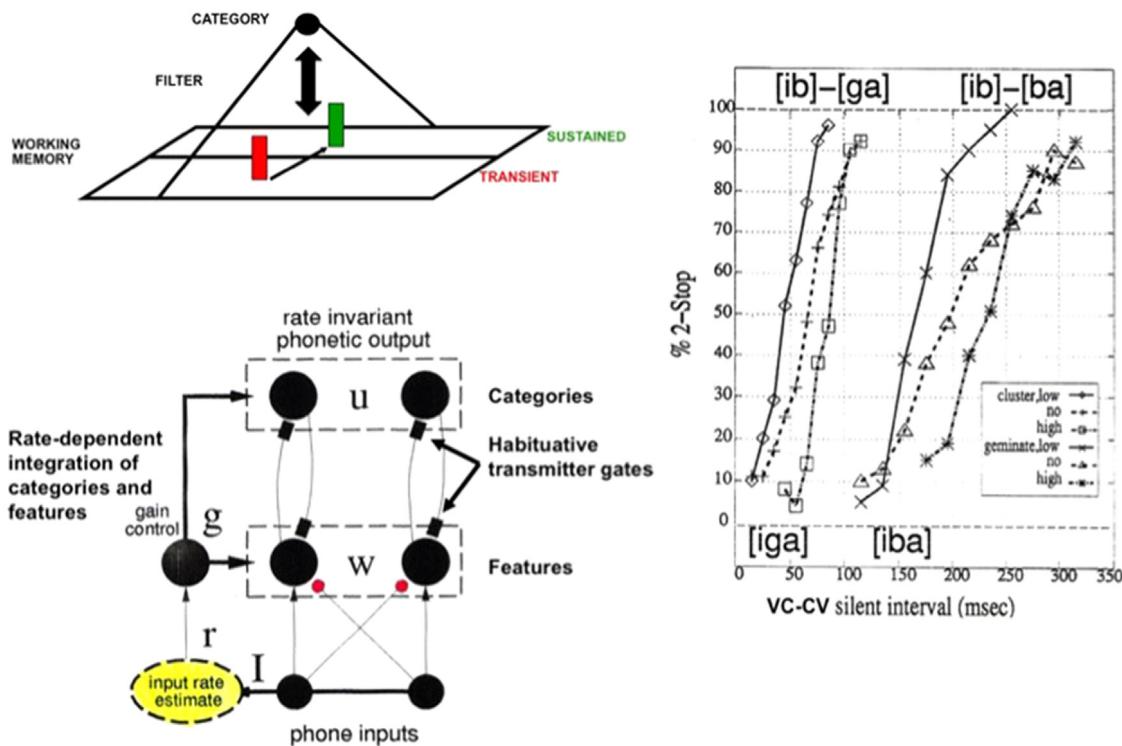


Fig. 34. (Upper left panel) In parallel transient (T) and sustained (S) working memories, T-to-S gain control helps to maintain invariant T/S activity ratios within a syllable during variable speech rates. This invariant enables the same category to be activated as speech rate changes in the PHONET model of Boardman et al. (1999). (Lower left panel) ARTPHONE model circuit showing the T-activated input rate estimate that automatically increases or decreases the integration rate of working memory (features) and list chunking (category) cells, thereby helping to preserve speech codes across syllables during variable rate speech. [Reprinted with permission from Grossberg et al. (1997).] (Right panel) Psychometric functions for the probability of perceiving two distinct stop consonants in two different [VC]-[CV] syllables in six conditions during which the silent interval between syllables was varied, averaged over eight participants. [Reprinted with permission from Repp (1980).]

models (Fig. 13(c)). In contrast, as in the auditory continuity illusion (Section 15.2) and phonemic restoration (Section 16.3), it takes more time for a resonance to reach threshold in response to a given cue than for a second occurrence of that cue, in this instance /b/ in [ib]-[ba], to keep the resonance active. Thus, the second occurrence of /b/ in this case keeps the resonance active for much longer than one would expect if only feedforward neuronal processing were involved. As a result, [iba] is heard for much longer silent intervals than [iga]. The Repp (1980) data may thus be understood as an example of resonant reset vs. resonant fusion.

In the percept of [ib]-[ba], why does the resonance that supports the sound /b/ ever end, given that there is no explicit reset cue? ART proposes that chemical transmitters are released within the resonating pathways between processing levels (Fig. 34, left column, lower row, square synapses). Persistent activity-dependent transmitter release during resonance inactivates, habituates, or depresses them (Grossberg, 1972b, 1980), thereby weakening the feedback signals until the resonance collapses. Thus, a resonance can be terminated by resonant reset or habituative collapse.

How do rate-dependent integration and rate-invariant working memory enter this story? This can be seen from finer aspects of the Repp (1980) data, which measured three category boundary curves for each of the VC-CV syllables [ib]-[ga] and [ib]-[ba] (Fig. 34, right column): A low anchor condition had its distribution of silent intervals skewed to be shorter, a high anchor condition had them skewed to be longer, and a no anchor condition had no skew. The three curves show that listeners track the mean physically silent interval to make their judgments in response to their conscious percepts. These rate-dependent category boundary shifts thus emerge from rate-sensitive integration of working memory and chunking networks to create a rate-invariant working memory.

These six data curves were simulated as emergent properties of model dynamics. Many other data about conscious speech

percepts can also be explained by using this combination of Item-List resonant feedback, rate-dependent gain control of working memory integration rate, and habituative transmitter gating of the feedback signals that support resonance.

17.7. From masked priming to resonance in lexical decision and word frequency tasks

The comparison between resonant fusion and resonant reset that plays an important role in explaining the Repp (1980) data on category boundary shifts also helps to explain data about the way in which masking stimuli can influence error rates and reaction times during lexical decision tasks. In a classical set of experiments, Schvaneveldt and McDonald (1981) presented word and non-word targets after semantically related (R), neutral (N), or unrelated (U) word primes. Two kinds of data were reported: error rates in tachistoscopic experiments with a backward mask, and reaction times in non-masked conditions. In the tachistoscopic experiments, the target was displayed for approximately 33.3 ms and was followed by a masking pattern consisting of a string of number signs (#). In the reaction time experiments, the target remained visible until the subject responded. The subjects were instructed to respond as rapidly and accurately as possible.

Grossberg and Stone (1986b) explained the paradoxical pattern of experimental results in terms of how the ART Matching Rule works in different priming situations, including the inability of top-down expectations to act before the mask interferes with the persistence of word and non-word target representations in working memory. In the unmasked condition, word vs. non-word primes could have different effects on reaction times due to the ensuing match vs. mismatch events caused by allowing enough time for top-down expectations to act.

A similar combination of mechanisms was used to explain data about word frequency effects in recognition and recall (Underwood & Freund, 1970), notably why forced choice recognition of (old-new) word pairs leads to more errors for old high-frequency words paired with new high-frequency words (H-H) than for old low-frequency words paired with new low-frequency words (L-L), yet more errors occur for old low-frequency words paired with new high-frequency words (L-H) than for the converse (H-L). Various other lexical decision and word frequency data have been explained using ART concepts (e.g., Dunbar, 2012; Glotin et al., 2010; McLennan et al., 2003; Vitevitch & Donoso, 2011).

18. From invariant working memory storage to volitionally-variant productions and percepts

This section summarizes how speaker- and rate-invariant working memories and list chunking networks can recall their learned representations as voice-specific and rate-variant productions that are matched against auditory top-down expectations, and consciously heard when a stream-shroud resonance develops in response to a good enough match. The networks are successively generalized to enable variable-rate productions without and with a learned rhythm, and without and with a singing voice. Further development of these modeling concepts and mechanisms is needed to process realistic speech and music.

18.1. LIST PARSE: Volitionally-controlled variable-rate sequential performance

The first example proposes how recall from working memory can occur in an individual's usual speaking voice with the production rate controlled by a volitional GO signal from the basal ganglia. This example uses the LIST PARSE model (Fig. 35; Grossberg & Pearson, 2008) to illustrate how multiple processing stages are coordinated to enable variable-rate production from rate-invariant working memory and list chunking networks. In particular, read-out rates from cognitive working memory in ventrolateral prefrontal cortex, and motor working memory in dorsolateral prefrontal cortex, need to be coordinated with the current production, so that the next command does not occur before the current production is sufficiently executed. As a result of this coordinated timing, the read-out of an ART sensory expectation, in the case of speech and language, can also be coordinated with the expected time that auditory feedback from the production will be consciously heard, or in the case of vision, can be coordinated with the expected time that a familiar scene will be consciously seen after an eye movement.

In order for such coordination to work, the read-out of the next working memory command must be timed to take into account when the current production is almost executed, even if the performance can occur at variable rates. How does the knowledge about the state of a current production influence the timing of a working memory read-out, given that these two processes occur in separate networks and compute different kinds of information? The VITE and DIRECT models (Fig. 27) describe how a motor trajectory approaches its target position as its difference vector (D) approaches zero. However, the difference vector does not include information about the speed with which the movement is being made. The outflow velocity command (DG) that multiplies D with the GO signal G does incorporate this information (Fig. 27, left panel). However, DG can be small at both the beginning and the end of a movement, since G can grow from zero when a movement is launched. How does the network distinguish between a small DG at the beginning vs. the end of a movement?

LIST PARSE proposes that the network achieves this distinction by computing the *deceleration* of the trajectory velocity profile

as D, and DG, both approach zero (Fig. 35, right panel). To do this, in addition to inputting to the present position command P, where it is integrated to reach the target position T, DG also inputs to two additional cell populations. It is slowly integrated through time by activity B and more rapidly integrated through time by A. Then B excites the rehearsal wave stage R, while A inhibits it. The difference B-A estimates the deceleration, and causes a rehearsal wave burst when it is positive, but sufficiently small, thereby reading-out the next command from the working memory at the appropriate time. The entire command hierarchy can then coordinate variable-rate read-out of working memory commands with trajectory generator dynamics to control smooth sequential productions. LIST PARSE was used to quantitatively simulate psychological data from humans about immediate serial recall, and immediate, delayed, and continuous distractor free recall; and neurophysiological data from monkeys about read-out of planned arm movement sequences.

The inputs to the cognitive working memory in LIST PARSE (I_i in Fig. 35, right panel) do not include the speaker-invariant and rate-invariant pre-processing stages that were discussed in Section 17. These pre-processing networks can readily be joined to the front end of LIST PARSE.

18.2. lisTELOS: Basal ganglia volitional control and Item-Order-Rank working memory

Volitional control by the basal ganglia plays an important role in the LIST PARSE model (Fig. 35, left panel, and variables V, F, and G in the right panel). Such volitional signals play a critical role in opening cognitive and motoric gates, which is necessary to enable the corresponding processes to unfold (Alexander, DeLong, & Strick, 1986; Graybiel, 1997), notably in self-initiated behaviors (Levy & Dubois, 2006; Taniwaki et al., 2006). In particular, basal ganglia gates open and close in a coordinated manner across the several brain systems that control working memory storage, choice, read-out, and performance of an action.

As one example of basal ganglia gating, consider the problem of how to control a decision about where next to move the eyes, so that this movement can be coordinated with the read-out of a sensory expectation of what the brain expects to see when the eyes do move to that target position. Typically, there are multiple possible visual target positions to which the eyes can move. The TELOS model (Brown et al., 2004) predicted how agreement between prefrontal and parietal representations of a target position causes a parietal–prefrontal resonance that selects this target position, and opens the correct basal ganglia gate, while also enabling basal ganglia-mediated release, in a different part of the brain, of a contextually-appropriate movement command to that position. This parietal–prefrontal resonance is one of the resonances that is not accessible to conscious experience (Section 8). Subsequent neurophysiological data of Buschman and Miller (2007) supported this prediction by describing such a parietal–prefrontal resonance during movement control, and Pasupathy and Miller (2004) additionally described different time courses of activation in the prefrontal cortex and basal ganglia that are consistent with how basal ganglia-mediated gating of prefrontal cortex occurs in TELOS. The TELOS model was also used to simulate the learning of six different eye movement tasks. After learning was complete, the model's learned parameters enabled it to quantitatively simulate neurophysiological data describing the dynamics of 17 functionally different identified cell types when the corresponding eye movement tasks were performed in monkeys. The predicted functional roles of these cell types remain to be experimentally tested.

The lisTELOS model (Silver et al., 2011) extended the TELOS model to explain and predict how sequences, or lists, of eye movements can be carried out, while continuing to simulate

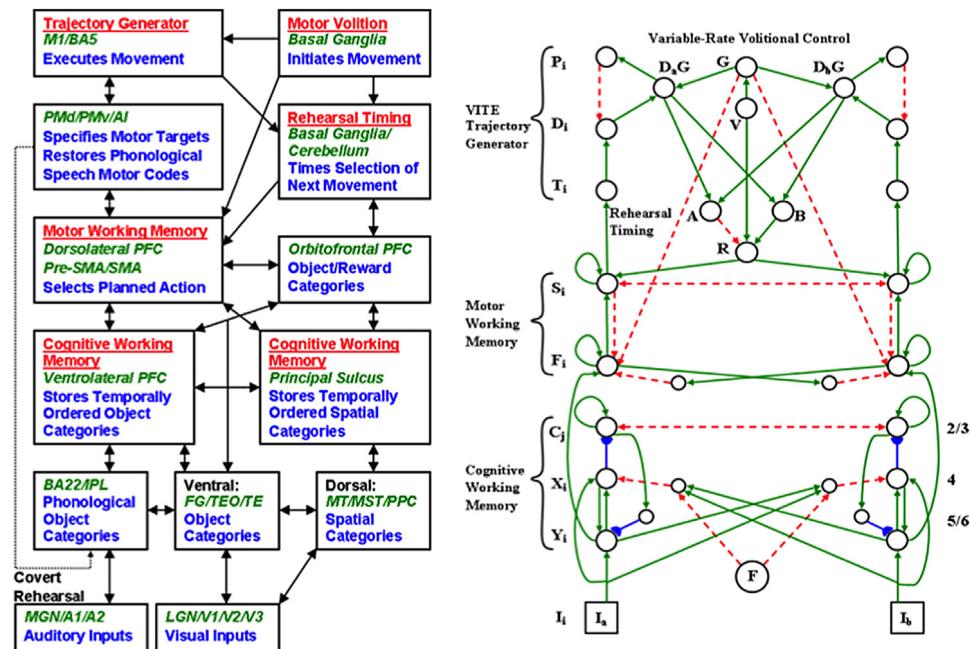


Fig. 35. LIST PARSE model. (Left panel) The brain processes that are incorporated in the model are written in red and underlined. (Right panel). The LIST PARSE model circuit. LIST PARSE models interactions between Cognitive Working Memory, Motor Working Memory, VITE Trajectory Generator, and Variable-Rate Volitional Control circuits, whose anatomical substrates are summarized in the left panel. The Cognitive Working Memory network is assumed to be within the deeper layers of ventrolateral PFC (4, 5/6), and the corresponding list chunking network is assumed to be within the superficial layers (2/3). Green solid arrows are excitatory, red dashed arrows are inhibitory, and blue lines ending in hemidisks are adaptive. Only 1-item chunks (C_i) and their feedback connections within a single Cognitive Working Memory channel are shown, whereas the model simulates chunks corresponding to words of variable lengths in layer 2/3, and feedback from layer 2/3 to layer 5/6 of the cognitive working memory is broadly distributed. Also, only the excitatory projections from Cognitive Working Memory to the Motor Working Memory ($Y_i \rightarrow F_i$) are shown. Several volitional gain control signals determine model dynamics. For example, the gain control signal F determines whether or not a sequence will be stored within the Cognitive Working Memory. The volitional signals V and G control variable-rate performance of the stored sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

everything that TELOS could. To accomplish this extension, lisTELOS predicted how basal ganglia gates can sequentially open and close in a coordinated manner across several brain systems. In particular, the dynamics of three different basal ganglia loops are coordinated for the planning, selection, and release of such an eye movement sequence. In addition, lisTELOS modeled how a sequence, or list, of eye movement target positions can be stored in an Item–Order–Rank spatial working memory. Because of its rank sensitivity, such a working memory can cause the eyes to move to and from the same position during performance of a sequence of eye movements. lisTELOS is supported by simulations of behavioral, electrophysiological, and anatomical, data, notably data demonstrating prefrontal cortical cells that are sensitive to the item, order, and rank properties of stored events in working memory.

This demonstration of how the basal ganglia control planning, choice, and performance of sequences is relevant to speech and language, and thus to the timed read-out of sensory expectations that will be matched during conscious hearing, due both to the fact that all linguistic, spatial, and motor working memories have a similar design (Section 16.4), and to the fact that all basal ganglia loops have a homologous anatomical organization (Alexander et al., 1986). When multiple basal ganglia loops are used to simulate the coordinated gating through time of perceptual, cognitive, and affective processes, a more complete description becomes possible of how several consciousness-supporting resonances evolve through time while a self is engaged with its world.

18.3. From learning of invariant lists to rhythmic intonated performance

LIST PARSE, and its various extensions, clarify how changing the amplitude of a volitional signal can, on a moment-by-moment

basis, vary the speed with which stored sequences are performed (Fig. 35). Albeit useful, this simple form of rhythmic performance is insufficient to explain the more sophisticated rhythms that occur during natural speech and singing. This article does not include a comprehensive analysis of rhythm (e.g., Ackermann, 2008; Deutsch, 2013; Fraisse, 1982; Krumhansl, 2000; Ladefoged & Disner, 2012), notably the role of beat in music. On the other hand, it does propose how parallel neural mechanisms for rate-invariant and speaker-normalized representations of speech, and for pitch-dependent and rhythm-dependent speech intonation (Ladefoged & Disner, 2012), may interact to achieve “online sequencing of syllables into fast, smooth, and rhythmically organized larger utterances” (Ackerman, 2008, p. 265), including how these several kinds of information are learned, stored, and combined during fluent performance and conscious awareness thereof. This analysis also proposes elements of how speaker-normalized representations may be sung with multiple tunes. These neural mechanisms and their design principles extend and complement the discoveries of processing streams, components, and their interactions that have been already been proposed (e.g., Armony, Aube, Angulo-Perkins, Peretz, & Concha, 2015; Norman-Haignere, Kanwisher, & McDermott, 2015; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000; Sieve & Okada, 2015).

These competences depend upon a brain design that supports a flexible relationship between stored order information and the pitch and rhythm with which it is performed. In particular, when a listener hears a sufficiently short segment of novel speech that is uttered with one rhythm, he or she can immediately perform it with the same rhythm, or with a different rhythm that the listener chooses. Thus, the rate-invariant order information in the sequence can flexibly be decoupled from, and later recoupled with, one of several possible rhythmic productions. I like to call this property *factorization of order and rhythm* (Grossberg, 1986) for the following reasons: The speech items and their

order can be stored in a rate-invariant Item–Order–Rank working memory (Section 16). The fact that a sufficiently short stored order sequence, after one hearing, can be performed with the same rhythm, or with a different rhythm, shows that a separate system with working memory properties exists for storing the rhythm. The fact that the stored rhythm does not have to be used for the performance shows that these parallel order and rhythm working memories can operate independently from each other, under volitional control.

This property is related to the so-called *independence of lyrics and tunes* that occurs when hearing and performing vocal music (Besson, Faita, Peretz, Bennel, & Requin, 1998; Bonnel, Faita, Peretz, & Besson, 2001). A big caveat to such “independence” is that these two types of information interact during storage – when the heard order information influences the perceived timing and storage of a tune’s pitch contours and rhythm – and again during production – when order, pitch, and rhythm information are combined to yield the final performance (Sammel et al., 2010; Slevc & Okada, 2015). These interactions are the reason why the term “factorization” is preferred to “independence” of order, pitch, and rhythm herein. Indeed, the learned list chunks that are stored in a rate-invariant Item–Order–Rank working memory help to determine what rhythm is perceived and separately stored, much as how the learned chunks of a language may influence its perceived rhythm.

18.4. From speaking to singing

One more interaction between complementary processing streams will now be briefly noted. Order and rhythm information is stored within a working memory whose inputs are speaker-normalized. These working memories read-out their commands to representations that control the speaking voice of the performer. This inter-stream interaction builds upon the learned circular reaction that was discussed in Section 14 between speaking and hearing that occurs throughout life.

Learning also goes on throughout life of a parallel circular reaction that links learned spectral-pitch-and-timbre categories for the recognition of heard sounds (Table 2), which are not speaker-normalized, to the motor synergies that control the pitches generated by the vocal folds (Sundberg, 1977). Sequences of these pitch-sensitive categories can be stored in a spectral working memory as the melody of a song (Section 15.3). When a speaker- and rate-invariant working memory, a rhythm working memory, and a spectral working memory are concurrently read-out, coordinated by the basal ganglia (Section 18.2), then learned language sequences at a given rhythm can be produced with the series of pitches characteristic of a song.

The main point for conscious audition is that these working memories read out production commands at the same time as they read out top-down sensory expectations of the sounds that are expected to be heard. When these expectations are matched by the auditory feedback of the productions (e.g., Fig. 33, left panel), the song is consciously heard by coordinated stream-shroud and spectral-pitch-and-timbre resonances (Table 2). As noted in Sections 15.3 and 15.4, these resonances also support conscious percepts of heard sounds that are emitted by the external world. More modeling needs to be done to carry out large-scale simulations of such coordinated resonances during the control of real-world auditory behaviors.

19. The feeling of what happens: Cognitive–emotional resonances for conscious feelings

19.1. Resonance between object, value, and object–value categories: Beyond Chomsky and Skinner

The visual and auditory processes summarized above represent external information about the world, but do not evaluate

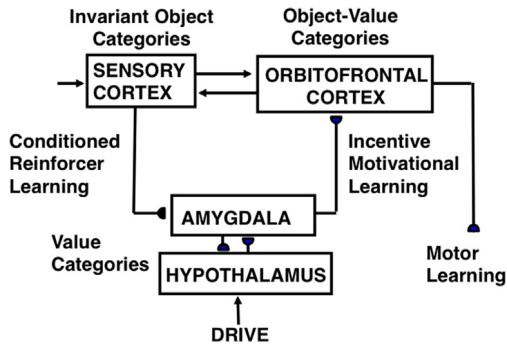


Fig. 36. Cognitive-Emotional-Motor (CogEM) model macrocircuit. CogEM models how invariant object categories in sensory cortex can activate value categories, also called drive representations, in the amygdala and hypothalamus, and object-value categories in the orbitofrontal cortex. Converging activation from an object category and its value category are needed to vigorously fire the corresponding object-value category. Achieving such convergence from the amygdala requires prior conditioned reinforcer learning and incentive motivational learning. When an object-value category fires, it can send positive feedback to its object representation and attentionally enhance it with value-derived activation. The motivationally-enhanced object representation can then better compete with other object representations via a recurrent on-center off-surround network (not shown) and draw attention to itself. Closing the feedback loop between object, value, and object-value categories via a cognitive-emotional resonance can induce a conscious percept of having a particular emotion, or feeling, towards the attended object, as well as knowing what it is. As this resonance develops, the object-value category can generate output signals that can activate cognitive expectations and action commands through other brain circuits.

Source: Adapted from Grossberg (1971) and subsequent CogEM articles.

how important this information is for survival or success. Interactions between perceptual/cognitive and evaluative reinforcement/emotional/motivational mechanisms accomplish this. What kind of resonance supports a conscious emotion? What kind of resonance enables an emotion to generate a conscious “feeling of what happens” (Damasio, 1999), knowing the source of that feeling, and being able to behave in an appropriate way in response to it? In particular, how does an emotion generate motivation that may help to search for, and acquire, goal objects that can satisfy that emotion? Under what circumstances can emotions affect behavior without being conscious? How do breakdowns in these resonances lead to prescribed mental disorders, such as autism and schizophrenia, and how does the affective meaning of familiar objects and events fail to organize behavior in a context-appropriate way when this happens?

A study of cognitive–emotional interactions, and the conscious resonances that they support, is needed to propose answers to these questions. Such resonances focus *motivated attention* upon valued objects in an adaptively timed way. Unlike visual and auditory conscious percepts, emotions or feelings do not represent *external* sensory cues, such as sights and sounds. Rather, they represent *internal* affective states of fear, pleasure, pain, happiness, sadness, and the like, that may be activated by a variety of external and internal signals.

When conscious emotions are triggered by viewing affectively meaningful visual scenes, *cognitive–emotional resonances* are proposed to support these feelings, our knowledge of them, and our ability to respond appropriately. These resonances occur using feedback interactions (Fig. 36) between *invariant object categories* for recognition of objects, also called object categories, in temporal cortex (Sections 7, 10 and 11); *value categories*, or drive representations, in the amygdala, where reinforcing and homeostatic inputs converge and from which motivational signals emerge; *homeostatic representations* in the hypothalamus (not labeled in Fig. 36) that register and control autonomic and other affective reactions; and *object–value categories* in orbitofrontal cortex to which object categories and value categories both send their output signals. The

object category representations may, in turn, synchronously interact with surface-shroud and stream-shroud resonances that support conscious awareness of the visual or auditory events that activate object categories (Sections 10 and 15).

Neural models of cognitive-emotional resonances began with the articles of Grossberg (1971, 1972a, 1972b) and Grossberg (1975) at a time when there was a major split between studies of cognition, as exemplified by the work of Chomsky (1957) in linguistics, and of emotion, as exemplified by the work of Skinner (1938) on instrumental conditioning. Skinner's anti-theoretical behaviorist stance opposed any model development. Chomsky's universal grammar formalism included no dynamics, let alone learning dynamics, and was devoid of any allusion to emotion. Instead, cognition and emotion studies were often antagonistic to each other, as illustrated by the critique of Skinner's behaviorism in Chomsky (1959). Such divided concepts of cognition and emotion have persisted during many historical periods, as described in the book Wittgenstein's Vienna (Janik & Toulmin, 1996). The Grossberg articles began a mechanistic synthesis of cognition and emotion by showing how the same kinds of neural equations and modules, suitably specialized (see Section 3), could interact during a cognitive-emotional resonance to join *what we know* with *what we feel*. As summarized below, the resurgence of affective neuroscience in the intervening years has supported all the predictions that grew out of this emerging synthesis.

19.2. Reinforcement learning: Conditioned reinforcer and incentive motivational learning

Many visual cues do not initially have affective meaning to an individual. These cues can acquire affective significance through reinforcement learning, whereby the pathways between the object, value, object-value, and homeostatic representations can be altered. For simplicity, consider only the simplest kind of reinforcement learning, called Pavlovian or classical conditioning (Kamin, 1968, 1969; Pavlov, 1927), during which a conditioned stimulus, or CS, that initially may have no emotional significance, is paired with an unconditioned stimulus, or US, that can from the start generate a strong emotional response. Before conditioning occurs, when a familiar CS occurs, the CS can selectively activate an object category. The object category can, in turn, send priming signals to most or all of the value categories, but cannot vigorously fire any of them. It can also send priming signals to an object-value category. Neither the object-value category nor the value categories can fire vigorously in response to a CS before reinforcement learning occurs.

When a CS is suitably paired in time with a US, reinforcement learning can begin. The US can activate the amygdala value category and hypothalamic homeostatic representations that correspond to the emotion it engenders. As a result, the CS-activated object category and US-activated value category are simultaneously activated, so that learning can occur in the pathways from the active object category to the active value category, thereby strengthening the learned connection between them. After sufficiently many learning trials, the CS can activate that value category. When this happens, the CS is said to have become a *conditioned reinforcer* (Fig. 36).

As a conditioned reinforcer is being learned, the value category that is activated by the US is able to send signals, called *incentive motivation* signals, to many object-value categories, including the object-value category of the CS. When this occurs, convergent signals from both the object category and the value category converge on the object-value category. Such convergent signals are sufficient to activate the object-value category. Now both the value category and an object-value category are simultaneously active, so learning can occur in the pathway from the value category to the object-value category, further strengthening it, while weakening pathway connections to inactive object-value

categories. This strengthening is called *incentive motivational learning* (Fig. 36).

After both conditioned reinforcer learning and incentive motivational learning occur, the CS-activated object category can prime its object-value category and activate its value category. Then convergent input to the primed object-value category from both the object category and the activated value category can fully activate the object-value category and, with it, motivated behaviors. After a value category is conditioned to sufficiently many CSs, activating just a value category can generate an *emotional set* by priming all of the object-value categories with which it was associated in the past.

After an object-value category gets activated, it can generate top-down priming signals to its object category via the ART Matching Rule (Figs. 12 and 36). The importance of this feedback pathway can be seen by considering situations where multiple visual cues occur simultaneously. If, after competition occurs between the object-value categories, the most active object-value category reads out its top-down attentional signal to a currently active object category, it can amplify and thereby help to select this motivationally consistent object category, while the competitive interactions in the off-surround of the ART Matching Rule attentionally *block* object categories of motivationally-irrelevant sensory cues (Grossberg & Levine, 1987; Pavlov, 1927).

Activation of this feedback circuit through temporal–amygdala–orbitofrontal interactions can create a cognitive–emotional resonance that focuses and maintains motivated attention upon a motivationally salient object category, while also supporting conscious feelings about this object, and enabling the release of motivationally-compatible actions (Tables 1 and 2). This kind of cognitive–emotional resonance was first predicted by the Cognitive–Emotional–Motor, or CogEM, model (Fig. 36; Grossberg, 1971, 1972a, 1972b, 1982, 1984b) and was used to quantitatively explain and simulate in these and later articles parametric properties of many cognitive–emotional data (e.g., Grossberg & Gutowski, 1987; Grossberg & Levine, 1987; Grossberg & Schmajuk, 1987). Predictions of CogEM have been supported by subsequent data, notably data concerning the anatomy and neurophysiology of the brain regions that the model interactions functionally explain.

Confidence in CogEM predictions about the role of cognitive–emotional resonances in conscious feeling, knowing, and acting grows with the amount of interdisciplinary data that this model and its extensions can explain and predict. To this end, further developments of the CogEM model, such as the Matching Objects to Internal Values Triggers Option Revaluations, or MOTIVATOR, model (Fig. 37; (Dranias, Grossberg, & Bullock, 2008; Grossberg, Bullock, & Dranias, 2008)), proposed how the amygdala/hypothalamus and basal ganglia carry out complementary roles during reinforcement learning, and simulated neurophysiological data about inferotemporal, amygdala, orbitofrontal, and hypothalamic cell dynamics during the learning process, including the discharge dynamics of known cell types that predict saccadic reaction times and CS-dependent changes in systolic blood pressure, while simulating data about food-specific satiety, Pavlovian conditioning, reinforcer devaluation, and simultaneous visual discrimination.

In particular, in response to a CS, the model amygdala and hypothalamus interact to calculate the expected current value of the subjective outcome that the CS predicts, constrained by the current state of deprivation or satiation. Model amygdala cells are called value categories because they respond selectively to different patterns of activation across hypothalamic homeostatic cells, thereby categorizing the homeostatic substrates of different emotions. These value categories also send top-down signals that obey the ART Matching Rule back to their hypothalamic sources, and thereby dynamically stabilize the learned amygdala value categories and attentionally bias homeostatic responses. The

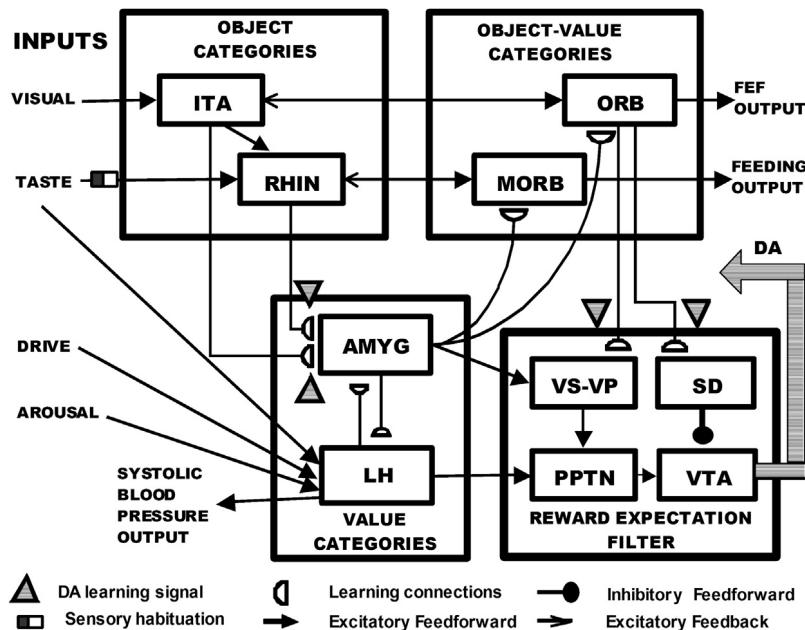


Fig. 37. MOTIVATOR model circuit. MOTIVATOR includes additional processes to the CogEM model, notably interactions with the reward expectation filter in the basal ganglia that generates dopaminergic Now Print signals in response to unexpected reinforcing cues. MOTIVATOR also models how hypothalamic homeostatic representations reciprocally interact with amygdala value categories via an ART circuit, and how both visual reinforcement learning, via cortical areas like the anterior inferotemporal cortex (ITA) and the lateral orbitofrontal cortex (ORB), and gustatory reinforcement learning, via cortical areas like the rhinal cortex (RHIN) and the medial orbitofrontal cortex (MORB), occur.

Source: Reprinted with permission from Dranias et al. (2008).

amygdala categories also project their expected value information to orbitofrontal object-value category cells that also receive inputs from inferotemporal object category cells. Activating orbitofrontal cells by conjoint inputs from object and value categories codes the subjective values of objects. These values guide behavioral choices via the output signals from orbitofrontal object-value categories.

The MOTIVATOR basal ganglia (Fig. 37) detect errors in CS-specific predictions of the value and timing of rewards. Excitatory inputs from the model pedunculopontine nucleus interact with timed inhibitory inputs from model striosomes in the ventral striatum to regulate dopamine burst and dip responses from cells in the model substantia nigra pars compacta and ventral tegmental area. These burst and dip reactions in response to unexpected reinforcers are broadcast as dopaminergic Now Print signals that modulate associative learning in multiple brain regions, as simulated within the TELOS and lisTELOS models (Section 18.2). It should also be noted that a single CS can become conditioned to more than one value category, and can elicit different motivated behaviors in response to the CS under different motivational conditions.

19.3. Interactions between sensory cortices, thalamus, amygdala, and orbitofrontal cortex

Many experiments have supported the prediction that drive-sensitive value category cells are found in the amygdala (e.g., Aggleton, 1993; LeDoux, 1993, 1996, 2012; Pessoa, 2008; Pessoa & Adolphs, 2010). Multimodal amygdala cells that are hunger- and satiety-selective (Muramoto, Ono, Nishijo, & Fukuda, 1993; Yan & Scott, 1996) and respond in proportion to the value of a food reward have been extensively studied in the primate and rodent (Nishijo, Ono, & Nishino, 1988; Toyomitsu, Nishijo, Uwano, Kurotsu, & Ono, 2002). Activated value categories can, in turn, activate the orbitofrontal cortex via learned incentive motivational pathways (Figs. 36 and 37). Neurobiological experiments have supported the prediction that an object-value category in orbitofrontal cortex receives a direct input from an object category

in a sensory cortex, and indirect incentive motivational inputs from the amygdala (e.g., Barbas, 1995), and that the strength of the incentive motivational input determines how vigorously an object-value representation is activated in response to its object category input (e.g., Baxter, Parker, Lindner, Izquierdo, & Murray, 2000; Rolls, 1999, 2000; Schoenbaum, Setlow, Saddoris, & Gallagher, 2003). In addition to neocortical inputs to the amygdala, a parallel pathway also includes direct subcortical inputs to the amygdala (LeDeux, 1996).

19.4. The feeling of what happens, core consciousness, dual competition, and survival circuits

Damasio (1999) has derived from clinical data what can be viewed as a heuristic version of the CogEM model, and has used it to describe what can be interpreted as cognitive-emotional resonances that support "the feeling of what happens". Translating from Damasio's Figure 6.1 into the CogEM circuit of Fig. 36, his concept of the "map of object X" becomes an invariant object category; the "map of the proto-self" becomes the value category and its multiple interactions; the "second-order map" becomes the object-value category; and the "map of object X enhanced" becomes the object category as it is attentively amplified by feedback from the object-value category. As this cognitive-emotional resonance develops through the excitatory feedback loop between object, value, and object-value categories, the attended object achieves emotional and motivational significance.

This cognitive-emotional resonance binds together complementary aspects of object categories and value categories in much the same way that a feature-category resonance binds together complementary properties of distributed features and their object categories (Section 7.5). In the CogEM model, the invariant object category can classify particular combinations of features in the world, but cannot compute their value to the organism. Activation of a value category sets the stage for thoughts and actions that have a particular emotional and motivational meaning, but does not represent any particular objects or events to which these feelings could

be directed. The *bound state* of the cognitive-emotional resonance combines facts about the external world with feelings about and motivations to act from the internal world. As Damasio (1999, p. 171), notes: "Attention is driven to focus on an object and the result is saliency of the images of that object in mind", leading to what Damasio calls *core consciousness*. Damasio (1999) also went on to write "I do not know how the fusing, blending, and smoothing are achieved..." (p. 180). The CogEM model began to provide explicit explanations of how this happens in the brain starting 28 years before Damasio wrote these words.

As in the case of surface-shroud resonances, a CogEM cognitive-emotional resonance is predicted to propagate to other brain regions with which its object, value, and object-value categories interact, including hypothalamus, visual cortex, and other pre-frontal cortical areas, including areas that control executive planning and action (Sections 16 and 17; Tables 1 and 2). A region like the hypothalamus is particularly important because it controls internal homeostatic states whose activation patterns correspond to different emotions (Figs. 36 and 37). As a result of such a cognitive-emotional resonance, "the feeling of what happens" can ascribe feelings to particular objects and events in the world, activate sensory expectations of what one would like to happen next, and activate behaviors to achieve valued goals.

The CogEM model is also consistent with more recent proposals. These include the "dual competition" model of Pessoa (2009, p. 160): "The proposed framework is referred to as the 'dual competition' model to reflect the suggestion that affective significance influences competition at both the perceptual and executive levels—and because the impact is caused by both emotion and motivation". These two competitive stages correspond to the object and object-value stages in CogEM.

The "survival circuit" concept of LeDoux (2012, p. 653) "integrates ideas about emotion, motivation, reinforcement, and arousal in the effort to understand how organisms survive and thrive by detecting and responding to challenges and opportunities in daily life". The CogEM and MOTIVATOR models propose how these processes are integrated to generate appropriate learned behaviors in response to unique environmental challenges (Table 1), especially when they are combined with perceptual, cognitive, and motor circuits in larger systems. Such larger systems have been embodied in adaptive mobile robots; e.g., Baloch and Waxman (1991).

19.5. Breakdowns during mental disorders: Theory of Mind, autism, and schizophrenia

When a CogEM circuit functions improperly, symptoms of mental disorders can result. For example, hypoactivity of amygdala or orbitofrontal cortex can prevent a cognitive-emotional resonance from occurring, thereby causing failures in Theory of Mind processes (Baron-Cohen, 1989; Perner, Frith, Leslie, & Leekam, 1989) in both autism and schizophrenia (Grossberg, 2000b; Grossberg & Seidman, 2006). Such failures include problems with activating motivationally directed goals and intentions. This happens in CogEM when a depressed emotional response in the amygdala also depresses the incentive motivational signals needed to activate prefrontal cortex in response to motivationally salient events (Fig. 36). The prefrontal cortex will then not adequately activate, and a hypofrontal condition can emerge during which working memory representations and plans are degraded, so social goals and plans will not form normally.

Depressed emotional responses to environmental and internally generated cues, including facial expressions, combined with insufficient motivational support for emotionally-appropriate plans and actions, helps to explain why individuals with autism may be unable to understand others' actions and may themselves

perform actions that are socially inopportune. Deficiencies in Theory of Mind are not, however, sufficient to explain all aspects of autistic development and behavior, as Grossberg and Seidman (2006) review.

19.6. ARTSCAN Search: Solving the Where's Waldo problem

Survival circuits can solve the Where's Waldo Problem by being able to search for and recognize important goal objects in multi-object scenes, including read-out of sensory expectations whose matching by acquired targets can elicit a conscious percept of the target, before launching appropriate goal-oriented actions towards them. The ARTSCAN Search model (Figs. 17 and 18; Chang et al., 2014) proposes how this may happen.

ARTSCAN Search shows how object categories that are learned with view-, size-, and *positionally-invariant* properties in the What cortical stream can activate *positionally-sensitive*, or-*dependent*, representations in the Where cortical stream that can be used to act upon a valued object in space. To accomplish this feat, ARTSCAN Search builds upon the ARTSCAN and CogEM models. In particular, the ARTSCAN model that was summarized in Section 12 used *Where-to-What* interactions to modulate the learning of invariant recognition categories. The ARTSCAN Search model supplements these *Where-to-What* interactions with *What-to-Where* interactions that can guide directed search and action towards valued goal objects whose representations are selectively amplified by motivated attention. ARTSCAN Search uses CogEM to find a desired target object in a scene through bottom-up stimulation by the scene (Figs. 17 and 38) and by a top-down motivated search (Figs. 18 and 39). ARTSCAN Search also simulates how "executive" prefrontal cognitive primes can drive such a search.

ARTSCAN Search can find Waldo by exploiting how ARTSCAN learns invariant recognition categories. Recall from Section 12 that ARTSCAN associates view-specific categories with view-invariant categories. The view-specific categories are also positionally-sensitive. Because this is the ARTSCAN model, top-down expectations are also learned from the view-invariant categories back to the view-sensitive categories. The pARTSCAN model generalizes this learning to view-, position-, and size-invariant categories, which can again send top-down priming signals to their view-specific and positionally-sensitive categories. These latter categories are a computational interface from an invariant category that represents Waldo to a target position for Waldo's current location, as summarized in Figs. 38 and 39.

ARTSCAN Search was tested by learning to recognize and search for CalTech 101 images of natural objects in scenes that contain non-overlapping objects, as well as by simulating reaction times measured in human search tasks (e.g., Brown & Denney, 2007). ARTSCAN Search cannot, however, quantitatively simulate how humans use *spatial and object contexts* to learn how to efficiently search for objects in a scene, as when seeing a refrigerator and sink leads to a higher expectation of next seeing a stove than a bed, and also where to look for the stove in a familiar kitchen. The ARTSCENE model (Grossberg & Huang, 2009) and ARTSCENE Search model (Huang & Grossberg, 2010) illustrate how humans accomplish these goals.

ARTSCENE explains how the gist of a scene can be rapidly learned and recognized as a large-scale texture category. Scene identity can be refined by scanning the scene with attentional shrouds to categorize finer texture categories within it. Recognition performance is then determined by using all the learned texture categories to vote for the best scenic label. The resonances that occur during such a recognition event is a type of feature-category resonance (Table 2). At the time ARTSCENE was published, it reduced the error rate of popular alternative scene classification models by 16.15%.

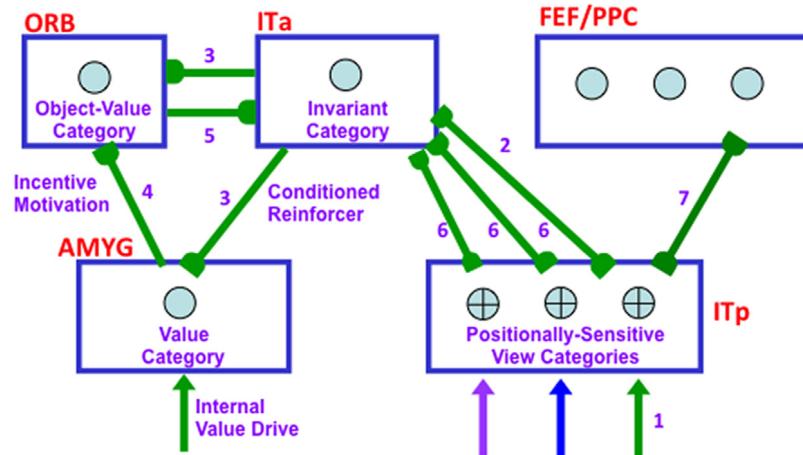


Fig. 38. How ARTSCAN Search models a cognitive–emotional resonance that can support search and orienting to a valued object in response to bottom-up processing of the object in a scene. The numbers indicate the order in which activity flows through the several modeled brain regions through time. 1. Multiple object inputs activate positionally-sensitive view-specific categories in posterior inferotemporal cortex (ITp). 2. The active ITp categories activate their invariant object categories in anterior inferotemporal cortex (ITa). 3. The active ITa categories try to activate the corresponding object–value categories in orbitofrontal cortex (ORB), but cannot, by themselves, do so. At the same time, the ITa categories try to activate the amygdala (AMYG) via a conditioned reinforcer pathway. 4. A value category in AMYG that receives a large internal value, or drive, input and a large conditioned reinforcer input from ITa wins the sensory-drive competition in the amygdala. This value category tries to activate ORB via an incentive motivational pathway. 5. Converging inputs to ORB from ITa and AMYG activate an object–value category there that activated the learned ITa-AMYG-ORB pathway. The active object–value category sends motivated feedback to the corresponding invariant category in ITa, thereby amplifying it with motivated attention and enabling it to win the competition for attended invariant categories. Closing this feedback loop enables a cognitive–emotional resonance to occur in ITa-AMYG-ORB, among other brain regions, thereby supporting “the feeling of what happens” and recognition of that feeling. 6. The winning ITa invariant category sends priming signals to all of its positionally-sensitive view-specific category representations in ITp. 7. The ITp category representation that also receives bottom-up external input can now win the competition among categories and activate a representation of its position in space that is represented in multiple brain areas, including frontal eye fields (FEF; see Brown et al. (2004)) and posterior parietal cortex (PPC; see Fazl et al. (2009)), thereby activating eye and other bodily movements towards the recognized and valued object in space.

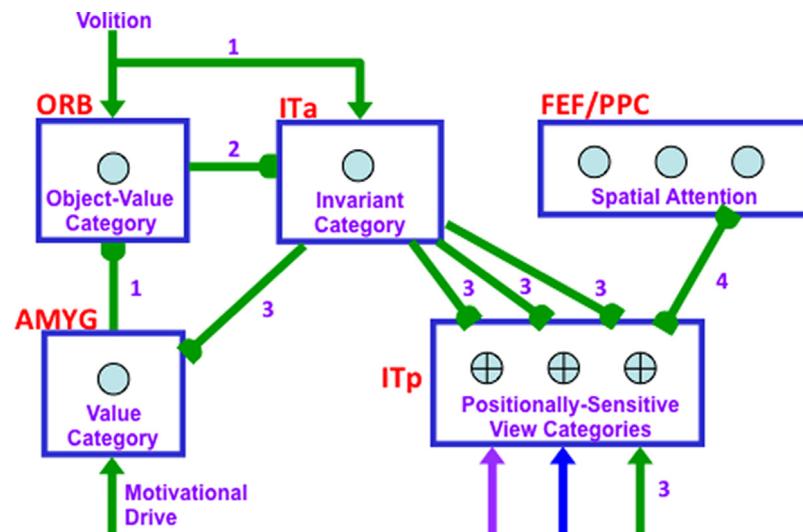


Fig. 39. How ARTSCAN Search models a cognitive–emotional resonance that can support search and orienting to a valued object in response to top-down value-driven search for the object. The numbers indicate the order in which activity flows through the several modeled brain regions through time. 1. An active value category in AMYG can prime a motivationally-compatible object–value category in ORB while a nonspecific volitional signal from the basal ganglia enables full activation of these cells. 2. As a result, when the object–value category fires, it sends top-down priming signals to the corresponding invariant object category in ITa. 3. ITa also receives the nonspecific volitional signal, which enables the primed ITa category to fire and activate all the corresponding positionally-sensitive view-specific category representations in ITp. 4. If one of these ITp categories also receives a bottom-up signal, then it is motivationally amplified, can win the competition among ITp categories, and activate a representation of its position in space in FEF, PPC, and related brain areas, thereby activating eye and other bodily movements towards it in space.

ARTSCENE Search goes further by incorporating spatial and object working memories, spatial and object list chunks, and spatial and object priming signals using interactions within the perirhinal and parahippocampal cortices, prefrontal cortex, inferotemporal cortex, and parietal cortex. The sequence-sensitive spatial and object list chunks guide context-sensitive searches using top-down expectation pathways that can support feature–category resonances when expected positions and objects are realized. ARTSCENE Search can quantitatively simulate key psychophysical data from experiments on contextual cueing, including spatial

and object cueing, positive and negative spatial cueing, and local and distant cueing effects, (e.g., Brockmole, Castelhano, & Henderson, 2006; Chun, 2000; Chun & Jiang, 1998; Jiang & Wagner, 2004; Lleras & von Mühlenen, 2004; Olson & Chun, 2002).

19.7. Conscious vs. non-conscious emotions

Many experiments have shown that emotional processes can be activated without causing conscious awareness or knowing about

them, yet these emotions can nonetheless influence subsequent behaviors. In the current theory, any stimulation that does not cause a vigorous and sustained cognitive-emotional resonance that includes, but is not restricted to, (temporal cortex)-amygdala-(orbitofrontal cortex) may lead to a non-conscious outcome. It may be due to insufficient sensory stimulation in terms of intensity or duration, attention-distracting tasks, or any other manipulation that prevents a full resonance from developing and being sustained. In particular, amygdala and hypothalamus can interact without necessarily leading to conscious awareness.

Tamietto and de Gelder (2010) have reviewed several different kinds of experimental evidence that led them to a similar viewpoint, but without mechanisms of adaptive resonance to derive mechanistic conclusions. They write that “a major difference between the two types of perception [conscious and non-conscious] may be the combined involvement of cortical areas and of cortico-subcortical interactions when stimuli are consciously perceived. *The dichotomy of conscious and non-conscious perception of emotional signals can thus be reformulated in neural terms as the integration of activity in subcortical and cortical structures*” [italics mine], where the amygdala figures prominently in their review of relevant emotional structures. Cognitive-emotional resonances in the CogEM and MOTIVATOR models link subcortical structures, such as the amygdala and hypothalamus, with cortical structures, such as the temporal and orbitofrontal cortices, into a bound state that supports both the feeling of what happens, and the knowledge of what event has caused it.

19.8. Conditioning and consciousness: Trace conditioning, hippocampus, and time cells

The need for a sufficiently sustained cognitive-emotional resonance to generate a conscious emotion is illustrated by trace conditioning experiments in humans. These experiments describe a link between consciousness and conditioning, including a role for hippocampus in enabling conditioning to bridge a temporal gap between a CS and a subsequent US. Such a temporal gap exists during *trace conditioning*, during which the CS terminates before the US begins, but not *delay conditioning*, where the CS and US overlap in time. Correspondingly, amnesic patients with hippocampal damage learn at a normal rate during delay conditioning, but not during trace conditioning. Clark and Squire (1998, p. 79) postulated that normal humans acquire trace conditioning because they have intact declarative or episodic memory and, therefore, can demonstrate conscious knowledge of a temporal relationship between CS and US: “trace conditioning requires the acquisition and retention of conscious knowledge [and] would require the hippocampus and related structures to work conjointly with the neocortex”.

Various other studies have linked consciousness and conditioning (Gabrieli et al., 1995; McGlinchey-Berroth, Brawn, & Disterhoft, 1999; McGlinchey-Berroth, Carrillo, Gabrieli, Brawn, & Disterhoft, 1997). For example, trace conditioning is facilitated by conscious awareness in normal control subjects, whereas delay conditioning is not, and amnesics with bilateral hippocampal lesions perform at a success rate similar to unaware controls for both delay and trace conditioning (Clark, Manns, & Squire, 2001). Amnesics were unaware of experimental contingencies, and poor performers on trace conditioning (Clark & Squire, 1998).

The nSTART, or neurotrophic Spectrally Timed ART, model (Fig. 40; Franklin and Grossberg (2008, 2016)) simulates these experimental results by embedding the CogEM model into a larger learning circuit that includes the hippocampus. The hippocampus in this model includes a circuit for adaptively timed learning, called a *spectral timing* circuit in the earlier START model (Grossberg & Merrill, 1992, 1996; Grossberg & Schmajuk, 1989). Such a spectral timing circuit shows how a population of cells with differently timed responses (the “spectrum”), none of which can

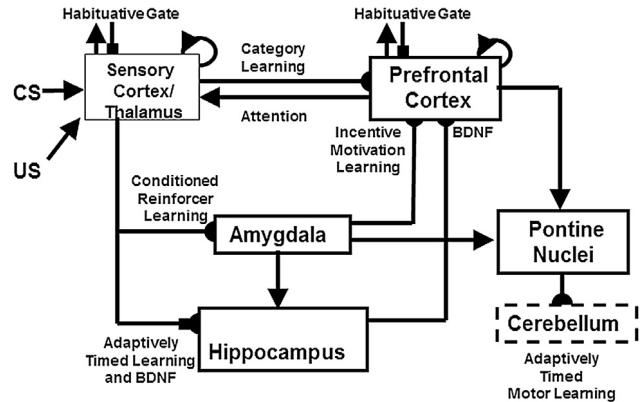


Fig. 40. Neurotrophic Spectrally Timed ART, or nSTART, model macrocircuit. The nSTART model augments CogEM with a hippocampal circuit that supports adaptively timed conditioning, thereby enables nSTART to simulate both delay and trace conditioning. The sensory cortex and thalamus send parallel pathways to the amygdala, for conditioned reinforcer learning, and to the hippocampus, for adaptively timed and BDNF-modulated learning, where BDNF stands for Brain-Derived Neurotrophic Factor and prolongs the effect of spectrally timed learning. The thalamic and sensory cortical circuits share many anatomical properties, so the nSTART model lumps the thalamus and sensory cortex together, for simplicity. Both the amygdala and the hippocampus provide learned incentive motivational feedback to the prefrontal (orbitofrontal) cortex, where again BDNF prolongs the duration of hippocampal feedback. This hippocampal feedback loop enables the model to bridge temporal gaps between CS and US during conditioning, and to do so in an adaptively timed way. BDNF also supports long-term memory consolidation within the thalamo-cortical, cortico-cortical, and cortico-thalamic pathways that subserve recognition and attention. In thalamus, sensory cortex, prefrontal cortex, and hippocampus, activity-dependent habituative transmitter gates modulate excitatory conductances. The pontine nuclei bridge signals from thalamus, amygdala and neocortex to cerebellar circuits, where adaptively timed motor learning takes place. The cerebellar motor learning is not simulated in nSTART, but see Fiala et al. (1996) and Grossberg and Merrill (1996) for examples of adaptively-timed cerebellar motor learning Key: an arrowhead indicates an excitatory synapse; a learned weight is indicated by a hemidisk; a synapse with a habituative gate is represented by a square; and a synapse with a habituative gate that is followed by a learned weight is represented as a square followed by a hemidisk.

Source: Reprinted with permission from Franklin and Grossberg (2016).

individually time a response over a long time interval (Fig. 41, bottom row), can together bridge the temporal gap between CS offset and US onset during trace conditioning (Fig. 41, top row), as well as during other learning paradigms wherein temporal gaps occur. This hippocampal timing circuit enables associations to form across such a temporal gap and become consolidated, while supporting a sustained cognitive-emotional resonance that underlies motivated attention, core consciousness, and “the feeling of what happens”.

The existence of a hippocampal spectral timing circuit that can support learning across CS-US temporal gaps is supported by psychological data (e.g., Gibbon, 1991; Millenson, Kehoe, & Gormezano, 1977; Smith, 1968) and neurophysiological data (e.g., Berger & Thompson, 1978; MacDonald, Lepage, Eden, & Eichenbaum, 2011), notably its *Weber law* property that larger inter-stimulus intervals between CS and US lead to learned response curves with broader variances (Fig. 41, top row), a property that is also called *scalar timing* in the psychological literature (Gibson, 1977), and has also been found in hippocampal neurophysiological data about “time cells” that together form a spectrum of differently timed cells that obey a Weber law (MacDonald et al., 2011), as predicted by the START model.

These theoretical results suggest that, contrary to Clark and Squire (1998), episodic memory may not be necessary to consciously experience emotions. On the other hand, the hippocampus is indeed part of the brain system that supports episodic learning and memory (Eichenbaum & Lipton, 2008). Both spatial and temporal representations coexist in the hippocampus, and both space and time are combined in episodic memories (Tulving, 1972).

Recent neural modeling proposes why spatial and temporal representations coexist in the entorhinal–hippocampal system by showing how they exploit similar circuit mechanisms to represent the large spaces and times that control spatial and adaptively timed behaviors (Grossberg & Pilly, 2012, 2014), a discovery that has been summarized with the phrase “neural relativity”. The homologous model circuits in the entorhinal–hippocampal cortices that represent space and time are called the Spectral Spacing and Spectral Timing models, respectively. The Spectral Spacing model shows how entorhinal grid cells and hippocampal place cells may be learned and control spatial navigation, among other spatial behaviors. The Spectral Timing circuit shows how adaptively timed reinforcement learning and motivated behaviors can be achieved. Thus, whereas the Spectral Timing circuit may help to sustain a cognitive–emotional resonance, the full resources of episodic memory may not be needed to support consciousness during trace conditioning, especially in conditioning paradigms where no spatial navigation occurs.

The predicted mechanistic homolog between spatial navigation and adaptively-timed conditioning circuits in the entorhinal–hippocampal system has quantitatively simulated challenging data in both domains, but also suggests new testable hypotheses. For example, the ability of Spectral Timing circuits to bridge long time intervals has been attributed to the metabotropic glutamate receptor (mGluR) system. Does such a biochemical substrate give rise to hippocampal time cells? Does a similar biochemical substrate enable Spectral Spacing to represent large spaces via entorhinal grid cells?

19.9. Why a Weber law? Distinguishing expected and unexpected disconfirmations

What is the functional utility of the Weber law property during behavior? Adaptively timed responding is essential for terrestrial animals that actively explore and learn about extended environments, since rewards and other goals are often delayed in time relative to the environmental cues that predict their future occurrence. The Spectral Timing model accomplishes this by predicting how the brain distinguishes *expected disconfirmations*, also called *expected non-occurrences*, of reward, which should not be allowed to interfere with acquiring a delayed reward, from *unexpected disconfirmations*, also called *unexpected non-occurrences*, of reward, which can trigger the usual consequences of predictive failure, including reset of working memory, attention shifts, frustrative reactions that can counter-condition unsuccessful cognitive–emotional associations, and the release of exploratory behaviors to find better sources of valued goal objects.

How does this distinction between expected and unexpected disconfirmations work in practice? Key questions that must be answered to understand this distinction include: What spares an animal from erroneously reacting to expected non-occurrences of reward as predictive failures? Why does an animal not become so frustrated by the immediate non-occurrence of a reward that it prematurely shifts its attentional focus and releases exploratory behavior aimed at finding the desired reward somewhere else, leading to relentless exploration for immediate gratification? Alternatively, if the animal does wait, but the reward does not appear at the expected time, then how does the animal then react to the unexpected non-occurrence of the reward by resetting its working memory, shifting its attention, becoming frustrated, and releasing exploratory behavior?

The key insight is that the process of registering ART-like sensory matches is not inhibited during either expected or unexpected non-occurrences within a sensory-cognitive ART circuit (Fig. 13). In either case, if the reward happened to appear earlier than expected, the animal could still perceive it during a cognitive–emotional resonance and release a consummatory response. Rather than directly

interfering with the matching process, the *effects of mismatches* upon reinforcement, attention, and exploration are prevented by inhibiting the orienting system (triangular region with vigilance parameter ρ inside it in Fig. 13). In other words, one role of the spectral timing circuit in the hippocampus is to inhibit the ART orienting system in the hippocampus (Fig. 13(c)) during an expected non-occurrence, and with it the arousal that would otherwise have led to undesired cognitive, emotional, and motor consequences.

Because a spectral timing response begins immediately after its triggering stimulus, and builds throughout the interstimulus interval, or ISI, between the CS and US (Fig. 41, top row), it can maintain inhibition of the orienting system until the expected time of occurrence of the reinforcing stimulus. This inhibition would not, however, occur in response to an unexpected non-occurrence because the spectral timing circuit would not be active then (Grossberg & Merrill, 1992, 1996; Grossberg & Schmajuk, 1989). As a result, an unexpected non-occurrence could lead to the usual sequence of cognitive, emotional, and motor responses to correct the predictive error.

The Weber law property of a hippocampal Spectral Timing circuit thus realizes three critical functional roles: By beginning right after its inducing stimulus, it can inhibit spurious orienting responses throughout its adaptively timed interval. By exciting orbitofrontal object–value categories (Fig. 40), it can also maintain motivated attention upon the correct orbitofrontal representation throughout this time interval. Finally, by peaking at the expected time of occurrence of the reinforcing cue, it can generate a peak amplitude of motivated attention at the correct time at the orbitofrontal representation that controls learned responses.

19.10. Effects of hippocampal, amygdala, and orbitofrontal lesions on learning and memory

In addition to explaining data about normal delay and trace conditioning, the nSTART model explains and simulates many subtle data about how learning and memory consolidation are influenced by different brain lesions (Franklin & Grossberg, 2016). In particular, the model proposes a unified explanation, and computer simulations, of why lesions of amygdala, hippocampus, and cortex have differential effects depending on the phase of learning when they occur. In particular, it explains why the hippocampus is typically needed for trace conditioning, but not delay conditioning, and what the exceptions reveal; why amygdala lesions made before or immediately after training decelerate conditioning while those made later do not; why thalamic or sensory cortical lesions degrade trace conditioning more than delay conditioning; why hippocampal lesions during trace conditioning experiments degrade recent but not temporally remote learning; why orbitofrontal cortical lesions degrade temporally remote but not recent or post-lesion learning; why temporally graded amnesia is caused by ablation of prefrontal cortex after memory consolidation; and how neurotrophins, notably Brain-Derived Neurotrophic Factor (BDNF), influence memory formation and consolidation. This unified explanation also overcomes problems of alternative memory consolidation models, and incorporates many new testable predictions about how memory consolidation works both in normal individuals and clinical patients.

20. Some other approaches to consciousness

The field of consciousness studies has exploded since the seminal article of Crick and Koch (1990), and boasts its own journals and conferences. Crick and Koch (1990) described two forms of consciousness “a very fast form, linked to iconic memory...; and a slower one [wherein] an attentional mechanism transiently binds together all those neurons whose activity relates

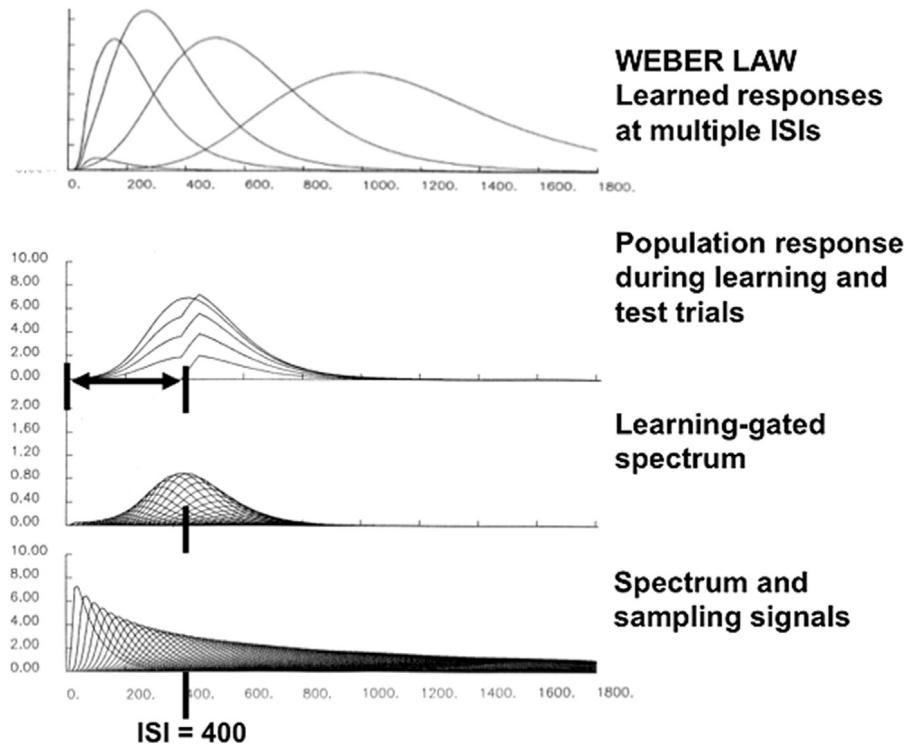


Fig. 41. (Bottom row) A spectrum of cell activations in response to the same CS input stimulus. Each cell in the spectrum responds at a different time. Each cell response acts as a sampling signal that enables its LTM trace to learn from a US only at times when the sampling signal is positive. (Third row) After conditioning trials, multiplying the spectral activities with their LTM traces leads to adaptive cell responses that tend to cluster near the interstimulus interval, or ISI, when the US was turned on. (Second row) When all of these adaptive cell responses are added to generate a population response, that population response is well-timed to peak at the ISI. The population response is shown after each of the first four learning trials and on the first recall trial. (Top row) When different ISIs occur, the population responses peak at their respective ISIs and have a width that scales with their ISI. This is the Weber law. When the population responses at multiple ISIs are shown, a non-zero optimal ISI can be seen with attenuated learning at both smaller and larger ISIs.
Source: Simulations adapted from Grossberg and Schmajuk (1989).

to the relevant features of a single visual object". This conclusion was consistent with available results about ART, but did not offer a linking hypothesis between brain dynamics and the perceptual, cognitive, and cognitive-emotional representations whose resonances support different conscious qualia. A great deal of additional experimental evidence for neural correlates of consciousness has been reported since 1990, but has typically led to theoretical conclusions that fail to make the crucial linking hypothesis between specific dynamical brain representations and specific conscious qualia (e.g., Baars, 2005; Dehaene, 2014; Dennett, 1991; Edelman & Tononi, 2000; Koch, Massimini, Boly, & Tononi, 2016; Tononi, 2004, 2015).

For example, the *neural global workspace* of Dehaene (2014), which builds upon the *global workspace* of Baars (2005), claims that "consciousness is global information broadcasting within the cortex [to achieve] massive sharing of pertinent information throughout the brain" (p. 13). Dehaene also makes a number of other useful observations, including that "the time that our conscious vision spends entertaining an interpretation is directly related to its likelihood, given the sensory evidence received" (p. 97) and that "the conditioning paradigm suggests that consciousness has a specific evolutionary role: learning over time, rather than simply living in the instant. The system formed by the prefrontal cortex and its interconnected areas, including the hippocampus, may serve the essential role of bridging temporal gaps" (p. 103). Such claims are consistent with the analyses in this article, but they do not describe the underlying organizational principles, neural mechanisms, or brain representations that embody subjective conscious aspects of experience.

In particular, the claim about "global information broadcasting" is consistent with analyses of how adaptive resonances, such as

surface-shroud and feature-category resonances, may synchronize the dynamics of multiple brain regions, but does not provide insights into the contents of conscious experiences, or the different kinds of resonances that support them, as summarized by Table 2. The observation that links conscious vision to likelihood is clarified by properties of the shunting competitive networks wherein winning conscious resonances are chosen, whether at the level of surface-shroud or stream-shroud resonances for representing individual percepts (Table 2; Sections 10 and 15.4) or of higher-order decisions about the most predictive list categories in a Masking Field (Figs. 31 and 33) for representing the most likely interpretation of a recent sequences of events, and the best prediction for what will happen next. The comment about the role of hippocampal–prefrontal interactions during conditioning does not provide insight into how or why spectrally timed learning occurs, notably the distinction between expected and unexpected disconfirmations during terrestrial exploration and learning, or the functional role, or neural mechanisms, that cause the Weber law property (Sections 19.8 and 19.9).

Continuing in the spirit of Edelman and Tononi (2000), Tononi (2004) defined a scalar function Φ , "the quantity of consciousness available to a system... as the value of a complex of elements. Φ is the amount of causally effective information that can be integrated across the informational weakest link of a subset of elements..." Tononi (2012, 2015) goes on to further develop postulates for this Integrated Information Theory (IIT) for physical systems that include consciousness, such as *intrinsic existence*, *compositionality*, *information*, *integration*, and *exclusion*. These postulates summarize some basic facts about consciousness, but do not explain them. For example, the *integration* postulate claims that each experience is unified and cannot be reduced to independent components,

much as an adaptive resonance is a bound state. The *exclusion* postulate claims that every experience is limited to particular things and not others, and has its own spatio-temporal grain, much as a resonance develops through time, focuses attention on some feature combinations while inhibiting others, and may be adaptively timed by spectral timing circuits. The *information* postulate claims that selective past causes and selective future effects are represented, much as a cognitive-emotional resonance can discover which combinations of features are predictive and which are accidental and attentionally blocked. And so on. Although these postulates tangentially touch upon some of the themes that the current article discusses, they have not demonstrated an ability to parametrically explain significant psychological or neurobiological databases. The current article provides ample evidence that no scalar-valued function can explain the brain principles, neural mechanisms, and dynamical events that give rise to specific conscious experiences.

Both Dehaene and Tononi used the word “information” as a critical component of their hypotheses. But what is “information”? The scientific concept of “information” in the mathematical sense of Information Theory (Shannon, 1948a, 1948b) requires that a set of *states* exist whose “information” can be computed, and that fixed probabilities exist for transitions between these states. In contrast, the brain is a self-organizing system that continually *creates new states* through development and learning, and whose probability structure is continually changing along with them. Without a theory that explains how these states arise, and how their transition probabilities may change through time in response to changing environmental statistics and internal representations thereof, the classical concept of information is useless. How such states arise is a key explanatory target of ART, and is one reason why ART can offer a classification of the resonances that are proposed to embody specific conscious experiences.

Dennett's highly cited book called *Consciousness Explained* (Dennett, 1991) argued against a Cartesian Theater model, a place in the brain where “it all comes together” and generates subjective judgments. Instead, Dennett advocated a Multiple Drafts model where discriminations are distributed in space and time across the brain, a concept that, without further elaboration, is too vague to have explanatory power. To the extent that Dennett supported his claim with a discussion of visual percepts, he has made some serious mistakes. For example, when discussing surface filling-in (see Section 6.5), he claimed: “The fundamental flaw in the idea of ‘filling-in’ is that it suggests that the brain is providing something when in fact the brain is ignoring something” (p. 356). “The brain doesn't have to ‘fill in’ for the blind spot...We don't notice these gaps, but they don't have to be filled in *because we're designed not to notice them*” (p. 355). In other words, Dennett argued that a physical process of filling-in does not occur. Given that Dennett put an example of neon color spreading on the back cover of his book, he clearly viewed this claim as an important part of his proposals about consciousness.

Dennett's proposal explained no data about how filling-in works and what its properties are, despite the fact that he learned about neon color spreading at one of my lectures about brightness and color perception where many properties of surface filling-in were explained, including model simulations of many perceptual data, including data about neon color spreading, that still have no explanation without invoking filling-in. Instead of offering explanations, Dennett offered sarcasm: “This idea of filling in is common in the thinking of even sophisticated theorists, and it is a dead giveaway to vestigial Cartesian materialism. What is amusing is that those who use the term often know better, but since they find the term irresistible, they cover themselves by putting in scare-quotes” (p. 344). In fact, many perceptual and neurobiological data support that filling-in occurs (e.g., Paradiso &

Nakayama, 1991; Takeichi, Shimojo, & Watanabe, 1992) and it is fully consistent with a Multiple Drafts perspective.

Various other authors have made interesting comments about consciousness based upon properties of perceptual and neurobiological data. Notable among them are Victor Lamme and his colleagues (e.g., Keizer, Hommel, & Lamme, 2015; Tsuchiya, Wilke, Frassle, & Lamme, 2016). However, these contributions have not proposed organizational principles or brain mechanisms whereby to understand the brain dynamics of individual conscious experiences, including how they may share similar principles of resonant dynamics, yet also occur in different brain regions where they embody solutions to environmental constraints that differentiate seeing, hearing, feeling, and knowing.

Perhaps the theory of Damasio (1999) comes closest to theoretically linking brain to mind by providing what is, in effect, a heuristic derivation of the CogEM model to explain his clinical data about cognitive-emotional interactions (Section 19.4). But this theory provided no mechanistic account, could therefore provide no data simulations, and did not situate this heuristic derivation within a larger theory of how brain resonances and consciousness may be linked.

21. Concluding remarks

21.1. Unified experiences of perceiving, knowing, and feeling

To elaborate the general prediction that “all conscious events are resonant events”, this article proposes how specific neural circuits and systems in different parts of the brain generate adaptive resonances that support conscious awareness and knowledge about external sensory inputs, such as specific sights and sounds, or internal sensory inputs, such as specific emotions. Each resonance supports an attentional focus upon the perceptual or affective representation that becomes conscious, and synchronizes with related attentive resonances that enable conscious recognition, or knowing, of the perceptual or affective event. Spatial, object, and motivated attention all play a role in these events. The article also suggests how conscious perceptual, cognitive, and affective representations can all resonate together, so that individuals can consciously know and experience feelings about what they consciously perceive (e.g., Sections 13 and 19). Each of the resonances that support different kinds of conscious experiences occur at brain interfaces that can trigger control of appropriate actions (Table 1), notably actions that lead to the closing of brain-environment perception/cognition/emotion-action loops.

Many psychological and neurobiological data have been explained and predicted using properties of these resonances, in both normal individuals and clinical patients. This article has unified the explanations of many previously treated data, explained additional data for the first time using its newly discovered concepts, and made new testable predictions. Many more data are explained in the original archival articles that introduced various of the synthesized concepts about vision, audition, emotion, and recognition. As such resonant dynamics unfold through time, they are bound together into a unified sense of self with the help of the coordinated opening of multiple cognitive and motoric gates by the basal ganglia. More modeling work needs to be done to demonstrate such basal ganglia coordination, in the spirit of, but moving beyond, the simulations of the lisTELOS model (Section 18.2).

Each resonance may involve multiple brain areas, and not all cells within a resonance may support conscious qualia. Only cells that can represent the features that discriminate perceived differences can do so. A conscious experience may not always be triggered even by active cells that can support such discriminations if inputs to a circuit that is capable of resonance are too brief or insufficiently intense, or these inputs unsuccessfully compete for attention across cell populations that are more fully activated. The activated cells can nonetheless often non-consciously influence future behaviors.

2.1.2. Using self-organizing and self-stabilizing expertise to control changing bodies

ART explains how perceptual and cognitive processes in the What ventral processing stream use excitatory matching and match-based learning (Table 4) to learn attentive representations of objects and events in the world. Match-based learning solves the stability-plasticity dilemma and can occur quickly without causing catastrophic forgetting, much as new faces can be learned quickly without forcing unselective forgetting of familiar faces. Excitatory matching can also support sustained resonance, and thus sometimes conscious awareness, when a good enough match with object information occurs.

Such match-based learning supports the creation of category representations at higher cortical levels that are increasingly invariant under changes in an object's views, positions, and sizes when it is registered on our retinas. That is, match-based learning can support *invariant category learning* (Section 12), which enables learning to categorize the world without causing a combinatorial explosion of memories. However, these properties are not sufficient for a brain to accomplish autonomous adaptation to a changing world, if only because positionally-invariant object category representations cannot, by themselves, be used to manipulate objects at particular positions in space.

Complementary spatial and motor processes in the Where/How dorsal cortical processing stream can be used to manipulate objects in space. To do so, they often use VAM-like inhibitory matching and mismatch learning (Table 4) to continually update spatial maps and sensory-motor gains as bodily parameters change through time (Sections 4, 14.1–14.4 and 18), whether for reaching or for speech production (e.g., Sections 14.1–14.4). These circuits cannot support an adaptive resonance under any conditions, and thus do not support conscious awareness.

Although each type of matching and learning in Table 4 is insufficient on its own to learn about the world and to effectively act upon it, together they can (Table 1). Perceptual and cognitive processes use excitatory matching and match-based learning to create self-stabilizing representations of objects and events that embody increasing expertise about the world, and conscious awareness of it. Complementary spatial and motor processes use inhibitory matching and mismatch learning to continually update spatial maps and sensory-motor gains to compensate for bodily changes throughout life. Together they provide a self-stabilizing perceptual and cognitive front end for conscious awareness and knowledge acquisition, which can intelligently manipulate more labile spatial and motor processes that enable our changing bodies to act effectively upon a changing world.

References

- Ackermann, J. (2008). Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, 31, 265–272.
- Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology*, 94, 2832–2843.
- Agam, Y., Galperin, H., Gold, B. J., & Sekuler, R. (2007). Learning to imitate novel motion sequences. *Journal of Vision*, 7, <http://dx.doi.org/10.1167/7.5.1>.
- Aggleton, J. P. (1993). The contribution of the amygdala to normal and abnormal emotional states. *Trends in Neurosciences*, 16, 328–333.
- Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences USA*, 90, 5718–5722.
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387, 401–406.
- Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, 9, 357–381.
- Ames, H., & Grossberg, S. (2008). Speaker normalization using cortical strip maps: A neural model for steady state vowel categorization. *Journal of Acoustical Society of America*, 124, 3918–3936.
- Amis, G., Carpenter, G. A., Ersoy, B., & Grossberg, S. (2009). Cortical learning of recognition categories: A resolution of the exemplar vs. prototype debate. CAS/CNS TR-2009-002.
- Andersen, R. A., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230, 456–458.
- Andersen, R., Essick, G., & Siegel, R. (1987). Neurons of area 7 activated by both visual stimuli and oculomotor behavior. *Experimental Brain Research*, 67, 316–322.
- Andersen, R. A., & Mountcastle, V. B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *The Journal of Neuroscience*, 3, 532–548.
- Armony, J. L., Aube, W., Angulo-Perkins, A., Peretz, I., & Concha, L. (2015). The specificity of neural responses to music and their relation to voice processing: An fMRI study. *Neuroscience Letters*, 593, 35–39.
- Auer, E. T., Jr., & Luce, P. A. (2008). Probabilistic phonotactics in spoken word recognition. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception*. Wiley Online Library, <http://dx.doi.org/10.1002/9780470757024.ch25>.
- Averbach, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences, USA*, 99(20), 13172–13177.
- Averbach, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003a). Neural activity in prefrontal cortex during copying geometrical shapes. I. Single cells encode shape, sequence, and metric parameters. *Experimental Brain Research*, 150, 127–141.
- Averbach, B. B., Crowe, D. A., Chafee, M. V., & Georgopoulos, A. P. (2003b). Neural activity in prefrontal cortex during copying geometrical shapes. II. Decoding shape segments from neural ensembles. *Experimental Brain Research*, 150, 142–153.
- Baars, B. (2005). Global workspace theory of consciousness: Towards a cognitive neuroscience of human experience? *Progress in Brain Research*, 150, 45–53.
- Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formation dynamics. *Vision Research*, 37, 3037–3059.
- Baloch, A. A., Grossberg, S., Mingolla, E., & Nogueira, C. A. M. (1999). A neural model of first-order and second-order motion perception and magnocellular dynamics. *Journal of the Optical Society of America A*, 16, 953–978.
- Baloch, A. A., & Waxman, A. M. (1991). Visual learning, adaptive expectations, and behavioral conditioning of the mobile robot MAVIN. *Neural Networks*, 4, 271–302.
- Banquet, J. P., & Grossberg, S. (1987). Probing cognitive processes through the structure of event-related potentials during learning: An experimental and theoretical analysis. *Applied Optics*, 26, 4931–4946.
- Barbas, H. (1995). Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, 19, 499–510.
- Baron-Cohen, S. (1989). The autistic child's theory of the mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 30, 285–297.
- Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D., & Murray, E. A. (2000). Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *Journal of Neuroscience*, 20, 4311–4319.
- Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology*, 14, 212–217.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312.
- Bellmann, A., Meuli, R., & Clarke, S. (2001). Two types of auditory neglect. *Brain*, 124, 676–687.
- Berger, T. W., & Thompson, R. F. (1978). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response, I: The hippocampus. *Brain Research*, 145, 323–346.
- Berke, J. D., Hetrick, V., Breck, J., & Green, R. W. (2008). Transient 23–30-Hz oscillations in mouse hippocampus during exploration of novel environments. *Hippocampus*, 18, 519–529.
- 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.
- Besson, M., Faita, F., Peretz, I., Bennel, A.-M., & Requin, J. H. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, 9, 494–498.
- Bhatt, R., Carpenter, G., & Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Research*, 47, 3173–3211.
- Bisley, J., Krishna, B., & Goldberg, M. (2004). A rapid and precise on-response in posterior parietal cortex. *Journal of Neuroscience*, 24, 1833–1838.
- Boardman, I., & Bullock, D. (1991). A neural network model of serial order recall from short-term memory. In: *Proceedings of the international joint conference on neural networks, IJCNN* (pp. 879–884). Vol. II.
- Boardman, I., Grossberg, S., Myers, C., & Cohen, M. (1999). Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception and Psychophysics*, 6, 1477–1500.
- Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, 411, 798–801.
- Bonnell, A.-M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception and Psychophysics*, 63, 1201–1213.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178.
- Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, 13, 767–782.
- Bradski, G., Carpenter, G. A., & Grossberg, S. (1994). STORE working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, 71, 469–480.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

- Bricat, S. L., & Miller, E. H. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nature Neuroscience*, 18, 576–581.
- Britt, R., & Starr, A. (1976). Synaptic events and discharge patterns of cochlear nucleus cells: Frequency modulated tones. *Journal of Neurophysiology*, 39, 179–194.
- Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 699–706.
- Bronkhorst, A. W., & Plomp, R. (1988). The effect of head-induced interaural time and level differences on speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 83, 1508.
- Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, 17, 471–510.
- Brown, J. M., & Denney, H. I. (2007). Shifting attention into and out of objects: Evaluating the processes underlying the object advantage. *Perception and Psychophysics*, 69, 606–618.
- Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2, 1–19.
- Buffalo, E. A., Fries, P., Landman, R., Buschman, T. J., & Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences*, 108, 11262–11267.
- Bullock, D., Cisek, P., & Grossberg, S. (1998). Cortical networks for control of voluntary arm movements under variable force conditions. *Cerebral Cortex*, 8, 48–62.
- Bullock, D., & Grossberg, S. (1988). Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review*, 95, 49–90.
- Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, 5, 408–435.
- Burr, D. C., & Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 365, 504–515.
- Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, 315, 1860–1862.
- Buschman, T. J., & Miller, E. K. (2009). Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations. *Neuron*, 63, 386–396.
- 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.
- Cao, Y., & Grossberg, S. (2012). Stereopsis and 3D surface perception by spiking neurons in laminar cortical circuits: A method of converting neural rate models into spiking models. *Neural Networks*, 26, 75–98.
- Cao, Y., Grossberg, S., & Markowitz, J. (2011). How does the brain rapidly learn and reorganize view- and positionally-invariant object representations in inferior temporal cortex? *Neural Networks*, 24, 1050–1061.
- Caplovitz, G. P., & Tse, P. U. (2007). V3A processes contour curvature as a trackable feature for the perception of rotational motion. *Cerebral Cortex*, 17, 1179–1189.
- Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, 37, 54–115.
- Carpenter, G. A., & Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.
- Carpenter, G. A., & Grossberg, S. (1993). Normal and amnesia learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16, 131–137.
- Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research*, 40, 1203–1215.
- Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, 14, 147–153.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200–219.
- Chalmers, D. J. (2000). What is a neural correlate of consciousness? In T. Metzinger (Ed.), *Neural correlates of consciousness: Empirical and conceptual questions*. Cambridge, MA: MIT Press.
- Chang, H.-C., Grossberg, S., & Cao, Y. (2014). Where's Waldo? How perceptual cognitive, and emotional brain processes cooperate during learning to categorize and find desired objects in a cluttered scene. *Frontiers in Integrative Neuroscience*, <http://dx.doi.org/10.3389/fnint.2014.0043>.
- 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, 761–774.
- Chey, J., Grossberg, S., & Mingolla, M. (1997). Neural dynamics of motion grouping: From aperture ambiguity to object speed and direction. *Journal of the Optical Society of America*, 14, 2570–2594.
- Chiu, Y. C., & Yantis, S. (2009). A domain-independent source of cognitive control for task sets: shifting spatial attention and switching categorization rules. *The Journal of Neuroscience*, 29, 3930–3938.
- Chomsky, N. (1957). *Syntactic structures*. The Hague/Paris: Mouton.
- Chomsky, N. (1959). Reviews: Verbal behavior by B. F. Skinner. *Language*, 35, 26–58.
- Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4, 170–178.
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, 36, 28–71.
- Cisek, P., Grossberg, S., & Bullock, D. (1998). A cortico-spinal model of reaching and proprioception under multiple task constraints. Boston University. *Journal of Cognitive Neuroscience*, 10, 425–444.
- Clark, R. E., Manns, J. R., & Squire, L. R. (2001). Trace and delay eyeblink conditioning: contrasting phenomena of declarative and nondeclarative memory. *Psychological Science*, 12, 304–308.
- Clark, R. E., & Squire, L. R. (1998). Classical conditioning and brain systems: The role of awareness. *Science*, 280, 77–81.
- Clarke, S., & At, A. (2013). Sound localization disorders and auditory neglect. In G. G. Celesia (Ed.), *Handbook of clinical neurophysiology*: 10 (pp. 423–434). Amsterdam: Elsevier.
- Cohen, M. A., & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception and Psychophysics*, 36, 428–456.
- Cohen, M. A., & Grossberg, S. (1997). Parallel auditory filtering by sustained and transient channels separates coarticulated vowels and consonants. *IEEE Transactions on Speech and Audio Processing*, 5, 301–318.
- Cohen, M. A., Grossberg, S., & Stork, D. G. (1988). Speech perception and production by a self-organizing neural network. In Y. C. Lee (Ed.), *Evolution, learning, cognition, and advanced architectures* (pp. 217–231). Hong Kong: World Scientific Publishers.
- Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of Acoustical Society of America*, 98, 862–879.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of a pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207–210.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences*, 2, 263–275.
- Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, 375, 121–123.
- Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in The Making of Consciousness*. New York: Houghton Mifflin Harcourt.
- Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology*, 159, 203–221.
- Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: The role of interaural time differences. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 617–629.
- Deadwyler, S. A., West, M. O., & Lynch, G. (1979). Activity of dentate granule cells during learning: Differentiation of perforant path inputs. *Brain Research*, 169, 29–43.
- Deadwyler, S. A., West, M. O., & Robinson, J. H. (1981). Entorhinal and septal inputs differentially control sensory-evoked responses in the rat dentate gyrus. *Science*, 211, 1181–1183.
- Dehaene, S. (2014). *Consciousness and the brain*. New York, NY: Viking Press.
- Delgutte, B., & Kiang, N. Y. S. (1984a). Speech coding in the auditory nerve, I: Vowel-like sounds. *Journal of Acoustical Society of America*, 75, 866–878.
- Delgutte, B., & Kiang, N. Y. S. (1984b). Speech coding in the auditory nerve, III: Voiceless fricative consonants. *Journal of Acoustical Society of America*, 75, 887–896.
- Delgutte, B., & Kiang, N. Y. S. (1984c). Speech coding in the auditory nerve, IV: Sounds with consonant like dynamic characteristics. *Journal of Acoustical Society of America*, 75, 897–906.
- Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron*, 37, 347–359.
- Dennett, D. (1991). *Consciousness explained*. Boston, MA: Little, Brown.
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London*, 353, 1245–1255.
- Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, 57, 835–868.
- Deutsch, D. (1975). Two-channel listening to musical scales. *Journal of Acoustical Society of America*, 57, 1156–1160.
- Deutsch, D. (Ed.) (2013). *The psychology of music* (3rd ed.). Amsterdam: Elsevier.
- DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neurosciences*, 11, 219–226.
- Dranias, M., Grossberg, S., & Bullock, D. (2008). Dopaminergic and non-dopaminergic value systems in conditioning and outcome-specific revaluation. *Brain Research*, 1238, 239–287.
- Dresp, B., Durand, S., & Grossberg, S. (2002). Depth perception from pairs of overlapping cues in pictorial displays. *Spatial Vision*, 15, 255–276.
- Dresp-Langley, B., & Grossberg, S. (2016). Neural computation of surface border ownership and relative surface depth from ambiguous contrast inputs. *Frontiers in Psychology*, <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.01102/full>.
- Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience*, 1, 17–22.
- Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, 255, 90–92.
- Dunbar, G. (2012). Adaptive Resonance Theory as a model of polysemy and vagueness in the cognitive lexicon. *Cognitive Linguistics*, 23, 507–537.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517.
- Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.

- Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, 123, 161–177.
- Eichenbaum, H., & Lipton, P. A. (2008). Towards a functional organization of the medial temporal lobe memory system: Role of the parahippocampal and medial entorhinal cortical areas. *Hippocampus*, 18, 1314–1324.
- Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, 2, 704–716.
- Fang, L., & Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45–82.
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin and Review*, 9, 59–79.
- Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, 58, 1–48.
- Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage*, 23, 840–848.
- Fiala, J. C., Grossberg, S., & Bullock, D. (1996). Metabotropic glutamate receptor activation in cerebellar Purkinje cells as substrate for adaptive timing of the classically conditioned eye blink response. *Journal of Neuroscience*, 16, 3760–3774.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local association field. *Vision Research*, 33, 173–193.
- Fischer, B. (1973). Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Research*, 13, 2113–2120.
- Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology*, 65, 77–117.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
- Francis, G., & Grossberg, S. (1996a). Cortical dynamics of boundary segmentation and reset: Persistence, afterimages, and residual traces. *Perception*, 35, 543–567.
- Francis, G., & Grossberg, S. (1996b). Cortical dynamics of form and motion integration: Persistence, apparent motion, and illusory contours. *Vision Research*, 36, 149–173.
- Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089–1104.
- Franklin, D. J., & Grossberg, S. (2008). Cognitive-emotional learning by neocortex, amygdala, and hippocampus: Timing, neurotrophins, amnesia, and consciousness. In *Proceedings of the twelfth international conference on cognitive and neural systems, ICCNS*. Boston University, May.
- Franklin, D. J., & Grossberg, S. (2016). A neural model of normal and abnormal learning and memory: Adaptively timed conditioning, hippocampus, amnesia, neurotrophins, and consciousness. *Cognitive, Affective, and Behavioral Neuroscience*, <http://dx.doi.org/10.3758/s13415-016-0463-y>, <http://rdcu.be/m8um>.
- Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32, 209–224.
- Gabrieli, J. D., McGlinchey-Berroth, R., Carroll, M. C., Gluck, M. A., Cermak, L. S., & Disterhoft, J. F. (1995). Intact delay-eyeblink classical conditioning in amnesia. *Behavioral Neuroscience*, 109, 819–827.
- Galantucci, B., Fowley, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, 13, 361–377.
- Gancarz, G., & Grossberg, G. (1999). A neural model of the saccadic eye movement control explains task-specific adaptation. *Vision Research*, 39, 3123–3143.
- Gao, E., & Suga, N. (1998). Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proceedings of the National Academy of Sciences*, 95, 12663–12670.
- Gaudiano, P., & Grossberg, S. (1991). Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories. *Neural Networks*, 4, 147–183.
- Gaudiano, P., & Grossberg, S. (1992). Adaptive vector integration to endpoint: Self-organizing neural circuits for control of planned movement trajectories. *Human Movement Science*, 11, 141–155.
- Georgopoulos, A. P., Kalaska, J. F., Caminiti, R., & Massey, J. T. (1982). On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *Journal of Neuroscience*, 2, 1527–1537.
- Georgopoulos, A. P., Schwartz, A. B., & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, 233, 1416–1419.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84, 279–325.
- Gibbon, J. (1991). The origins of scalar timing. *Learning and Motivation*, 22, 3–38.
- Glotin, H., Warnier, P., Dandurand, F., Dufau, S., Lete, B., Touzet, C., et al. (2010). An Adaptive Resonance Theory account of the implicit learning of orthographic word forms. *Journal of Physiology-Paris*, 104, 19–26.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305–320.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20–25.
- Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154–156.
- Gottlieb, J., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481–484.
- Graybiel, A. M. (1997). The basal ganglia and cognitive pattern generators. *Schizophrenia Bulletin*, 23, 459–469.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534–537.
- Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, 18, 88–94.
- Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, 324, 1207–1210.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of Acoustical Society of America*, 61, 1270–1277.
- Grindley, G. C., & Townsend, V. (1965). Binocular masking induced by a moving object. *Quarterly Journal of Experimental Psychology*, 17, 97–109.
- Grossberg, S. (1968a). A prediction theory for some nonlinear functional-differential equations, II: Learning of patterns. *Journal of Mathematical Analysis and Applications*, 22, 490–522.
- Grossberg, S. (1968b). Some nonlinear networks capable of learning a spatial pattern of arbitrary complexity. *Proceedings of the National Academy of Sciences*, 55, 368–372.
- Grossberg, S. (1968c). Some physiological and biochemical consequences of psychological postulates. *Proceedings of the National Academy of Sciences*, 60, 758–765.
- Grossberg, S. (1969). On the production and release of chemical transmitters and related topics in cellular control. *Journal of Theoretical Biology*, 22, 325–364.
- Grossberg, S. (1971). On the dynamics of operant conditioning. *Journal of Theoretical Biology*, 3(3), 225–255.
- Grossberg, S. (1972a). A neural theory of punishment and avoidance, I: Qualitative theory. *Mathematical Biosciences*, 15, 39–67.
- Grossberg, S. (1972b). 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. (1975). A neural model of attention, reinforcement, and discrimination learning. *International Review of Neurobiology*, 18, 263–327.
- Grossberg, S. (1976a). Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, 23, 121–134.
- Grossberg, S. (1976b). Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics*, 23, 187–202.
- Grossberg, S. (1978a). A theory of human memory: Self-organization and performance of sensory-motor codes, maps, and plans. In R. Rosen, & F. Snell (Eds.), *Progress in theoretical biology: Vol. 5* (pp. 233–374). New York: Academic Press.
- Grossberg, S. (1978b). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, 3, 199–219.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1–51.
- Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, 89, 529–572.
- Grossberg, S. (1984a). Outline of a theory of brightness, color, and form perception. In Degreef E., & J. van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 59–85). Amsterdam: North Holland.
- Grossberg, S. (1984b). Some psychophysiological and pharmacological correlates of a developmental, cognitive, and motivational theory. In R. Karrer, J. Cohen, & P. Tueting (Eds.), *Brain and information: Event related potentials* (pp. 58–142). New York: New York Academy of Sciences.
- Grossberg, S. (1986). The adaptive self-organization of serial order in behavior: Speech, language, and motor control. In E. C. Schwab, & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines, Vol. 1: Speech perception* (pp. 187–294). New York: Academic Press.
- Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception, II: Binocular theory. *Perception and Psychophysics*, 41, 117–158.
- 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 and Psychophysics*, 55, 48–120.
- Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional figures. *Psychological Review*, 104, 618–658.
- Grossberg, S. (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, 12, 163–186.
- Grossberg, S. (2000a). The complementary brain: Unifying brain dynamics and modularity. *Trends in Cognitive Sciences*, 4, 233–246.
- Grossberg, S. (2000b). The imbalanced Brain: From normal behavior to schizophrenia. *Biological Psychiatry*, 48, 81–98.
- Grossberg, S. (2003a). 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. (2003b). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423–445.
- Grossberg, S. (2007). Consciousness CLEARS the mind. *Neural Networks*, 20, 1040–1053.
- Grossberg, S. (2009). Cortical and subcortical predictive dynamics and learning during perception, cognition, emotion and action. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 364, 1223–1234.

- Grossberg, S. (2013a). Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, *37*, 1–47.
- Grossberg, S. (2013b). Recurrent neural networks. *Scholarpedia*, http://www.scholarpedia.org/article/Recurrent_neural_networks.
- Grossberg, S. (2014). How visual illusions illuminate complementary brain processes: illusory depth from brightness and apparent motion of illusory contours. *Frontiers in Human Neuroscience*, <http://dx.doi.org/10.3389/fnhum.2014.00854>.
- Grossberg, S. (2016a). Cortical dynamics of figure-ground separation in response to 2D pictures and 3D scenes: How V2 combines border ownership, stereoscopic cues, and Gestalt grouping rules. *Frontiers in Psychology: Perception Science*, <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.02054/full>.
- Grossberg, S. (2016b). Neural dynamics of the basal ganglia during perceptual, cognitive, and motor learning and gating. In J.-J. Soghomonian (Ed.), *The basal ganglia: Novel perspectives on motor and cognitive functions* (pp. 457–512). Berlin: Springer.
- Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 418–503.
- Grossberg, S., Bullock, D., & Dranias, M. (2008). Neural Dynamics Underlying Impaired Autonomic and Conditioned Responses Following Amygdala and Orbitofrontal Lesions. *Behavioral Neuroscience*, *122*, 1100–1125.
- Grossberg, S., Govindarajan, K. K., Wyse, L. L., & Cohen, M. A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. *Neural Networks*, *17*, 511–536.
- Grossberg, S., & Gutowski, W. E. (1987). Neural dynamics of decision making under risk: Affective balance and cognitive-emotional interactions. *Psychological Review*, *94*, 300–318.
- Grossberg, S., & Hong, S. (2006). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision*, *19*, 263–321.
- Grossberg, S., & Howe, P. D. L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research*, *43*, 801–829.
- Grossberg, S., & Huang, T.-R. (2009). ARTSCENE: A neural system for natural scene classification. *Journal of Vision*, *9*(6), 1–19.
- Grossberg, S., Hwang, S., & Mingolla, E. (2002). Thalamocortical dynamics of the McCollough effect: Boundary-surface alignment through perceptual learning. *Vision Research*, *42*, 1259–1286.
- Grossberg, S., & Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: A neural model of phonemic restoration using subsequent context in noise. *Journal of Acoustical Society of America*, *130*, 440–460.
- Grossberg, S., & Kazerounian, S. (2016). Phoneme restoration and empirical coverage of Interactive Activation and Adaptive Resonance models of human speech processing. *Journal of Acoustical Society of America*, *140*, 1130–1153.
- Grossberg, S., Kuhlmann, L., & Mingolla, E. (2007). A neural model of 3D shape-from-texture: Multiple-scale filtering, boundary grouping, and surface filling-in. *Vision Research*, *47*, 634–672.
- Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: Ballistic eye movements*. Amsterdam; New York: North-Holland.
- Grossberg, S., & Levine, D. S. (1987). Neural dynamics of attentionally modulated Pavlovian conditioning: Blocking, inter-stimulus interval, and secondary reinforcement. *Applied Optics*, *26*, 5015–5030.
- Grossberg, S., Markowitz, J., & Cao, Y. (2011). On the road to invariant recognition: Explaining tradeoff and morph properties of cells in inferotemporal cortex using multiple-scale task-sensitive attentive learning. *Neural Networks*, *24*, 1036–1049.
- Grossberg, S., & McLoughlin, N. (1997). Cortical dynamics of 3-D surface perception: Binocular and half-occluded scenic images. *Neural Networks*, *10*, 1583–1605.
- Grossberg, S., & Merrill, J. W. L. (1992). A neural network model of adaptively timed reinforcement learning and hippocampal dynamics. *Cognitive Brain Research*, *1*, 3–38.
- Grossberg, S., & Merrill, J. W. L. (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement. *Journal of Cognitive Neuroscience*, *8*, 257–277.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, *92*, 173–211.
- Grossberg, S., Mingolla, E., & Ross, (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, *101*, 470–489.
- Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*, 2521–2553.
- Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, *107*, 735–767.
- Grossberg, S., & Pearson, L. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review*, *115*, 677–732.
- Grossberg, S., & Pessoa, L. (1998). Texture segregation, surface representation, and figure-ground separation. *Vision Research*, *38*, 2657–2684.
- Grossberg, S., & Pilly, P. K. (2012). How entorhinal grid cells may learn multiple spatial scales from a dorsoventral gradient of cell response rates in a self-organizing map. *PLoS Computational Biology*, *8*(10), 31002648. <http://dx.doi.org/10.1371/journal.pcbi.1002648>.
- Grossberg, S., & Pilly, P. K. (2014). Coordinated learning of grid cell and place cell spatial and temporal properties: multiple scales, attention, and oscillations. *Philosophical Transactions of the Royal Society B*, *369*, 20120524.
- Grossberg, S., & Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, *40*, 1413–1432.
- Grossberg, S., & Repin, D. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, *16*, 1107–1140.
- Grossberg, S., & Schmajuk, N. A. (1987). Neural dynamics of attentionally-modulated Pavlovian conditioning: Conditioned reinforcement, inhibition, and opponent processing. *Psychobiology*, *15*, 195–240.
- Grossberg, S., & Schmajuk, N. A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*, *2*, 79–102.
- Grossberg, S., & Seidman, D. (2006). Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates. *Psychological Review*, *113*, 483–525.
- Grossberg, S., & Seitz, A. (2003). Laminar development of receptive fields, maps, and columns in visual cortex: The coordinating role of the subplate. *Cerebral Cortex*, *13*, 852–863.
- Grossberg, S., Srinivasan, K., & Yazdanbakhsh, A. (2014). Binocular fusion and invariant category learning due to predictive remapping during scanning of a depthful scene with eye movements. *Frontiers in Psychology: Perception Science*, <http://dx.doi.org/10.3389/fpsyg.2014.01457>.
- Grossberg, S., & Stone, G. O. (1986a). Neural dynamics of attention switching and temporal order information in short-term memory. *Memory and Cognition*, *14*, 451–468.
- Grossberg, S., & Stone, G. O. (1986b). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review*, *93*, 46–74.
- 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., & Todorovic, D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics*, *43*, 241–277.
- Grossberg, S., & Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research*, *1218*, 278–312.
- Grossberg, S., & Vladusich, T. (2010). How do children learn to follow gaze, share joint attention, imitate their teachers, and use tools during social interactions? *Neural Networks*, *23*, 940–965.
- Grossberg, G., & Williamson, J. R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings 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.
- Grossberg, S., Yazdanbakhsh, A., Cao, Y., & Swaminathan, G. (2008). How does binocular rivalry emerge from cortical mechanisms of 3-D vision? *Vision Research*, *48*, 2232–2250.
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*, 594–621.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*, 280–301.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*, 611–633.
- Gutschalk, A., & Dykstra, A. (2015). Auditory neglect and related disorders. *Handbook of Clinical Neurology*, *129*, 557–571.
- Harnad, S. (1990). The symbol grounding problem. *Physica B*, *42*, 335–346.
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337.
- He, Z. J., & Nakayama, K. (1992). Surface features in visual search. *Nature*, *359*, 231–233.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197.
- Heilman, K. M., Bowers, D., Coslett, H. B., Whelan, H., & Watson, R. T. (1985). Directional hypokinesia. *Neurology*, *35*, 855–859.
- Heitger, F., & von der Heydt, R. (1993). A computational model of neural contour processing: figure-ground segregation and illusory contours. In *Proceedings of the 4th international conference on computer vision* (pp. 32–40). Berlin, Germany.
- Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, *15*, 1–18.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Hong, S., & Grossberg, S. (2004). A neuromorphic model for achromatic and chromatic surface representation of natural images (with S. Hong). *Neural Networks*, *17*, 787–808.
- Houghton, G. (1990). The problem of serial order: a neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 287–319). London: Academic Press.
- Huang, T.-R., and, & Grossberg, S. (2010). Cortical dynamics of contextually cued attentive visual learning and search: Spatial and object evidence accumulation. *Psychological Review*, *117*, 1080–1112.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.

- Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General*, 118, 258–279.
- Husain, M., & Nachev, P. (2007). Space and the parietal cortex. *Trends in Cognitive Sciences*, 11, 30–36.
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, 43, 171–216.
- Ito, M., Westheimer, G., & Gilbert, C. D. (1998). Attention and perceptual learning modulate contextual influences on visual perception. *Neuron*, 20, 1191–1197.
- Janik, A. S., & Toulmin, S. E. (1996). *Wittgenstein's Vienna*. Chicago, Illinois: Ivan R. Dee Publisher.
- Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cueing: Configuration or individual locations? *Perception and Psychophysics*, 66, 454–463.
- Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). The functional equivalence of verbal and spatial memory in serial short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 21, 1008–1018.
- Kamin, L. J. (1968). Attention-like processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation*. Miami: University of Miami press.
- Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell, & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.
- Karnath, H.-O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal not the posterior parietal lobe. *Nature*, 411, 950–953.
- Karnath, H.-O., & Rorden, C. (2012). The anatomy of spatial neglect. *Neuropsychologia*, 50, 1010–1017.
- Kazerounian, S., & Grossberg, S. (2014). Real-time learning of predictive recognition categories that chunk sequences of items stored in working memory. *Frontiers in Psychology: Language Sciences*, <http://dx.doi.org/10.3389/fpsyg.2014.01053>.
- Keizer, A. W., Hommel, B., & Lamme, V. A. F. (2015). Consciousness is not necessary for visual feature binding. *Psychonomic Bulletin and Review*, 22, 453–460.
- Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology*, 23, 141–221.
- Kelly, F. J., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics*, 62, 1596–1619.
- Kentros, C. G., Agniotri, N. T., Streater, S., Hawkins, R. D., & Kandel, E. R. (2004). Increased attention to spatial context increases both place field stability and spatial memory. *Neuron*, 42, 283–295.
- Koch, C., Massimini, M., Boly, M., & Tononi, G. (2016). Neural correlates of consciousness: Progress and problems. *Nature Reviews Neuroscience*, 17, 307–321.
- Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*, 11, 16–22.
- Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., & Somers, D. C. (2012). Auditory spatial attention representations in the human cerebral cortex. *Cerebral Cortex*, 24, 773–784.
- Kooistra, C. A., & Heilman, K. M. (1989). Hemispatial visual inattention masquerading as hemianopia. *Neurology*, 20, 303–330.
- Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., et al. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and fMRI. *Science*, 284, 167–170.
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126, 159–179.
- Krupa, D. J., Ghazanfar, A. A., & Nicolelis, M. A. L. (1999). Immediate thalamic sensory plasticity depends on corticothalamic feedback. *Proceedings of the National Academy of Sciences*, 96, 8200–8205.
- Kunisaki, O., & Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, 85–91.
- Lafedoged, P., & Disner, S. F. (2012). *Vowels and consonants* (3rd ed.). Chichester, West Sussex, UK: Wiley-Blackwell.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, 10, 494–501.
- LeDoux, J. (1993). Emotional memory systems in the brain. *Behavioral Brain Research*, 58, 69–79.
- LeDoux, J. (1996). *The emotional brain*. New York: Simon & Shuster.
- LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73, 653–676.
- Leveille, J., Versace, M., & Grossberg, S. (2010). Running as fast as it can: How spiking dynamics form object groupings in the laminar circuits of visual cortex. *Journal of Computational Neuroscience*, 28, 323–346.
- Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48, 635–654.
- Levy, R., & Dubois, B. (2006). Apathy and the functional anatomy of the prefrontal cortex–basal ganglia circuits. *Cerebral Cortex*, 16, 916–928.
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321, 1502–1507.
- Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size invariant object represent in inferior temporal cortex. *Neuron*, 67, 1062–1075.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, 106, 623–642.
- Lleras, A., & von Mühlenen, A. (2004). Spatial context and top-down strategies in visual search. *Spatial Vision*, 17, 465–482.
- Llinás, R., Ribary, U., Contreras, D., & Pedroarena, C. (1998). The neuronal basis for consciousness. *Philosophical Transactions of the Royal Society of London B*, 353, 1841–1849.
- Logothetis, N. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society of London B*, 353, 1801–1818.
- Lu, Z.-L., & Dosher, B. A. (2004). Perceptual learning retunes the perceptual template in foveal orientation identification. *Journal of Vision*, 4, 5, <http://dx.doi.org/10.1167/4.1.5>.
- Luce, P. A., & McMennan, C. T. (2008). Spoken Word Recognition: The challenge of variation. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception*. Wiley Online Library, <http://dx.doi.org/10.1002/9780470757024.ch24>.
- Lucck, C. J., Zeki, S., Friston, K. J., Deiber, M.-P., Cope, P., Cunningham, V. J., et al. (1989). The colour centre in the cerebral cortex of man. *Nature*, 340, 386–389.
- Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron*, 90, 152–164.
- MacDonald, C. J., Lepage, K. Q., Eden, U. T., & Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontiguous events. *Neuron*, 71, 737–749.
- MacNeilage, P. F. (1998). The frame/content theory of evolution of speech. *Behavioral and Brain Sciences*, 21, 499–546.
- Magezi, D. A., & Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *Journal of Neurophysiology*, 104, 1997–2007.
- Manassi, M., Lonchamp, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, 16, 35. <http://dx.doi.org/10.1167/16.3.35>.
- Mann, V., & Repp, B. (1980). The influence of vocalic context on perception of the /sh/-s/ context. *Perception and Psychophysics*, 28, 213–228.
- Marshall, J. C. (2001). Auditory neglect and right parietal cortex. *Brain*, 124, 645–646.
- Matthot, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, 21, 1793–1798.
- Mattingley, J. B., Bradshaw, J. L., & Bradshaw, J. A. (1995). The effects of unilateral visuospatial neglect on perception of Muller-Lyer illusory figures. *Perception*, 24, 415–433.
- Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science*, 275, 671–674.
- Mattingley, J. B., Husain, M., Rorden, C., Kennard, C., & Driver, J. (1998). Motor role of human inferior parietal lobe revealed in unilateral neglect patients. *Nature*, 392, 179–182.
- McAdams, S. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed.) (pp. 35–68). Amsterdam: Elsevier.
- McGlinchey-Berroth, R., Brawn, C., & Disterhoft, J. F. (1999). Temporal discrimination learning in severe amnesia patients reveals an alteration in the timing of eyeblink conditioned responses. *Behavioral Neuroscience*, 113, 10–18.
- McGlinchey-Berroth, R., Carrillo, M. C., Gabrieli, J. D., Brawn, C. M., & Disterhoft, J. F. (1997). Impaired trace eyeblink conditioning in bilateral, medial-temporal lobe amnesia. *Behavioral Neuroscience*, 111, 873–882.
- McGlinchey-Berroth, R., Milberg, W. P., Verfaellie, M., Alexander, M., & Kilduff, P. (1993). Semantic priming in the neglected field: Evidence from a lexical decision task. *Cognitive Neuropsychology*, 10, 79–108.
- McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 306–321.
- McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2003). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539–553.
- Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, 10, 903–907.
- Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *Journal of Vision*, 8, 1–17.
- Melcher, D. (2009). Selective attention and the active remapping of object features in trans-saccadic perception. *Vision Research*, 49, 1249–1255.
- Mendelson, J., Schreiner, C., Sutter, M., & Grasse, K. (1993). Functional topography of cat primary auditory cortex: responses to frequency-modulated sweeps. *Experimental Brain Research*, 94, 65–87.
- Mesulam, M.-M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical Transactions of the Royal Society B*, 354, 1325–1346.
- Millenson, J. R., Kehoe, E. J., & Gormezano, I. (1977). Classical conditioning of the rabbit's nictitating membrane response under fixed and mixed CS-US intervals. *Learning and Motivation*, 8, 351–366.
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics*, 25, 457–465.
- Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions Royal Society of London B*, 298, 85–95.
- Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6, 414–417.
- Mitroff, S. R., & Scholl, B. J. (2005). Forming and updating object representations without awareness: Evidence from motion-induced blindness. *Vision Research*, 45, 961–967.
- Møller, A. R. (1983). *Auditory physiology*. New York: Academic Press.
- Mormann, F., & Koch, C. (2007). Neural correlates of consciousness. *Scholarpedia*, 2(12), 1740.
- Morris, R. G. M., & Frey, U. (1997). Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, 350, 1469–1503.

- Mort, D. J., Malhotra, P., Mannan, S. K., Rorden, C., Pambakian, A., Kennard, C., & Husain, M. (2003). The anatomy of visual neglect. *Brain*, *126*, 1966–1997.
- Muramoto, K., Ono, T., Nishijo, H., & Fukuda, M. (1993). Rat amygdaloid neuron responses during auditory discrimination. *Neuroscience*, *52*, 621–636.
- Näätänen, R. (1982). Processing negativity: An evoked-potential reflection. *Psychological Bulletin*, *92*, 605–640.
- Näätänen, R., Simpson, M., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychiatry*, *14*, 53–98.
- Nachev, P., & Husain, M. (2006). Disorders of visual attention and the posterior parietal cortex. *Cortex*, *42*, 766–773.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *4*, 435–450.
- Nakayama, K., & Shimojo, S. (1990). Da Vinci stereopsis: Depth and subjective occluding contours from unpaired image points. *Vision Research*, *30*, 1811–1825.
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*, 264–265.
- Neville, H. J., Coffey, S. A., Lawson, D. S., Fischer, A., Emmorey, K., & Bellugi, U. (2002). Neural systems mediating American sign language: Effects of sensory experience and age of acquisition. *Brain and Language*, *57*, 285–308.
- Nishijo, H., Ono, T., & Nishino, H. (1988). Topographic distribution of modality-specific amygdalar neurons in alert monkey. *Journal of Neuroscience*, *8*, 3556–3569.
- Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*, 1281–1296.
- Ogawa, T., & Komatsu, H. (2004). Target selection in area V4 during a multidimensional visual search task. *The Journal of Neuroscience*, *24*, 6371–6382.
- O'Herron, P., & von der Heydt, R. (2009). Short-term memory for figure-ground organization in the visual cortex. *Neuron*, *61*, 801–809.
- Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*, 273–302.
- Otto, T., & Eichenbaum, H. (1992). Neuronal activity in the hippocampus during delayed non-match to sample performance in rats: Evidence for hippocampal processing in recognition memory. *Hippocampus*, *2*, 323–334.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761–781.
- Palma, J., Grossberg, S., & Versace, M. (2012). Persistence and storage of activity patterns in spiking recurrent cortical networks: Modulation of sigmoid signals by after-hyperpolarization currents and acetylcholine. *Frontiers in Computational Neuroscience*, *6*, 42. <http://dx.doi.org/10.3389/fncom.2012.00042>.
- Palma, J., Versace, M., & Grossberg, S. (2012). After-hyperpolarization currents and acetylcholine control sigmoid transfer functions in a spiking cortical model. *Journal of Computational Neuroscience*, *32*, 253–280.
- Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 309–328.
- Paradiso, M. A., & Nakayama, K. (1991). Brightness perception and filling-in. *Vision Research*, *31*, 1221–1236.
- Parker, J. L., & Dostrovsky, J. O. (1999). Cortical involvement in the induction, but not expression, of thalamic plasticity. *The Journal of Neuroscience*, *19*, 8623–8629.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.
- Pasupathy, A., & Miller, E. K. (2004). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*, 873–876.
- Pavlov, I. P. (1927). *Conditioned reflexes*. Oxford University Press.
- Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of the mind: Knowledge, belief, and communication. *Child Development*, *60*, 688–700.
- Pessoa, L. (2008). On the relation between emotion and cognition. *Nature*, *9*, 148–158.
- Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*, 160–166.
- Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: from a "low road" to "many roads" of evaluating biological significance. *Nature Reviews Neuroscience*, *11*, 773–782.
- Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *The Journal of Neuroscience*, *9*, 1749–1763.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of Acoustical Society of America*, *24*, 175–184.
- Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*, 283–290.
- Piaget, J. (1945). *La Formation du Symbole Chez L'enfant*. Paris: Delachaux Niestle, S.A.
- Piaget, J. (1951). *Play, dreams and imitation in childhood* (C. Gattegno and C. F. M. Hodgson, Trans.). London: Routledge and Kegan Paul.
- Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.
- Pickles, J. O. (1988). *An introduction to the physiological of hearing* (2nd ed.). New York: Academic Press.
- Pilly, P. K., & Grossberg, S. (2012). How do spatial learning and memory occur in the brain? Coordinated learning of entorhinal grid cells and hippocampal place cells. *Journal of Cognitive Neuroscience*, *24*, 1031–1054.
- Pinna, B., & Grossberg, S. (2005). The watercolor illusion and neon color spreading: A unified analysis of new cases and neural mechanisms. *Journal of the Optical Society of America A*, *22*, 2207–2221.
- Poliemi, J. R., Balasubramanian, M., & Schwartz, E. L. (2006). Multi-area visuotopic map complexes in macaque striate and extra-striate cortex. *Vision Research*, *46*, 3336–3359.
- Pollen, D. A. (1999). On the neural correlates of visual perception. *Cerebral Cortex*, *9*, 4–19.
- Ponce, C. R., Lomber, S. G., & Born, R. T. (2008). Integrating motion and depth via parallel pathways. *Nature Neuroscience*, *11*, 216–223.
- Port, R. F., & Dalby, J. (1982). Consonant-vowel ratio as a cue for voicing in English. *Perception and Psychophysics*, *32*, 141–152.
- Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.
- Posner, M. I., Walker, J. A., Friedrich, R. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, *4*, 1863–1874.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381–410.
- Prescott, T. J., Gurney, K., & Redgrave, P. (2003). Basal ganglia. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed.). Cambridge, MA: MIT Press.
- Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). *Nature Neuroscience*, *10*, 1492–1499.
- Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues and Gestalt rules. *Neuron*, *47*, 155–166.
- Raizada, R., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, *13*, 100–113.
- Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling-in of artificially induced scotomas in human vision. *Nature*, *350*, 699–702.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*, 516–521.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*, 718–724.
- Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences*, *97*, 11800–11806.
- Rayleigh, L. (1907). On our perception of sound direction. *Philosophical Magazine*, *13*, 2–4–232.
- Remez, R. E. (2003). Establishing and maintaining perceptual coherence: Unimodal and multimodal evidence. *Journal of Phonetics*, *31*, 293–304.
- Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, *12*, 24–29.
- Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review*, *101*, 129–156.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- Repp, B. H. (1980). A range-frequency effect on perception of silence in speech. In *Status report on speech research SR-61* (pp. 151–165). New Haven, CT: Haskins Laboratories.
- Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*, 853–863.
- Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*, 168–185.
- Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*, 703–714.
- Rhode, W. S., & Smith, P. H. (1986a). Encoding timing and intensity in the ventral cochlear nucleus of the cat. *Journal of Neurophysiology*, *56*, 261–287.
- Rhode, W. S., & Smith, P. H. (1986b). Physiological studies on neurons in the dorsal cochlear nucleus of the cat. *Journal of Neurophysiology*, *56*, 287–307.
- Robertson, I. H., Manly, T., Beschin, N., Daini, R., Haeske-Dewick, H., Homberg, V., et al. (1997). Auditory sustained attention is a marker of unilateral spatial neglect. *Neuropsychologia*, *35*, 1527–1532.
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*, 376–381.
- Roggeveen, A., Pilz, K., Bennett, P., & Sekuler, A. (2009). Individual differences in object based attention. *Journal of Vision*, *9*, 143.
- Rolls, E. T. (1999). *The brain and emotion*. Oxford: Oxford University Press.
- Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *10*, 284–294.
- Rueckert, L., & Grafman, J. (1998). Sustained attention deficits in patients with lesions of parietal cortex. *Neuropsychologia*, *36*, 653–660.
- Sachs, M. B., & Young, E. D. (1979). Encoding of steady state vowels in the auditory nerve: Representations in terms of discharge rate. *Journal of the Acoustical Society of America*, *66*, 470–479.
- Sammiller, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., et al. (2010). Processing of unfamiliar songs: A functional magnetic resonance adaptation study. *The Journal of Neuroscience*, *30*, 3572–2478.
- Sams, M., Paavilainen, P., Alho, K., & An, R. (1985). Auditory frequency discrimination and event-related potentials. *Clinical Neurophysiology*, *62*, 437–448.
- Samuel, A. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 474–494.
- Samuel, A. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1124–1131.
- Saygin, A. P., & Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cerebral Cortex*, *18*, 2158–2168.
- Schoenbaum, G., Setlow, B., Saddoris, M. P., & Gallagher, M. (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron*, *39*, 855–867.
- Schvaneveldt, R. W., & McDonald, J. E. (1981). Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 243–256.

- Schwab, E. C., Sawusch, J. R., & Nusbaum, H. C. (1981). The role of second formant transitions in the stop-semivowel distinction. *Perception and Psychophysics*, 21, 121–128.
- Schwartz, E. L. (1984). Spatial mapping and spatial vision in primate striate and inferotemporal cortex. In L. Spillmann, & B. R. Woolen (Eds.), *Sensory experience, adaptation and perception* (pp. 73–104). Hillsdale, NJ: Erlbaum.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesion. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.
- Seitz, A. R., & Watanabe, T. (2003). Is subliminal learning really passive? *Nature*, 422, 6927.
- Schiller, P. H., & Lee, K. (1991). The role of the primate extrastriate area V4 in vision. *Science*, 251, 1251–1253.
- Shannon, C. E. (1948a). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.
- Shannon, C. E. (1948b). A mathematical theory of communication. *Bell System Technical Journal*, 27, 623–656.
- Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science*, 348, 1352–1355.
- Silver, M. R., Grossberg, S., Bullock, D., Histed, M. H., & Miller, E. K. (2011). A neural model of sequential movement planning and control of eye movements: Item-order-rank working memory and saccade selection by the supplementary eye fields. *Neural Networks*, 26, 29–58.
- Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in Cognitive Sciences*, 13, 488–495.
- Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, 9, 16–20.
- Singer, W. (1998). Consciousness and the structure of neuronal representations. *Philosophical Transactions of the Royal Society B*, 353, 1829–1840.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Oxford, UK: Appleton-Century.
- Sleve, L. R., & Okada, B. M. (2015). Processing structure in language and music: a case for shared reliance on cognitive control. *Psychological Bulletin and Review*, 22, 637–652.
- Smith, M. C. (1968). CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology*, 3, 679–687.
- Solakov, E. N. (1968). *Mechanisms of memory*. Moscow University Press.
- Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial vision processing in frontal cortex. *Nature*, 444, 374–377.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, 240, 338–340.
- Squire, L. R., & Cohen, N. J. (1984). Human memory and amnesia. In G. Lynch, J. McGaugh, & N. M. Weinberger (Eds.), *Neurobiology of learning and memory* (pp. 3–64). New York: Guilford.
- Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American*, 236, 82–91.
- Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *Journal of Neurophysiology*, 27, 5326–5337.
- Takeichi, H., Shimojo, S., & Watanabe, T. (1992). Neon flank and illusory contour: Interaction between the two processes leads to color filling-in. *Perception*, 21, 313–324.
- Tamietti, M., & de Gelder, B. (2010). Neural bases of the non-conscious perception of emotional signals. *Nature Reviews Neuroscience*, 11, 697–709.
- Tanca, M., Grossberg, S., & Pinna, B. (2010). Probing perceptual antinomies with the watercolor illusion and explaining how the brain sees them. *Seeing and Perceiving*, 23, 295–333.
- Taniwaki, T., Okayama, A., Yoshiura, T., Togao, O., Nakamura, Y., Yamasaki, T., et al. (2006). Functional network of the basal ganglia and cerebellar motor loops *in vivo*: Different activation patterns between self-initiated and externally triggered movements. *NeuroImage*, 31, 745–753.
- Theeuwes, J., Mathot, S., & Kingstone, A. (2010). Object-based eye movements: the eyes prefer to stay within the same object. *Attention, Perception, and Psychophysics*, 72, 597–601.
- Thorell, L. G., De Valois, R. L., & Albrecht, D. G. (1984). Spatial mapping of monkey V1 cells with pure color and luminance stimuli. *Vision Research*, 24, 751–769.
- Tian, B., & Rauschecker, J. (1984). Processing of frequency-modulated sounds in the cat's anterior auditory field. *Journal of Neurophysiology*, 71, 1959–1975.
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32, 1349–1357.
- Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, 29, 757–767.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*. <http://dx.doi.org/10.1186/1471-2202-5-42>.
- Tononi, G. (2012). Integrated information theory of consciousness: an updated account. *Archives Italiennes de Biologie*, 150, 290–326.
- Tononi, G. (2015). Integrated information theory. *Scholarpedia*, 10(1), 4164.
- Tootell, R. B. H., Silverman, M. S., Switkes, E., & DeValois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218, 902–904.
- Toyomitsu, Y., Nishijo, H., Uwano, T., Kuratsu, J., & Ono, T. (2002). Neuronal responses of the rat amygdala during extinction and reassociation learning in elementary and configural associative tasks. *European Journal of Neuroscience*, 15, 753–768.
- 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.
- Tsuchiya, N., Wilke, M., Frassle, S., & Lamme, V. A. F. (2016). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, 19, 757–770.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
- Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. *Perception*, 24, 127–153.
- Umeno, M. M., & Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye fields, I: predictive visual responses. *Journal of Neurophysiology*, 78, 1373–1383.
- Underwood, B. J., & Freund, J. S. (1970). Word frequency and short term recognition memory. *American Journal of Psychology*, 83, 343–351.
- van Tuijl, H. F. J. M. (1975). A new visual illusion: Neonlike color spreading and complementary color induction between subjective contours. *Acta Psychologica*, 39, 441–445.
- Varin, D. (1971). Fenomini di contrasto e diffusione chromatica nell'organizzazione spaziale del campo percettivo. *Revista di Psychologica*, 65, 101–128.
- Vinogradova, O. S. (1975). Functional organization of the limbic system in the process of registration of information: Facts and hypotheses. In R. L. Isaacson, & K. H. Pribram (Eds.), *The hippocampus*: Vol. 2 (pp. 3–69). Plenum Press.
- Vitevitch, M. S., & Donoso, A. (2011). Processing of indexical information requires time: Evidence from Change deafness. *The Quarterly Journal of Experimental Psychology*, 64, 1484–1493.
- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408.
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260–1262.
- von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, 40, 1955–1967.
- Wallis, T. S., & Arnold, D. H. (2009). Motion-induced blindness and motion-streak suppression. *Current Biology*, 19, 325–329, 1484–1493.
- Wanning, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience*, 14, 1243–1244.
- Warren, R., & Sherman, A. (1974). Phonemic restorations based on subsequent context. *Perception & Psychophysics*, 16, 150–156.
- Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature*, 413, 844–848.
- Williams, L. R., & Jacobs, D. W. (1997). Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9, 837–858.
- Wolfe, J. M. (1994). Guided search 2. A revised model of visual search. *Psychonomic Bulletin and 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.
- Yan, J., & Scott, T. R. (1996). The effect of satiety on responses of gustatory neurons in the amygdala of alert cynomolgus macaques. *Brain Research*, 740, 193–200.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 121–134.
- Young, E. D., & Sachs, M. B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *Journal of Acoustical Society of America*, 66, 1381–1403.
- Zeki, S. (1983). Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, 9, 741–765.
- Zhang, N. R., & von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *The Journal of Neuroscience*, 30, 6482–6496.
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, 20, 6594–6611.
- Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. *The Journal of Neuroscience*, 27, 12292–12307.