

REGULAR ARTICLE

## Grandmother cohorts: multiple-scale brain compression dynamics during learning of object and sequence categories

Stephen Grossberg<sup>a,b,c</sup>

<sup>a</sup>Center for Adaptive Systems, Boston University, Boston, MA, USA; <sup>b</sup>Graduate Program in Cognitive and Neural Systems, Boston University, Boston, MA, USA; <sup>c</sup>Departments of Mathematics, Psychology, and Biomedical Engineering, Boston University, Boston, MA, USA

### ABSTRACT

This article summarises neural models of how invariant object categories are learned under conditions where eye movements freely scan a scene, and of how list categories, or chunks, are learned in response to sequences, or lists, of events that are temporarily stored in working memory. During both kinds of category learning, distributed information is compressed by a category learning process into compact groups of cells that can selectively respond to this information. Adaptive Resonance Theory models describe these learning processes and enable, in a limiting case, winner-take-all grandmother cells to be learned, even in a single trial, and rapidly activated during recall trials by bottom-up direct access. The degree of compression can vary with the amount of contextual evidence, and a distributed network of multiple categories may be needed to represent an object or event, all computed using localist mechanisms. Such a network is called a *grandmother cohort* in the current article.

### ARTICLE HISTORY

Received 14 May 2016  
Accepted 1 July 2016

### KEYWORDS

Grandmother cell;  
grandmother cohort;  
category learning; working  
memory; Adaptive  
Resonance

### 1. From grandmother cells to multiple-scale brain compression dynamics

Wikipedia defines a grandmother cell as follows: “The grandmother cell is a hypothetical neuron that represents a complex but specific concept or object. It activates when a person sees, hears, or otherwise sensibly discriminates a specific entity, such as his or her grandmother”. A grandmother cell is a specific kind of recognition category. This article summarises two kinds of brain processes wherein category learning, and thus information compression, occur. The first kind emphasises spatial processing, since it involves the learning of invariant object categories. The second kind emphasises temporal processing, since it involves the learning of sequence categories, or list chunks, of sequences of events that are temporarily stored in a short-term working memory. An invariant object category can be activated when a given object is seen on the retina at multiple views, positions, and sizes. A list chunk can be activated when a given sequence, whether linguistic, spatial, or motor, is stored in working memory. Such a list chunk can, for example, in the case of linguistic working memory represent a phoneme, syllable, or word; in the case of spatial working memory represent a sequence of target positions for eye, arm, or navigational movements; and in the case of motor working memory represent a motor skill such as a sequence of reaching or dance moves. All three types of working memory and

list chunking networks are predicted to be realised by similar network designs; see Section 6.6.

In all these cases, even though the learned categories compress distributed information of various kinds, thereby illustrating “localist” representations, the information that they are compressing, and the control dynamics that regulate the learning processes, often involve information that is distributed across multiple scales across an entire neural system. Because these learning processes use information that is distributed across multiple spatial scales, a compressive event, even when the limiting case of a “grandmother” cell may exist, is only a small part of the dynamics whereby learning, recognition, and prediction occur. It should, however, be emphasised that, even across all of these scales, “local” mechanisms are at work, meaning mechanisms that use only information that is locally available at the location in the network where the computation takes place. Despite the physical necessity of such a constraint, it is not realised in many probabilistic, notably Bayesian, theories wherein non-local computations are explicitly or implicitly assumed, and thus are not physically plausible.

### 2. Adaptive Resonance Theory: the CLEARS processes

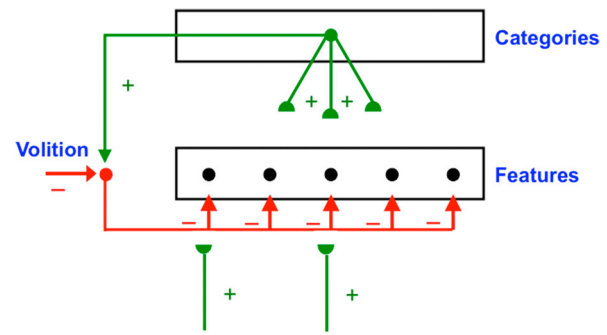
These localist models are part of Adaptive Resonance Theory, or ART, which is celebrating its 40th anniversary

this year (Grossberg, 1976a, 1976b, 1978a, 1980). ART clarifies how humans are able to rapidly learn enormous amounts of new information, on their own, throughout life, without being forced to just as rapidly forget what they already know. When such unselective forgetting does occur, it is often called *catastrophic forgetting*, or alternatively, *catastrophic interference*.

Grossberg (1980) has called the problem whereby the brain learns quickly and stably without catastrophically forgetting its past knowledge the *stability–plasticity dilemma*. ART proposes how the stability–plasticity dilemma may be solved, and hereby clarifies how humans can stably accumulate memories over a lifetime that cohere into an emerging self. ART has been incrementally developed into a cognitive and neural theory of how the brain autonomously learns to attend, recognise, and predict objects and events in a changing world, without experiencing catastrophic forgetting. At present, ART currently has the broadest explanatory and predictive range of available cognitive and neural theories. Grossberg (2013a, 2016b) provides heuristic reviews of ART design principles, neural mechanisms, and interdisciplinary data explanations. Here only the minimal amount of background will be provided to make our main points.

To solve the stability–plasticity dilemma, ART specifies mechanistic links between brain processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony, and proposes how these processes are related to one another in specific ways that psychological and neurobiological data have supported. These CLEARs mechanisms clarify how many animals become 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) and *one-to-many maps* (representations that enable us to expertly know many things about individual objects and events).

ART accomplishes these properties by proposing how learned recognition categories read-out learned expectations that focus attention via top-down, modulatory on-center, off-surround networks of cells that obey the membrane equations of neurophysiology, also called shunting laws (Figure 1). Such a top-down expectation supports a matching process, called the ART Matching Rule, which can focus attention on learned combinations of cues, called *critical feature patterns*. Critical feature patterns are the features that enable categories and expectations to successfully predict the objects and events that they represent. This on-center off-surround form of the ART Matching Rule realises a kind of self-normalising “biased competition” that has helped to explain several different kinds of data (e.g. Bhatt, Carpenter, &



**Figure 1.** 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 sensitise cells in its excitatory on-center (green pathways with hemi-circular 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, while other featural cells are inhibited (“two-against-one”).

Grossberg, 2007; Carpenter & Grossberg, 1987, 1991; Desimone, 1998; Grossberg, 2013a; Reynolds & Heeger, 2009).

ART explains how such top-down attentive matching may help to solve the stability–plasticity dilemma as follows: 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 bottom-up recognition categories and top-down expectations – hence the name *adaptive resonance* – while suppressing outliers that could have caused catastrophic forgetting.

### 3. Three types of attention: prototype, surface, and boundary attention

The above type of object attention is also called *prototype attention* because a critical feature pattern represents the prototype that is learned from the multiple sets of features that can activate that object category. ART also mechanistically describes two other kinds of attention: *boundary attention* and *surface attention*. When spatial attention focuses on an object boundary, it can flow along that boundary, even if it is defined by an illusory contour, as reported in neurophysiological experiments by Roelfsema, Lamme, and Spekreijse (1998) and Wannig, Stanisor, and Roelfsema (2011), and simulated using the LAMINART laminar cortical model of perceptual grouping and attention (Grossberg & Raizada, 2000). In addition, when

spatial attention focuses on part of an object surface, surface attention can flow from one part of the surface to another via a process of surface filling-in (Cao & Grossberg, 2005; Fazl, Grossberg, & Mingolla, 2009; Grossberg & Todorovic, 1988). This kind of surface attention, which is computed within the posterior parietal cortex (PPC) of the dorsal Where/How cortical processing stream (Goodale & Milner, 1992), will be shown in Section 5.2 to play an important role in modulating the learning of invariant object categories within the inferotemporal (IT) cortex of the ventral What cortical processing stream (Mishkin, 1982; Mishkin, Ungerleider, & Macko, 1983).

Such an invariant object category embodies properties of a grandmother cell in the sense that it is selectively activated by a specific object in the world. Many possible views, positions, or sizes of this object on the retina may activate the same invariant object category; hence, its “invariance”. A number of classical concepts defined grandmother cells in terms of cells that exhibit various invariant properties (e.g. Barlow, 1972; Gross, 2002; Konorski, 1967). An invariant object category is not as specific as cells that may respond, say, to just one of several possible views of the object. As noted in Section 5, cells at an earlier stage of cortical processing may respond more selectively to a particular object view at a particular retinal position. Such view-specific cells have also been found in the IT cortex (e.g. Desimone, Albright, Gross, & Bruce, 1984; Perrett et al., 1984, 1992), whereas cells with more invariant response properties have been found in other areas of IT cortex (e.g. Booth & Rolls, 1998; Logothetis, Pauls, & Poggio, 1995). Thus, in considering the concept of “grandmother cell”, it is important to note, in addition to the selectivity of cell response, whether that selectivity is elicited by a specific object with a particular view or position in space relative to the observer, or to that object even if it is viewed under less restrictive conditions.

## 4. Linking prototype and surface attention, category learning, and recognition

### 4.1. Learning an invariant object category as the eyes freely scan a scene

The modulatory role of spatial attention on invariant object category learning was discovered during the development of the family of ARTSCAN models, whose culmination, at least to the present, is the 3D ARTSCAN Search model (Chang, Grossberg, & Cao, 2014; Grossberg, Srinivasan, & Yazdanbakhsh, 2014); see Figure 2. 3D ARTSCAN Search proposes solutions to basic problems that confront a human or other primate as it learns to recognise objects in a 3D scene while freely

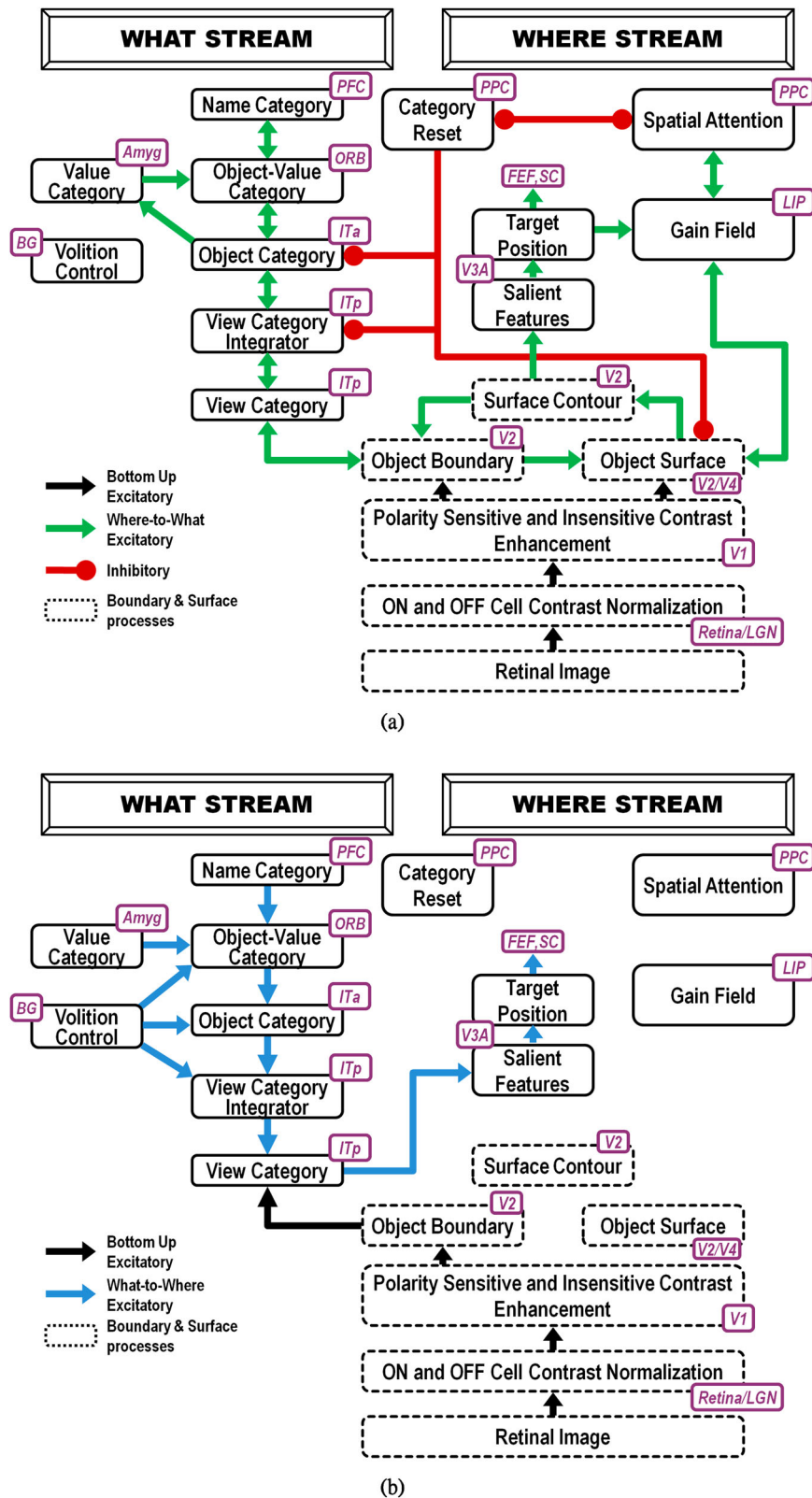
scanning the scene with eye movements. One such problem is that, as our eyes scan a scene, two successive eye movements may focus on different parts of the same object or on different objects. How does the brain avoid learning to erroneously classify views of different objects together, without an external 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.

In fact, such scanning eye movements may be used to learn the object concept in the first place. For this to be possible, successive eye movements need to foveate different views of the same object, and not just randomly jump from location to location in the scene. Scanning eye movements do seem to prefer looking at different parts of the same object before jumping to focus on a different object (Theeuwes, Mathot, & Kingstone, 2010), and the current theory proposes how this happens as part of the way in which surface attention works. Many other paradoxical data may also be explained by these concepts, including how spatial attention can increase the perceived brightness of a surface (Carrasco, Penpeci-Talgar, & Eckstein, 2000; Reynolds & Desimone, 2003), how predictive remapping of eye position occurs (Duhamel, Colby, & Goldberg, 1992; Gottlieb, Kusunoki, & Goldberg, 1998; Melcher, 2007), and what sort of category invariance can be learned (Grossberg, Markowitz, & Cao, 2011; Zoccolan, Kouh, Poggio, & DiCarlo, 2007), all of them providing accumulating evidence in support of model hypotheses.

The ARTSCAN models show how these properties may be accomplished by introducing the concept of a *surface-shroud resonance* (Fazl et al., 2009; Grossberg, 2009), which the model proposes to be the mechanism whereby the brain maintains sustained spatial attention upon the surface of an object that is being learned. A surface-shroud resonance explains how the brain can learn to associate multiple views of the *same* object, and only that object, with an emerging invariant object category, without an external teacher, as the eyes scan a scene. Such a surface-shroud resonance is an example of a distributed representation, albeit one that is realised by local neuronal mechanisms.

### 4.2. Surface-shroud resonances triggered between V4 and PPC

Several perceptual and cognitive scientists have reported that spatial attention can fit itself to the shape of an attended object. The term *attentional*



**Figure 2.** ARTSCAN Search model macrocircuit. A surface-shroud resonance between an object surface representation within cortical areas V2/V4 and a form-fitting attentional shroud in a spatial attention region of PPC is predicted to initiate a conscious visual percept of surface qualia. This circuit (a) can learn invariant object categories in cortical area ITa using Where-to-What modulation of category learning from the active shroud in PPC, and (b) can search for a valued goal object using What-to-Where interactions. See text for details. Reprinted with permission from Chang et al. (2014).



*shroud* for such a form-fitting distribution of spatial attention was introduced by Tyler and Kontsevich (1995). Grossberg (2009) predicted that, in addition to possible roles for spatial attention in visual perception, an active shroud in cortical area PPC that focuses sustained spatial attention upon an object surface in cortical area V4 also regulates the learning of invariant recognition categories in anterior inferotemporal (ITa) cortex, among other cortical areas.

An attentional shroud can form when bottom-up topographic excitatory signals from an active surface representation in V4 compete in PPC for spatial attention, while activated cells in PPC also send top-down topographic excitatory signals back to V4. Taken together, these signals form a *recurrent* on-center off-surround network that is capable of contrast-enhancing and normalising 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 sustains spatial attention upon the selected object surface. Once triggered, a surface-shroud resonance can propagate top-down to earlier cortical areas such as V1 and V2, and bottom-up to higher cortical areas such as prefrontal cortex, or PFC. Another important property of a surface-shroud resonance is that it is predicted to support conscious visual percepts (Grossberg, 2016b). Thus, a surface-shroud resonance supports paying spatial attention to consciously visible surface qualia.

#### 4.3. Why do not all occluders look transparent? Recognition vs. seeing

Why is such a surface-shroud resonance assumed to be triggered between V4 and PPC? Many theoretical and experimental results are consistent with this hypothesis. In particular, the FACADE and 3D LAMINART models have explained and predicted interdisciplinary data about many visual percepts using mechanisms that realise the assumption that cortical areas V2 and V4 resolve a basic design tension between recognition and seeing that prevents all occluding objects from looking transparent (Grossberg, 1994, 1997).

In particular, V4 is predicted to be the cortical region where figure-ground-separated 3D surface representations of *unoccluded* opaque object regions are completed and seen, and where percepts of 3D *transparent* surfaces are seen, whereas V2 is the cortical region where the completed object boundaries and surfaces of both occluding and *occluded* object regions may be amodally recognised (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). Many neurophysiological data about V2 and V4 are also consistent with this hypothesis. See Grossberg (2016a, 2016b) for reviews. Thus, a surface-shroud resonance is triggered between V4 and PPC because V4 is predicted to be the cortical stage at which visible figure-ground-separated 3D surface representations are computed.

Before reviewing how a surface-shroud resonance is proposed to modulate the learning of invariant object categories, more information is needed about how ART can learn recognition categories. From the perspective of ARTSCAN, these ART recognition categories are view- and position-selective categories that are learned in posterior inferotemporal (ITp) cortex. The additional machinery of ARTSCAN is needed to show how learning can link several such ART categories in ITp to an invariant object recognition category in ITa, and conversely, while eye movements freely scan a scene. Although these object categories in ITp and ITa individually compress important information about the object, a distributed network of categories across IT, as well as other brain regions, is needed to fully represent the object. Such a network is called a *grandmother cohort*.

#### 4.4. Feature-category resonances support category learning and recognition

The original development of the ART model proposed how such specific recognition categories are learned. This learning and recognition process is supported by *feature-category resonances*; that is, by resonances that coherently bind distributed patterns of critical feature patterns into recognition categories. Evidence in support of this kind of category learning and recognition has accumulated from modelling studies of such varied processes as texture segregation (Bhatt et al., 2007); visual scene recognition (Grossberg & Huang, 2009); contextual search of visual scenes (Huang & Grossberg, 2010); learning of view-, position-, and size-invariant IT recognition categories, and recoding of IT categories during target swapping (Cao, Grossberg, & Markowitz, 2011); gamma and beta oscillations during top-down matches and mismatches (Grossberg, 2013a; Grossberg & Versace, 2008); learning of concrete or abstract recognition categories under variable vigilance (Amis, Carpenter, Ersoy, & Grossberg, 2009; Grossberg et al., 2011); fast visual attention switching during the recognition of letter or number targets during rapid serial visual presentations (Grossberg & Stone, 1986a); 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 laws for top-down prototype attention obeying the ART Matching Rule derived from studies of this type (e.g. Bhatt et al., 2007), are also supported by studies by other authors, notably Reynolds and Heeger (2009) in their “normalization model of attention”.

How such ART recognition categories are learned is proposed to occur as follows. When a bottom-up input pattern is received at feature processing level, it can activate its target cells, if nothing else is happening. The activated cells send bottom-up signals through an adaptive filter to a category processing level. The category cells compete via a contrast-enhancing and normalising recurrent on-center off-surround network. The chosen category can, in turn, send a top-down expectation back to the feature level. When the top-down expectation pattern is received at the feature level, it provides excitatory modulatory, or priming, signals to feature cells in its on-center, and driving inhibitory signals to feature cells in its off-surround via the ART Matching Rule. The on-center is modulatory because the off-surround also inhibits the on-center cells (Figure 1), and these excitatory and inhibitory inputs are approximately balanced (“one-against-one”). When both a bottom-up input pattern and a top-down expectation are both active, feature cells that receive both bottom-up excitatory inputs and top-down excitatory priming signals can fire (“two-against-one”), while cells in the off-surround 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.

#### 4.5. Art hypothesis testing and learning of a predictive recognition category

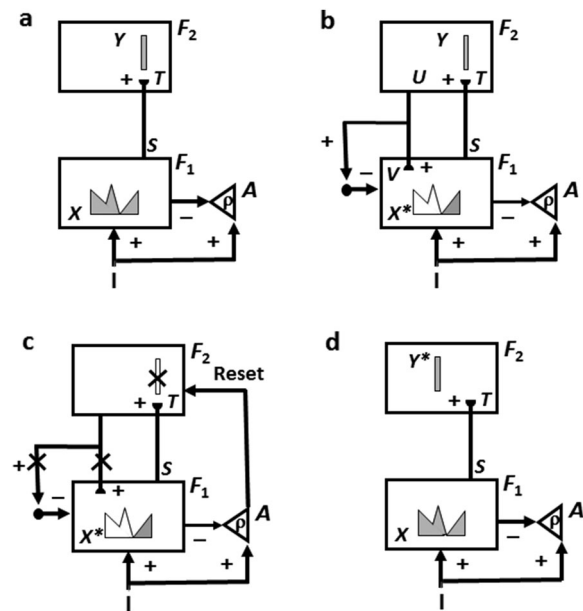
ART builds upon these bottom-up category selection and top-down expectation matching processes to search for, and learn, new recognition categories using cycles of resonance and reset. A new recognition category is said to form when no previous category is activated in response to a particular input pattern, and the search process leads to previously uncommitted cells with which to represent the input pattern.

Figure 3 schematises how the ART hypothesis testing and learning cycle this is proposed to occur in the simplest two-level network. Figure 3(a) shows that an input pattern  $I$  is transformed across feature detectors at the feature level  $F_1$  into an activity pattern  $X$ , while

its inputs activate the orienting system  $A$  with a gain  $\rho$  that is called vigilance. Active feature detectors in  $X$  inhibit  $A$  while they generate a bottom-up output pattern  $S$ .  $S$  is multiplied by learned adaptive weights within the bottom-up adaptive filter to form the input pattern  $T$  that activates the category level  $F_2$ . At  $F_2$ , the category cells compete to pick a winning category  $Y$ . This competition is realised by a recurrent on-center off-surround network that contrast-enhances and normalises its input patterns. This contrast-enhancing process compresses the input pattern into a category choice. In the case of maximal compression, a winner-take-all choice is made for which only the cell population that received the biggest total input survives the competition. Such a choice is a kind of “grandmother cell”.

Figure 3(b) shows that the chosen category  $Y$  generates top-down signals  $U$  that are multiplied by adaptive weights to form the prototype  $V$ . Prototype  $V$  encodes the learned expectation of the active  $F_2$  category. The prototype signals are positive at the critical features that are also used to choose  $Y$  via the bottom-up filter. Expectation  $V$  is matched at  $F_1$  with the bottom-up input pattern  $I$  using the ART Matching Rule. If  $V$  mismatches  $I$  at  $F_1$ , then a subset of the short-term memory (STM) activity pattern  $X$ , which is denoted by  $X^*$  (the gray pattern), survives the matching process at  $F_1$ . The subset pattern  $X^*$  is active at cells whose input features  $I$  are confirmed by  $V$ , while mismatched features (white area) are inhibited by the off-surround.

When  $X$  changes to  $X^*$ , total inhibition decreases from  $F_1$  to  $A$ . Figure 3(c) shows that, if inhibition



**Figure 3.** How ART searches for a recognition category using cycles of resonance and reset. See text for details. Adapted with permission from Carpenter and Grossberg (1993).

decreases sufficiently to satisfy the vigilance parameter  $\rho$  (i.e. there is not enough inhibition from  $X^*$  to overcome the excitation by  $I$ ), then  $A$  releases a nonspecific arousal burst to  $F_1$ ; that is, “novel events are arousing”. Arousal resets  $F_2$  by inhibiting the currently active category  $Y$ . Intuitively, if the expectation that is read-out by  $Y$  is disconfirmed by the currently active input, then  $Y$  is inhibited and a search for a better-matching category is instated.

Figure 3(d) shows that, after  $Y$  is inhibited,  $X$  is disinhibited, and thus reinstated. The inhibited category  $Y$  then stays inhibited as  $X$  activates a different category  $Y^*$ . Search for a new  $F_2$  category continues until a better-matching or novel category is selected. When search ends, an attentive feature-category resonance triggers fast learning of the attended data in the adaptive weights within both the bottom-up adaptive filter and the top-down expectation.

It is because a feature-category resonance can trigger fast learning of the attended, or matched, feature pattern, that the name *adaptive resonance theory* is used. Resonance can only occur if the bottom-up and top-down patterns are sufficiently well matched. In order for a top-down expectation of a totally new recognition category to match the features that activate it, all its top-down adaptive weights initially have large values, so that it can match *any* feature pattern. If this did not happen, then the top-down expectation could mismatch the novel input pattern, thereby triggering a search, and preventing the system from ever settling down to learn a new category. These adaptive weights are pruned as learning proceeds, leading to the selection of an attended critical feature pattern across learning trials with that category.

It has been mathematically proved that this kind of *match learning* within an ART model can solve the stability–plasticity dilemma by enabling the learning of stable categories which code arbitrary sequences of events presented in any order; e.g. Carpenter and Grossberg (1987, 1991). This kind of category learning goes on within the *attentional system* of ART.

ART shows how memory search overcomes a potentially devastating defect of match learning: If learning can occur only if there is a sufficiently good match, then how is anything really novel ever learned? Here is where the complementary properties of attentional matching and orienting search are crucial. As noted above, a sufficiently bad mismatch in the attentional system between an active top-down expectation and a bottom-up input, say because the input represents an unfamiliar experience, can drive a memory search that activates the complementary *orienting system*, which is sensitive to unexpected and unfamiliar events.

Output signals from the orienting system rapidly reset the recognition category that has been reading out the poorly matching top-down expectation. The cause of the mismatch is hereby removed, thereby freeing the system to activate a different recognition category. As noted in Figure 3(d), such a reset event triggers memory search, or hypothesis testing, which automatically leads to the selection of a recognition category that can better match the input. If no such recognition category exists, say because the bottom-up input represents a truly novel experience, then the search process automatically activates an as yet uncommitted population of cells with which to learn to categorise the novel information.

When a top-down expectation mismatches a bottom-up input pattern, how many categories need to be searched to find a known category, or uncommitted population of cells, with which to match and learn the input pattern? The first reply is that the search is selective, not exhaustive. It only selects category cells whose top-down expectations are as close to the input pattern as possible. In particular, when an input pattern is presented to the system, it may mismatch an active top-down expectation just because of what the system was previously processing. Then the next category to be activated will be the one that best matches the new input pattern, if the input pattern is familiar. The second reply is that system parameters can be chosen to create very fast searches, with few if any additional mismatches, by biasing the search to choose uncommitted cells right after a mismatch occurs (Carpenter & Grossberg, 1987, Section 25).

ART suggests that the orienting system includes the nonspecific thalamus and the hippocampal system. Carpenter and Grossberg (1993) and Grossberg and Versace (2008) summarise psychological, psychophysiological, and neurophysiological data that support this prediction.

#### 4.6. Memory consolidation and fast bottom-up recognition of the best category

As sequences of inputs are practised over learning trials, the search process eventually converges upon stable categories. This is a form of dynamically regulated *memory consolidation*. Carpenter and Grossberg (1993) and Franklin and Grossberg (in press) describe how ablations of the hippocampal system, which is part of the orienting system, can lead to learning and memory problems that are found in ablated animals and patients with medial temporal amnesia.

During category learning in normal individuals, it has been mathematically proved (e.g. Carpenter & Grossberg, 1987) that familiar inputs directly access the

category whose prototype provides the globally best match, without undergoing any search, while unfamiliar inputs may continue to activate the orienting system to trigger memory searches for better categories until they too become familiar. In other words, ART provides a solution of both the *catastrophic forgetting problem* and the *local minimum problem* that various other algorithms, such as back propagation (Baldi & Hornik, 1989; Gori & Tesi, 1992), do not solve. This process of search and category learning continues until the memory capacity, which can be chosen arbitrarily large, is fully utilised.

The direct one-pass bottom-up access to familiar categories also clarifies data showing rapid one-shot classification of familiar scenes (Thorpe, Fize, & Marlot, 1996). In ART, once learning is complete, top-down expectations merely confirm the bottom-up choice, rather than requiring more search.

#### 4.7. Concrete or general grandmother cells: vigilance control

What combinations of features or other information are bound together into categories by the search and learning cycle in Figure 3? One critical factor is the gain control process, called *vigilance* control, which can be influenced by environmental feedback or internal volition (Carpenter & Grossberg, 1987). Low vigilance permits the learning of general categories with abstract prototypes. High vigilance forces a memory search to occur for a new category when even small mismatches exist between an exemplar and the category that it activates. As a result, in the limit of high vigilance, the category prototype may encode the critical feature pattern of an individual exemplar. Significantly, either concrete or general categories can be learned using winner-take-all choices at the category level. Thus, one can learn either a concrete or general critical feature pattern with which to recognise a particular grandmother (Grandma Leitner!) or a general grandmother using such a process.

Vigilance is computed within the orienting system of an ART model (Figure 3(b)–(d)). It is here that bottom-up excitation from all the active features in an input pattern  $I$  are compared with inhibition from all the active features in a distributed feature representation  $X$  or  $X^*$  across  $F_1$ . If the ratio of the total activity across the active features in  $F_1$  (i.e. the “matched” features) to the total activity due to all the features in  $I$  is less than the vigilance parameter  $\rho$  (Figure 3(b)), then a nonspecific reset wave is activated (Figure 3(c)) that can drive the search for another category with which to classify the exemplar.

In particular, let  $\rho$  multiply the bottom-up inputs  $I$  to the orienting system; that is,  $\rho$  is the *gain* of the inputs to the orienting system. The orienting system is then activated when the total excitatory input  $\rho|I|$  is greater than the total inhibition  $|X^*|$  from the features  $X^*$  across  $F_1$  that survive top-down matching; that is, when  $\rho|I| - |X^*| > 0$ , where  $| \cdot |$  denotes the number of positive inputs or matched features. This inequality can be rewritten as  $\rho > |X^*| |I|^{-1}$  to show that the orienting system is activated whenever  $\rho$  is chosen higher than the ratio of active  $X^*$  matched features in  $F_1$  to total features in  $I$ . In other words, the vigilance parameter controls how bad a match will be tolerated before search for a new category is initiated. If the vigilance parameter is low, then many exemplars can all influence the learning of a shared prototype, by chipping away at the features that are not shared with all the exemplars, thereby leading to learning of an abstract prototype whose critical features are shared among all the coded exemplars. If the vigilance parameter is high, then even a small difference between a new exemplar and a known prototype (e.g. F vs. E) can drive the search for a new category with which to represent F. In all of these cases, the system may be tuned, if desired, to make winner-take-all, and thus highly selective, category choices, thereby learning a kind of “grandmother cell”. If the recurrent competition for activity at the category level is less severe, then distributed patterns of categories can be chosen to represent the system’s input patterns.

#### 4.8. Minimax learning via match tracking: learning the most general predictive categories

One dynamical way to control vigilance is by *match tracking* (Carpenter & Grossberg, 1991, 1993). During match tracking, in response to a predictive error (e.g. the category that previously learned to code E is chosen and triggers an “E” prediction in response to the input pattern F), the vigilance parameter  $\rho$  increases just enough to trigger reset and search for a better-matching category. Match tracking gives up the minimum amount of generalisation in the learned categories to search for a better-matching category. In other words, vigilance “tracks” the degree of match between input exemplar and matched prototype. Because match tracking increases vigilance by the minimum amount to trigger a reset and search for a new category, it realises a Minimax Learning Rule that conjointly *maximises* category generality while it *minimises* predictive error. Match tracking thus uses the least memory resources that can correct errors in classification.

The baseline level of vigilance is initially set at the lowest level that has led to predictive success in the past, so that ART models try to learn the most general



category that is consistent with the data. This tendency can, for example, lead to the type of overgeneralisation that is seen in young children until further learning leads to category refinement. However, because vigilance can vary during match tracking in a manner that reflects current predictive success, recognition categories capable of encoding widely differing degrees of generalisation or abstraction can be learned by a single ART system. Thus a single ART system may be used, say, to learn abstract prototypes with which to recognise abstract categories of faces, dogs, or grandmothers, as well as “exemplar prototypes” with which to recognise individual views of these objects, depending on task requirements.

#### **4.9. Cholinergic modulation of category learning via nucleus basalis vigilance control**

Finer details about how vigilance may be controlled have been proposed by the Synchronous Matching ART, or SMART, model (Grossberg & Versace, 2008). SMART predicts how vigilance may be altered by acetylcholine (ACh) when the nucleus basalis of Meynert is activated via the nonspecific thalamus (van Der Werf, Witter, & Groenewegen, 2002; Kraus, McGee, Littman, Nicol, & King, 1994) which, in turn, is activated by corticothalamic mismatches with one or more specific thalamic nuclei (Figure 4).

In general, it is known that cholinergic modulation is an essential ingredient in cortical plasticity (e.g. Kilgard & Merzenich, 1998). Saar, Grossman, and Barkai (2001) have shown, in addition, that acetylcholine (ACh) release reduces the after-hyperpolarization (AHP) current and increases cell excitability in layer 5 cortical cells. In SMART, this increased layer 5 excitability due to predictive mismatch may cause reset via the layer 5-to-6<sup>1</sup>-to-4 circuit that realises part of the ART Matching Rule (Figure 4), even in cases where top-down feedback may earlier have sufficiently matched bottom-up input, which is a key property of vigilance control. The increase of ACh hereby promotes search for finer recognition categories in response to environmental feedback, even when bottom-up and top-down signals have a pretty good match in the nonspecific thalamus based on similarity alone. Recent neurobiological experiments suggest that increased ACh (and attention) may refine perceptual representations by adding specificity. Palma, Grossberg, and Versace (2012) and Palma, Versace, and Grossberg (2012) review these experiments and simulate recurrent on-center off-surround networks composed of spiking shunting neurons to illustrate how ACh may modulate the transformation and STM storage of input

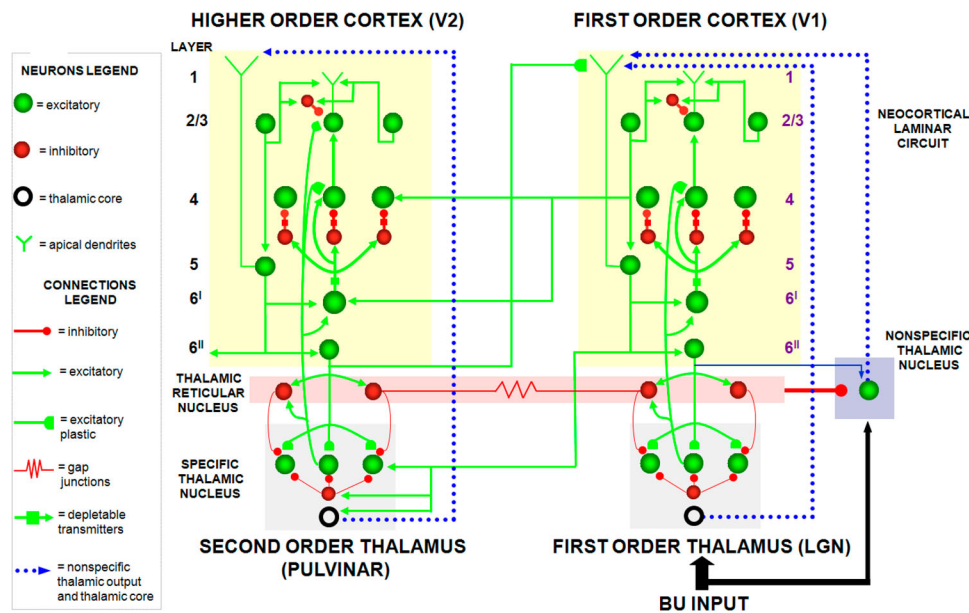
patterns in a manner compatible with how vigilance control can explain these data.

In summary, recent variants of ART show how spiking neurons in laminar cortical circuits with identified neurons may learn concrete or general recognition categories via a cycle of attentional matching and orienting search that is regulated by cholinergically modulated vigilance control. All of these categories are learned using localist representations that may, in the limit of strong contrast-enhancement by a category-selecting recurrent on-center off-surround network, encode their critical feature patterns using grandmother cells.

#### **4.10. How distributed does a recognition category need to be?**

Can a single winner-take-all category code all the critical features that may be needed to control and predict all aspects of an object's recognition during goal-oriented behaviours with it? In this regard, it should be noted that distributed ARTMAP shows how, during supervised learning, which is sensitive to the predictive success of a recognition category in generating a desired outcome, the ART learning cycle tries to discover the optimal degree of compression that is needed to categorise the predictively successful critical feature patterns that are hidden within the database that is currently being learned (Carpenter, 1997; Carpenter, Milenova, & Noeske, 1998). Various neurophysiological data suggest that recognition of a complex object, such as a face, may be distributed over a population of cells (Gross, 2008). Such a conclusion needs to be approached with caution for several reasons. First, laminar cortical models have illustrated how similar circuit designs, suitably specialised at different cortical levels, can carry out distinct category learning processes. Very simple categories can be learned in cortical area V1, including the orientationally and positionally selective simple and complex cells, which approximate grandmother cells (cf. Bowers, 2009).

How such cells can develop has been described in laminar cortical models of Grossberg and Seitz (2003) and Grossberg and Williamson (2001) using the same kinds of laminar ART, or LAMINART, circuits that may also be used for higher-order types of category learning. In addition, as this article reviews below, object recognition becomes increasingly invariant at successive processing stages in IT cortex and beyond, beginning with more view-, size-, and position-selective categories and progressing to increasingly view-, size-, and position-invariant categories. Although, taken together, all of these cells represent a distributed categorical representation of the object category, each of them could, in



**Figure 4.** The SMART model clarifies how laminar neocortical circuits in multiple cortical areas interact with specific and nonspecific thalamic nuclei to regulate learning on multiple organisational levels, ranging from spikes to cognitive dynamics. The thalamus is subdivided into specific first-order and second-order nuclei, nonspecific nucleus, and thalamic reticular nucleus (TRN). The first-order thalamic matrix cells (shown as an open ring) provide nonspecific excitatory priming to layer 1 in response to bottom-up input, priming layer 5 cells and allowing them to respond to layer 2/3 input. This allows layer 5 to close the intracortical loop and activate the pulvinar (PULV). V1 layer 4 receives inputs from two parallel bottom-up thalamocortical pathways: a direct LGN→4 excitatory input, and a  $6^I \rightarrow 4$  modulatory on-center, off-surround network that contrast-normalises the pattern of layer 4 activation via the recurrent  $4 \rightarrow 2/3 \rightarrow 5 \rightarrow 6^I \rightarrow 4$  loop. V1 activates the bottom-up V1→V2 corticocortical pathways from V1 layer 2/3 to V2 layers  $6^I$  and 4, as well as the bottom-up corticothalamocortical pathway from V1 layer 5 to the PULV, which projects to V2 layers  $6^I$  and 4. In V2, as in V1, the layer  $6^I \rightarrow 4$  pathway provides divisive contrast normalisation to V2 layer 4 cells. Corticocortical feedback from V2 layer  $6^{II}$  reaches V1 layer 1, where it activates apical dendrites of layer 5 cells. Layer 5 cells, in turn, activate the modulatory  $6^I \rightarrow 4$  pathway in V1, which projects a V1 top-down expectation to the LGN. TRN cells of the two thalamic sectors are linked via gap junctions, which synchronise activation across the two thalamocortical sectors when processing bottom-up stimuli. The nonspecific thalamic nucleus receives convergent bottom-up excitatory input from specific thalamic nuclei and inhibition from the TRN, and projects to layer 1 of the laminar cortical circuit, where it regulates mismatch-activated reset and hypothesis testing in the cortical circuit. Corticocortical feedback connections from layer  $6^{II}$  of the higher cortical area terminate in layer 1 of the lower cortical area, whereas corticothalamocortical feedback from layer  $6^{II}$  terminates in its specific thalamus and on the TRN. This corticothalamocortical feedback is matched against bottom-up input in the specific thalamus. Reprinted with permission from Grossberg and Versace (2008).

principle, be coded by a compact sets of cells, albeit sets of cells that are sensitive to different object properties, much as Hasselmo, Rolls, and Baylis (1989) have observed in the case of face recognition that cells selective for facial expression tend to be located within the superior temporal sulcus, whereas cells selective for identity tend to be located on the inferior temporal gyrus. As noted above, such a distributed network of compressed categories with which to represent an object or event is herein called a grandmother cohort.

#### 4.11. Bound states and resonant hierarchies of grandmother cohorts

The concept of a grandmother cell, or even a grandmother cohort, may lead one to mistakenly have a passive picture of categorisation in mind. Such a viewpoint has a strong antidote when one remembers that

even a grandmother cohort is part of bound states in resonant hierarchies that include multiple cortical regions.

This can most naturally be understood by noting that a basic kind of complementary ignorance occurs during category learning, and is overcome through adaptive resonance. When a feature pattern is activated at a level like  $F_1$  (Figure 3), its individual features have no meaning for the same reason that individual pixels in a picture are meaningless. A picture become meaningful only through spatial context to which its individual pixels belong. The same is true for features at a level like  $F_1$ . A recognition category that is selectively activated at a level like  $F_2$  provides such a context, but there is no information at this level about what features are represented by the category. 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

$F_2$ . 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 resonance hereby becomes a coherent and synchronous *bound state* that carries information about both what category is active and what features are represented by it. Without such a resonant bound state, even a category as unambiguous as a grandmother cell cannot, by itself, answer the question: What do I represent? In addition to such bound states between individual categories and their distributed features, the SMART model (Grossberg & Versace, 2008) proposes how hierarchies of cortical areas may synchronise and thereby bind together complex grandmother cohorts with the critical feature patterns that embody what they represent.

## 5. Learning of invariant object categories using Where-to-What stream interactions

### 5.1. From view-specific to view-invariant category learning

The discussion about ART above considered only one form of object attention (Posner, 1980), the kind that focuses attention upon the critical feature pattern of a view- and position-specific category prototype. In order to explain how invariant category learning occurs, ARTSCAN proposes how this kind of object attention in the What cortical stream, which I call *prototype attention* (Section 3), interacts with a particular form of spatial attention (Duncan, 1984) in the Where/How cortical stream, which I call *surface attention*, to learn invariant object categories during free viewing of a scene with eye movements. The ART dynamics schematised in Figure 3 learn the view- and position-specific categories that are bound together through the kind of coordination shown in Figure 2 into view- and position-invariant object categories, where the degree of position-invariance is limited by the cortical magnification factor and the multiple spatial scales that underlie it (Grossberg et al., 2011).

This binding process begins when a view-specific category of a novel object is learned. As the eyes move around an object surface, multiple view-specific categories are learned of the object, say in ITp, and are associated with the emerging invariant object category that forms, say in ITa (see Figure 2). When all of these categories have been learned, the object will be represented by a distributed network of categories each of which is itself represented by a compact set of cells, that is, by a grandmother cohort.

Imagine that the first view-specific category to be activated in ITp also activates some cells in ITa that will

become the view-invariant category via associative learning with multiple view-specific categories. As the perceived view of the object changes sufficiently due to its relative motion with respect to the observer, this view-specific category will be reset to enable the next view-specific category to be activated and learned. When the first view-specific category is reset, the input from ITp to ITa that initially activated the emerging view-invariant category is removed. How does the brain prevent the invariant object category from also being inhibited at this time, if only due to the collapse of its original source of activation? How does the brain maintain activity of the active ITa cells while they are associated with multiple view-specific categories of a single object, so that the ITa cells can learn to become a view-invariant category with which to represent this object?

### 5.2. Attentional shroud inhibits invariant object category reset during object search

As noted in Section 4.2, a surface-shroud resonance achieves shroud persistence using positive feedback signals between the attended surface representation in V4 and the shroud's surface-fitting spatial attention in PPC, combined with competitive interactions across PPC. The winning shroud can hereby focus and sustain spatial attention upon the object to be learned. The active shroud protects the view-invariant category from getting reset, even while view-specific categories are reset, by inhibiting the reset mechanism while spatial attention is focused on the object, via the inhibitory connections from Spatial Attention to Category Reset in Figure 2. Functional magnetic resonance imaging (fMRI) data support the hypothesis that the reset mechanism is found in the parietal cortex, as is the attentional shroud, as noted below.

Figure 2 illustrates how an active surface-shroud resonance also supports several other processes: as it maintains inhibition of the reset mechanism while attention remains focused on the object surface, it generates a sequence of saccadic eye movement targets to explore the attended surface (see Salient Features and Target Position stages in the model circuit), and it uses corollary discharges from these target position commands to predictively update gain fields that maintain the stability of the shroud in head-centered coordinates as the eyes move. As will be discussed in greater detail below, all of these processes involve distributed cortical representations that are computed using localistic laws.

When spatial attention shifts from an object, its shroud collapses, thereby disinhibiting the Category Reset stage in PPC. 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 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.

### **5.3. Human and monkey data support shroud reset properties: a further test**

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 corresponding to a shift in categorisation 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 categorisation rules). The transient signal is “domain-independent” because the parietal reset mechanism can be inhibited by spatial attention to *any* object, and can reset *any* active invariant category when it is disinhibited. These data illustrate the fact that both sustained (shroud) and transient (reset) mechanisms in parietal cortex are used to control how spatial attention focuses on objects and switches between them, again illustrating how cortical representations may be spatially distributed even though they obey localist laws.

Cao et al. (2011) have developed the positional ARTSCAN, or pARTSCAN, extension of the ARTSCAN model to explain how both view- and position-invariant categories can be learned, and have used this extended model to simulate neurophysiological data of Li and DiCarlo (2008; see also Li & DiCarlo, 2010) which shows that features from different objects can be merged within invariant IT categories 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 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 test more

fully how spatial attention may modulate the learning of invariant object categories. The prediction is that there would not be a transient parietal burst under the usual target swapping conditions, hence features from different objects could be merged through learning. However, 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 be either completely prevented, or at least significantly attenuated.

### **5.4. Predictive remapping by gain fields maintains shroud stability as the eyes move**

When an eye movement occurs on an object surface, why does not the shroud shift so much that it is spuriously reset? ARTSCAN proposes that this is accomplished by computing shrouds in head-centered coordinates that do not move when the eyes move. This is proposed to occur by transforming surface representations, which are computed in retinotopic coordinates, into head-centered coordinates. The gain field signals that update a head-centered shroud are assumed to occur very quickly, even 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). As noted above, predictive remapping is often explained as being achieved by *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 lateral intraparietal (LIP) cortex in Figure 2. Gain fields are populations of cells that are activated by outflow eye movement signals and used to transform retinotopic coordinates into head-centered coordinates.

Conscious visual percepts shift with the positions foveated by the eyes, and are thus computed in retinotopic coordinates. Shrouds, in contrast, are computed in head-centered coordinates. Data of Burr and Morrone (2011) illustrates this subtlety, while also supporting the above conclusion that an attentional shroud is computed in spatial coordinates: “We firstly report recent evidence from imaging studies in humans showing that many brain regions are tuned in spatiotopic coordinates, but only for items that are actively attended”. The

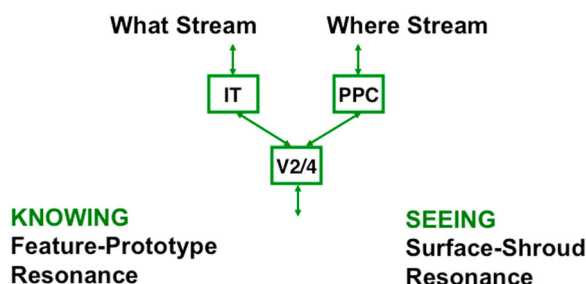


coexistence of both retinotopic and head-centered representations of object information illustrates yet another sense in which object representations can be distributed, yet defined by localist laws.

### 5.5. Synchronous resonances for seeing and knowing

When we consciously see a familiar object, we also typically *know* what it is. ART proposes that these two kinds of awareness are due to different kinds of resonances, with conscious seeing supported by surface-shroud resonances, and knowing what is seen is supported by visual feature-category resonances (Section 4.4). ART proposes that both surface-shroud resonances and feature-category resonances interact with shared visual cortical areas, such as V4. Because of this shared featural interface, both kinds of resonances can synchronise with each other, often with gamma oscillations (Fries, 2009; Grossberg & Versace, 2008). Surface-shroud resonances include Where/How stream regions such as PPC, whereas feature-category resonances include What stream regions such as IT (Figure 5). Thus, when we know what we see, a resonance may synchronously bind What stream recognition categories with Where/How stream spatial attentional resources, all converging on visual surface representations that can support conscious qualia. This is yet another kind of distributed representation of an object, whose various processes again obey localist laws and whose categories together form a grandmother cohort.

The above discussion also illustrates how, even when winner-take-all categories exist at multiple levels of a recognition hierarchy, widespread distributed representations in multiple brain regions, and across complementary cortical processing streams, interact to enable these categories to be rapidly learned and stably remembered.



**Figure 5.** 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.

## 6. Item-list resonances for recognition of speech and language

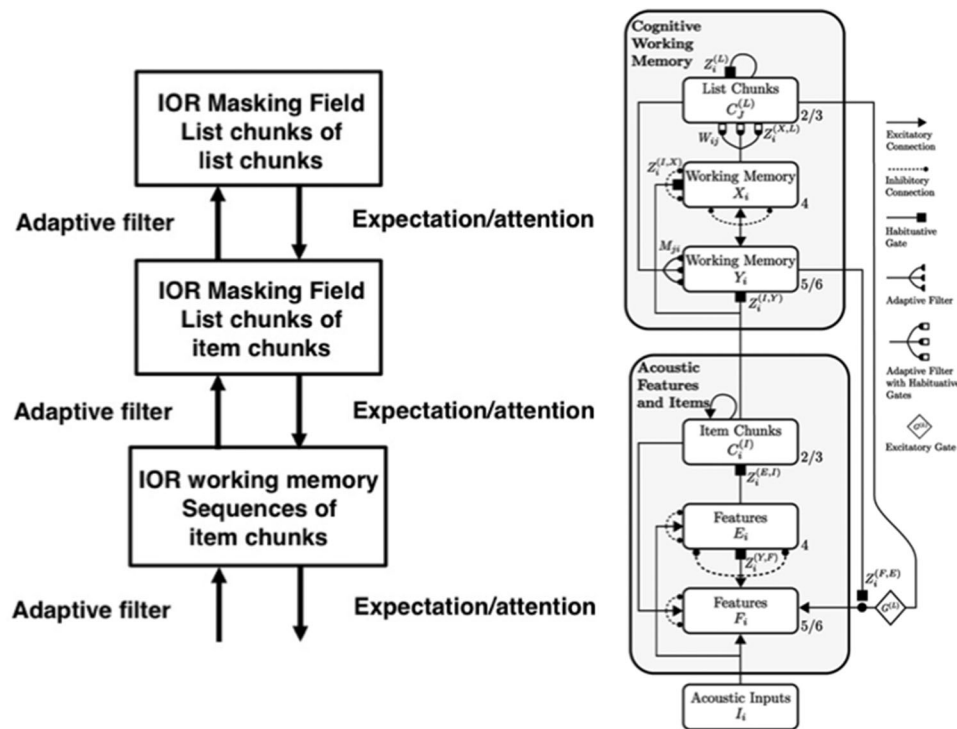
### 6.1. ART in the temporal domain: Item-and-Order working memories and list chunks

ART mechanisms also help to clarify how speech and language are rapidly learned, stably remembered, flexibly performed, and consciously heard. These ART models form part of an increasingly comprehensive neural theory of speech learning, perception, recognition, and recall whose various models go by names such as the PHONET, ARTPHONE, ARTWORD (Figure 6, left panel), and conscious ARTWORD, or cARTWORD (Figure 6, right panel). They clarify how sequences of acoustic *item chunks* that are stored temporarily in a working memory are categorised by *list chunks*. The term working memory is here used to mean the temporary storage of *sequences* of events through time, not just the persistent storage of a single event. The list chunks are sequence categories that may approximate grandmother cells when the amount of available evidence is sufficient to fully activate them. What sorts of neural circuits realise working memories and list chunks?

Grossberg (1978a, 1978b) introduced a neural model of working memory on which the more recent models listed above 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 unities item chunks of individual sounds. Each item chunk 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 these item-selective cells (Figure 7). The *relative activities* of different cell populations code the temporal order in which the items will be rehearsed, with the largest activities rehearsed earliest; hence, the name *Item-and-Order* working memory for this class of models. Another name for an Item-and-Order working memory is *competitive queuing* (Houghton, 1990); see Section 6.6.

A *primacy gradient*, in which the first item stored has the largest activity, the second the next largest activity, and so on, can be recalled in the correct order. When an Item-and-Order working memory can also store repeated items in a sequence, it is called an *Item-Order*

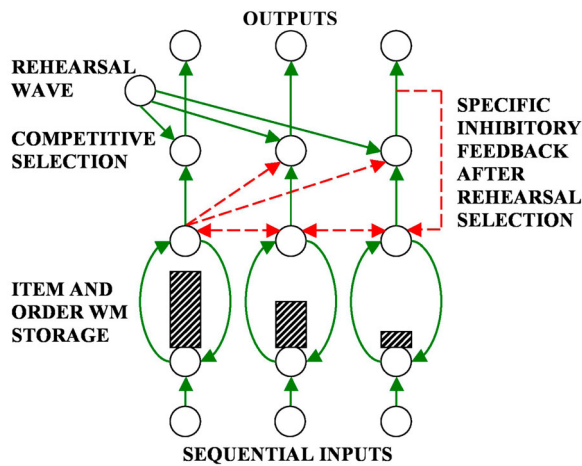


**Figure 6.** (Left panel) Elaboration of the ARTWORD model macrocircuit. Interactions among three speech processing levels are capable of working memory storage, chunking, and performance of word sequences. Each level consists of an Item-Order-Rank working memory. The second and third levels are, in 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 of item chunks. The inputs from the second level to the third level enable it to store list chunks of list chunks, in particular sequences of words. (Right panel) cARTWORD model circuit. cARTWORD includes a hierarchy of two cortical processing levels that model different cortical regions. Each level is organised into laminar cortical circuits that share a similar laminar organisation. Cells in both levels are organised into 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 categorise, or chunk, distributed patterns across these deeper layers into unitised 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. Reprinted with permission from Grossberg and Kazerounian (2011).

*Rank* working memory (Bradski, Carpenter, & Grossberg, 1994; Grossberg & Kazerounian, 2016; Grossberg & Pearson, 2008; Silver, Grossberg, Bullock, Histed, & Miller, 2011). Item-Order-Rank working memories, including the role of primacy gradients in correct list storage and recall, are supported by both psychophysical and neurophysiological data (e.g. Averbek, Chafee, Crowe, & Georgopoulos, 2002; Farrell & Lewandowsky,

2002; Grossberg & Pearson, 2008; Page & Norris, 1998; Silver et al., 2011).

These stored working memory sequences are unitised 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;



**Figure 7.** 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. Adapted from Grossberg (1978a).

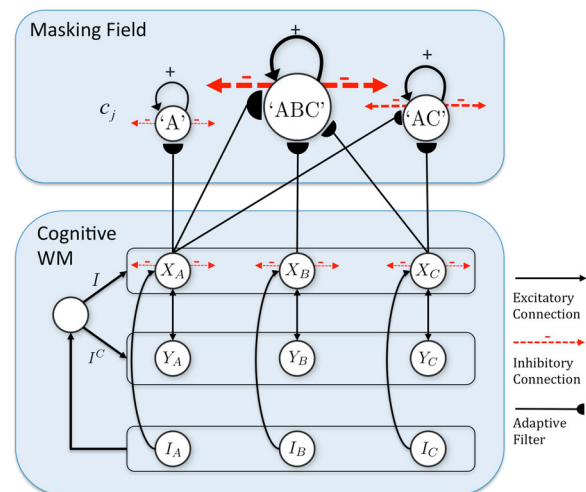
Grossberg & Kazerounian, 2011; Grossberg & Myers, 2000; Grossberg & Stone, 1986b; Kazerounian & Grossberg, 2014). They can approximate grandmother cells when they, and they alone, represent the sequence. Cohen and Grossberg (1986, 1987) have shown, however, how list chunks for sub-sequences and for the entire sequence may be co-activated to different degrees, depending upon the amount of evidence for them in the item chunk sequences that are currently stored in working memory. When all the items that are coded by a list chunk are currently active in working memory, and in the correct order, then a winner-take-all choice of this list chunk can occur. This is thus a grandmother cohort acting across time, as Section 6.5 further discusses.

## 6.2. Masking Fields: list chunks of variable-length sequences and the Magical Number 7

How are working memory sequences, such as syllables or words, selectively coded by list chunks? This can be achieved using a second Item-and-Order, or Item-Order-Rank, working memory that is called a Masking Field (Figure 8; Cohen & Grossberg, 1986, 1987; Grossberg, 1978a, 1984, 1986; Grossberg & Kazerounian, 2011, 2016; Grossberg & Myers, 2000; Kazerounian & Grossberg, 2014). A Masking Field is a specialised type of Item-and-Order working memory. As with all Item-

and-Order working memories, it is defined by a recurrent on-center off-surround network whose cells obey the membrane equations of neurophysiology. In a Masking Field, however, the "items" are *list chunks* that are selectively activated, via a bottom-up adaptive filter, by prescribed sequences of items that are stored in an Item-and-Order working memory at an earlier processing level (Figures 6 and 8). Masking Field cells represent list chunks because each of them is activated by a particular temporal sequence, or list, of items that is stored within the Item-and-Order working memory at the previous processing level.

Masking Field list chunks can represent lists (e.g. syllables or words) of multiple lengths because its cells interact within and between multiple spatial sizes, or scales. Cells of larger sizes selectively represent item sequences of greater length, and can inhibit smaller Masking Field cells that represent item sequences of lesser length. As items are stored in working memory, an adaptive filter activates the learned Masking Field list chunks that



**Figure 8.** An Item-and-Order working memory for the short-term sequential storage of item sequences can activate a multiple-scale Masking Field list chunking network through a bottom-up adaptive filter. As in Figure 6 (right column), the cognitive working memory uses two layers of cells with activities  $X$  and  $Y$ . The inputs are denoted by  $I$ . When embedded in a larger architecture like cARTWORD, these inputs are derived from item chunks, as in Figure 6 (right column). Adaptive filter weights from the  $X$  activities to the Masking Field learn to selectively activate list chunks within the Masking Field. For simplicity, the Masking Field shows a single list chunk that receives one input (for the list "A"), two inputs (for the list "AC"), or three inputs (for the list "ABC") from the cognitive working memory. The larger cell sizes and interaction strengths of the list chunks that categorise longer lists enable the Masking Field to choose the list chunk that currently receives the largest total input, and thus best predicts the sequence that is currently stored in the cognitive working memory. Reprinted with permission from Kazerounian and Grossberg (2014).

represent the most predictive item groupings at any time, while its recurrent inhibitory interactions suppress less predictive list chunks. Kazerounian and Grossberg (2014) have simulated how variable-length list chunks of a Masking Field can be learned as a list of items is stored in working memory in real time. See also Page and Norris (2009) for a variant of Masking Fields that can learn list chunks, and has been applied to explain the Hebb repetition effect and other properties of word learning.

An item that is stored in working memory is more properly called an *item chunk*, because, just like any chunk, it is a compressed representation, or category, of a spatial pattern of activity within a prescribed time interval. In the case of an item chunk, the spatial pattern of activity exists across acoustical feature detectors that process sounds through time, and which are compressed by an adaptive filter to activate item chunks. The prescribed time interval is short, commensurate with the duration of the shortest perceivable acoustic inputs, of the order of 10–100 msec. Thus the models in Figures 6 and 8 first compress spatial patterns of feature detectors into item chunks, and then compress sequences of the item chunks that are stored in working memory into list chunks.

### 6.3. Temporal chunking problem: learning words of variable length

Masking Fields were introduced to solve the *temporal chunking problem* (Cohen & Grossberg, 1986, 1987; Grossberg, 1978a, 1986). This problem concerns how an internal representation of an unfamiliar list of familiar speech units – for example, a novel word composed of familiar phonemes or syllables – can be learned under the type of unsupervised learning conditions that are the norm during daily experiences with language. Before a novel word, or list, can fully activate the adaptive filter, all of its individual items must first be presented. By the time the entire list is fully presented, all of its familiar sublists will thus have also been presented. A Masking Field prevents the familiarity of smaller sublists (e.g. MY, ELF, and SELF), which have already learned to activate their own list chunks, from forcing the novel longer list (e.g. MYSELF) to always be processed as a sequence of these smaller familiar chunks, rather than eventually as a newly learned unitised whole.

### 6.4. Self-similar competition solves the temporal chunking problem

A Masking Field accomplishes this using cells with multiple cell and receptive field sizes, or scales (Figure 8),

that are related to each other by a property of *self-similarity*; that is, each scale's properties, including its cell body sizes and their excitatory and inhibitory connection lengths and interaction strengths, are a multiple of the corresponding properties in another scale. Such a self-similarity property can develop as a result of simple activity-dependent growth laws (Cohen & Grossberg, 1986, 1987); see Grossberg and Kazerounian (2016) for a review.

Due to self-similar growth, larger list chunks selectively represent longer lists because they need more inputs, and thus more evidence, to fire. List chunks can hereby fire vigorously only when they receive enough evidence that the item sequence which they represent is currently stored in working memory. Once they fire, their stronger inhibitory interaction strengths than those of smaller list chunks can inhibit the smaller list chunks more than conversely ("asymmetric competition"). The properties of list chunk selectivity and asymmetric competition, acting together, realise the intuition that, other things being equal, the longest lists are better predictors of subsequent events than are shorter sublists, because a longer list embodies a more unique temporal context. The stronger inhibition from list chunks of longer, but unfamiliar, lists (e.g. MYSELF) enables them to inhibit the chunks that represent shorter, but familiar, sublists (e.g. MY), more than conversely, thereby enabling a solution of the temporal chunking problem.

These Masking Field properties naturally explain cognitive data such as the world word length effect during word superiority studies (Samuel, van Santen, & Johnston, 1982, 1983), the Magical Number Seven Plus or Minus Two of Miller (1956), who first asserted that "the memory span is a fixed number of chunks" (but see Baddeley, Thomson, & Buchanan, 1975; Grossberg & Pearson, 2008), and the Magical Number Four Plus or Minus One of Cowan (2001), who carefully distinguished between bottom-up and top-down signals as determinants of working memory storage. Grossberg and Kazerounian (2016) review these explanations, as well other speech and language data that ART concepts can explain.

### 6.5. The degree of list chunk compression covaries with the amount of evidence

The multiple-scale self-similar competition across a Masking Field is sensitive to working memory evidence that can partially activate its list chunks. The list chunks that represent sub-sequences and super-sequences of currently active working memory sequences can thus also be partially activated. Masking Field computer simulations (Cohen & Grossberg, 1986, 1987) demonstrated these properties. To summarise these simulations, call a



list chunk that codes a sequence of  $n$  items in working memory an  $n$ -chunk. Then, in a Masking Field whose longest list chunks categorise sequences of length 3 in working memory, activating one item in this sequence can maximally activate all the 1-chunks that code for this item, can somewhat less activate 2-chunks that include the item, and even less activate 3-chunks that also include that item. One can think of these 2-chunks and 3-chunks as being *primed* by a larger sequence that may or may not be stored in working memory during the next moments.

When a second item in the sequence is stored in working memory, the 2-chunk that categorises this sequence – including its order information – is maximally activated, whereas 1-chunks that receive inputs from the stored items and 3-chunks that include one or two of the items have commensurately less activation.

Finally, when all three items are stored in working memory, the 3-chunk that codes for all the items in the correct order wins the competition, and all other chunks are inhibited because this 3-chunk fully predicts the temporal context that the Masking Field can discriminate. The balance between input, recurrent excitation, and recurrent inhibition can also be altered by parameter changes to ensure that only the chunk whose sequence is active in working memory at any time can win the competition, and thereby suppress all less predictive chunks. These Masking Field simulations show how the degree of compression in a grandmother cohort network can covary with the amount of evidence for the sequences that are being categorised.

The SONNET 1 variant of the Masking Field model (Nigrin, 1990) has, for example, been used by Page (1994) to provide examples of how partial activation of list chunks can encode musical expectations that predict upcoming items in musical sequences.

### 6.6. A shared design for linguistic, motor, and spatial working memories: LTM invariance

How general can we expect these list chunk properties to be? This question reduces to a related question: What properties recommend Item-and-Order working memories more than other possible alternatives, over and beyond their broad explanatory and predictive range? Grossberg (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. These postulates are called the Long-term memory (LTM) Invariance Principle and the Normalisation Rule. They guarantee, for example,

that the first time a novel word, such as MYSELF, is stored in working memory, it does not destabilise previously learned list chunks that code for its familiar subwords MY, ELF, and SELF. Both postulates are automatically satisfied by a ubiquitously occurring neural design; namely, a recurrent on-center off-surround network (Figure 7) whose cells obey the membrane equations of neurophysiology, also called shunting dynamics.

When item 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. 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-normalise 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 Normalisation Rule. Taken together, the properties of competitive normalisation, rehearsal waves, and self-inhibitory feedback have explained many psychological and neurobiological data about working memory. Another name for the mechanism of choosing and inhibiting each item as it is rehearsed is *competitive queuing* (Houghton, 1990).

Indeed, the Grossberg (1978a) prediction that all working memories satisfy the LTM Invariance Principle and Normalisation Rule imply that all linguistic, spatial, and working memories should be realised by variants of recurrent shunting on-center off-surround networks with rehearsal wave and self-inhibitory regulation. It also implies that there should be a strong interaction between working memory and list chunking networks, since the working memories are designed to enable the list chunking networks to learn in a stable way. Thus, one would expect grandmother cohort properties of networks like Masking Fields to be found across all modalities that support working memory and list chunking. Grossberg and Kazerounian (2016) review psychological and neurobiological experimental evidence that support these predictions.

## 7. ART categories during object and list learning and performance

The above discussions illustrate how ART category learning mechanisms can enable the learning of specific and invariant object categories, and the learning of list chunks that represent sequences of items that are temporarily stored in working memory. In both domains, localist mechanisms of category learning are sufficient

to explain large psychological and neurobiological databases. The categories that are learned by these mechanisms tend to be compact, involving small numbers of cells or cell populations, even perhaps grandmother cells as a limiting case. Balanced against this property, however, is the fact that networks of such categories are often needed to represent multiple aspects of an object or of a stored sequence of events. Such networks are called grandmother cohorts to emphasise the balance between compression and distribution that is needed to learn predictive representations of the world in space and time.

## Disclosure statement

No potential conflict of interest was reported by the author.

## References

- Amis, G. P., Carpenter, G. A., Ersoy, B., & Grossberg, S. (2009). *Cortical learning of recognition categories: A resolution of the exemplar vs. prototype debate* (Technical report, CAS/CNS TR-2009-002). Boston, MA: Boston University.
- 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., Essick, G. K., & Siegel, R. M. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230, 456–458.
- 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.
- Averbeck, 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. doi:10.1073/pnas.162485599
- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575–589.
- Baldi, P., & Hornik, K. (1989). Neural networks and principal component analysis: Learning from examples and local minima. *Neural Networks*, 2, 53–58.
- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1, 371–394.
- Bhatt, R., Carpenter, G., & Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Research*, 47, 3173–3211. doi:10.1016/j.visres.2007.07.013
- Boardman, I., Grossberg, S., Myers, C., & Cohen, M. (1999). Neural dynamics of perceptual order and context effects for variable-rate speech syllables. *Perception & Psychophysics*, 61, 1477–1500. doi:10.3758/BF03213112
- Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, 8, 510–523. doi:10.1093/cercor/8.6.510
- Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, 116, 220–251, 215. doi:10.1037/a0014462
- 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.
- Burr, D. C., & Morrone, M. C. (2011). Spatiotopic coding and remapping in humans. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 366, 504–515. doi:10.1098/rstb.2010.0244
- Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and Da Vinci stereopsis. *Spatial Vision*, 18, 515–578. doi:10.1163/156856805774406756
- 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. doi:10.1016/j.neunet.2011.10.010
- 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. doi:10.1016/j.neunet.2011.04.004
- Carpenter, G. A. (1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks*, 10, 1473–1494.
- 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 amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences*, 16, 131–137.
- Carpenter, G. A., Milenova, B. L., & Noeske, B. W. (1998). Distributed ARTMAP: A neural network for fast distributed supervised learning. *Neural Networks*, 11, 793–813.
- 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.
- 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*. doi:10.3389/fnint.2014.0043
- Chiu, Y. C., & Yantis, S. (2009). A domain-independent source of cognitive control for task sets: Shifting spatial attention and switching categorization rules. *The Journal of Neuroscience*, 29, 3930–3938. doi:10.1523/JNEUROSCI.5737-08.2009
- Cohen, M. A., & Grossberg, S. (1986). Neural dynamics of speech and language coding: Developmental programs, perceptual grouping, and competition for short-term memory. *Human Neurobiology*, 5, 1–22.
- Cohen, M. A., & Grossberg, S. (1987). Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Applied Optics*, 26, 1866–1891.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114.

- Deneve, S., & Pouget, A. (2003). Basis functions for object-centered representations. *Neuron*, 37, 347–359.
- van Der Werf, Y. D., Witter, M. P., & Groenewegen, H. J. (2002). The intralaminar and midline nuclei of the thalamus. Anatomical and functional evidence for participation in processes of arousal and awareness. *Brain Research*, 39, 107–140. doi:10.1016/S0165-0173(02)00181-9
- Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society of London*, 353, 1245–1255. doi:10.1098/rstb.1998.0280
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46, 369–384.
- 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.
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517.
- Fang, L., & Grossberg, S. (2009). From stereogram to surface: How the brain sees the world in depth. *Spatial Vision*, 22, 45–82. doi:10.1163/156856809786618484
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin and Review*, 9, 59–79. doi:10.3758/BF03196257
- Fazl, A., Grossberg, S., & Mingolla, E. (2009). View-invariant object category learning, recognition, and search: How spatial and object attention are coordinated using surface-based attentional shrouds. *Cognitive Psychology*, 58, 1–48. doi:10.1016/j.cogpsych.2008.05.001
- Franklin, D. J., & Grossberg, S. (in press). A neural model of normal and abnormal learning and memory: Adaptively timed conditioning, hippocampus, amnesia, neurotrophins, and consciousness. *Cognitive, Affective, and Behavioral Neuroscience*.
- Fries, P. (2009). Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, 32, 209–224. doi:10.1146/annurev.neuro.051508.135603
- Gancarz, G., & Grossberg, G. (1999). A neural model of the saccadic eye movement control explains task-specific adaptation. *Vision Research*, 39, 3123–3143.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15, 20–25.
- Gori, M., & Tesi, A. (1992). On the problem of local minima in backpropagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 76–86.
- Gottlieb, J., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, 391, 481–484. doi:10.1038/35135
- Gross, C. G. (2002). Genealogy of the “grandmother cell”. *The Neuroscientist*, 8, 512–518. doi:10.1177/107385802237175
- Gross, C. G. (2008). Single neuron studies of inferior temporal cortex. *Neuropsychologia*, 46, 841–852. doi:10.1016/j.neuropsychologia.2007.11.009
- 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. doi:10.1093/cercor/11.1.37
- Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52, 213–257.
- 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*, 17, 199–219.
- Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87, 1–51.
- Grossberg, S. (1984). Unitization, automaticity, temporal order, and word recognition. *Cognition and Brain Theory*, 7, 263–283.
- 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. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55, 48–121.
- Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional figures. *Psychological Review*, 104, 618–658.
- 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. doi:10.1098/rstb.2008.0307
- Grossberg, S. (2013a). Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 37, 1–47. doi:10.1016/j.neunet.2012.09.017
- Grossberg, S. (2013b). Recurrent neural networks. *Scholarpedia*. Retrieved from [http://www.scholarpedia.org/article/Recurrent\\_neural\\_networks](http://www.scholarpedia.org/article/Recurrent_neural_networks)
- 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*. doi:10.3389/fpsyg.2015.02054
- Grossberg, S. (2016b). *Towards solving the hard problem of consciousness: The varieties variety of brain resonances and the conscious experiences that they support*. Manuscript submitted for publication.
- Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 481–503.
- Grossberg, S., & Huang, T.-R. (2009). ARTSCENE: A neural system for natural scene classification. *Journal of Vision*, 9(6), 1–19. doi:10.1167/9.4.6
- Grossberg, S., & Kazerounian, S. (2011). Laminar cortical dynamics of conscious speech perception: A neural model

- of phonemic restoration using subsequent context in noise. *The Journal of the Acoustical Society of America*, 130, 440–460. doi:10.1121/1.3589258
- Grossberg, S., & Kazerounian, S. (2016). Phoneme restoration and empirical coverage of interactive activation and adaptive resonance models of human speech processing. *Journal of the Acoustical Society of American*, 140(1130). doi:10.1121/1.494676
- 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. doi:10.1016/j.visres.2006.10.024
- Grossberg, S., & Kuperstein, M. (1986). *Neural dynamics of adaptive sensory-motor control: ballistic eye movements*. Amsterdam: North-Holland.
- 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. doi:10.1016/j.neunet.2011.04.001
- 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., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735–767. doi:10.1037/0033-295X.107.4.735
- Grossberg, S., & Pearson, L. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review*, 115, 677–732. doi:10.1037/a0012618
- Grossberg, S., & 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., & 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. doi:10.1093/cercor/13.8.852
- 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*. doi: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. doi:10.1016/j.visres.2003.12.009
- 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. doi:10.1016/j.brainres.2008.04.024
- Grossberg, S., & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research*, 45, 1725–1743. doi:10.1016/j.visres.2005.01.006
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, 32, 203–218.
- 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., & Grossberg, S. (2010). Cortical dynamics of contextually cued attentive visual learning and search: Spatial and object evidence accumulation. *Psychological Review*, 117, 1080–1112. doi:10.1037/a0020664
- 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*. doi:10.3389/fpsyg.2014.01053
- Kelly, F.J., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics*, 62, 1596–1618. doi:10.3758/BF03212158
- Kilgard, M. P., & Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, 279, 1714–1718.
- Konorski, J. (1967). *Integrative activity of the brain: An interdisciplinary approach*. Chicago, IL: University of Chicago Press.
- Kraus, N., McGee, T., Littman, T., Nicol, T., & King, C. (1994). Nonprimary auditory thalamic representation of acoustic change. *Journal of Neurophysiology*, 72, 1270–1277.
- 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. doi:10.1007/s10827-009-0211-1
- Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, 321, 1502–1507. doi:10.1126/science.1160028
- Li, N., & DiCarlo, J. J. (2010). Unsupervised natural visual experience rapidly reshapes size invariant object represent in inferior temporal cortex. *Neuron*, 67, 1062–1075. doi:10.1016/j.neuron.2010.08.029
- Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5, 552–563.
- Mathot, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, 21, 1793–1798. doi:10.1177/0956797610388813
- Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, 10, 903–907. doi:10.1038/nn1917
- Melcher, D. (2008). Dynamic, object-based remapping of visual features in trans-saccadic perception. *Journal of Vision*, 8, 2–2. doi:10.1167/8.14.2
- Melcher, D. (2009). Selective attention and the active remapping of object features in trans-saccadic perception. *Vision Research*, 49, 1249–1255. doi:10.1016/j.visres.2008.03.014



- Miller, G. A. (1956). The magical number seven plus or minus two. *Psychological Review*, 63, 81–97.
- 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.
- Nigrin, A. L. (1990). Real-time classification of temporal sequences with an adaptive resonance circuit. In *International joint conference on neural networks*, Washington, DC, 1 (pp. 525–528). Hillsdale, NJ: Erlbaum Associates.
- Page, M. P. A. (1994). Modelling the perception of musical sequences with self-organizing neural networks. *Connection Science*, 6, 223–246.
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781.
- Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B*, 364, 3737–3753. doi:10.1098/rstb.2009.0173
- 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. doi: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. doi:10.1007/s10827-011-0354-8
- Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London B*, 335, 23–30.
- Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1984). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B Biological Sciences*, 223, 293–317.
- Posner, M. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, 26, 381–410. doi:10.1146/annurev.neuro.26.041002.131112
- 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. doi:10.1016/j.neuron.2009.01.002
- 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.
- Saar, D., Grossman, Y., & Barkai, E. (2001). Long-lasting cholinergic modulation underlies rule learning in rats. *Journal of Neuroscience*, 21, 1385–1392.
- Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1982). Length effects in word perception: We is better than I but worse than you or them. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 91–105.
- Samuel, A. G., van Santen, J. P. H., & Johnston, J. C. (1983). Reply to Matthei: We really is worse than you or them, and so are ma and pa. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 321–322.
- Saygin, A. P., & Sereno, M. I. (2008). Retinotopy and attention in human occipital, temporal, parietal, and frontal cortex. *Cerebral Cortex*, 18, 2158–2168. doi:10.1093/cercor/bhm242
- 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. doi:10.1016/j.neunet.2011.10.004
- Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial vision processing in frontal cortex. *Nature*, 444, 374–377. doi:10.1038/nature05279
- 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. doi:10.3758/APP.72.3.597
- Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381, 520–522.
- 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.
- 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.
- Wannig, 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. doi:10.1038/nn.2910
- 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. doi:10.1523/JNEUROSCI.1897-07.2007