Incremental learning of perceptual and conceptual representations and the puzzle of neural repetition suppression

Stephen J. Gotts¹

Published online: 9 June 2016

© Psychonomic Society, Inc. (outside the USA) 2015

Abstract Incremental learning models of long-term perceptual and conceptual knowledge hold that neural representations are gradually acquired over many individual experiences via Hebbian-like activity-dependent synaptic plasticity across cortical connections of the brain. In such models, variation in task relevance of information, anatomic constraints, and the statistics of sensory inputs and motor outputs lead to qualitative alterations in the nature of representations that are acquired. Here, the proposal that behavioral repetition priming and neural repetition suppression effects are empirical markers of incremental learning in the cortex is discussed, and research results that both support and challenge this position are reviewed. Discussion is focused on a recent fMRI-adaptation study from our laboratory that shows decoupling of experience-dependent changes in neural tuning, priming, and repetition suppression, with representational changes that appear to work counter to the explicit task demands. Finally, critical experiments that may help to clarify and resolve current challenges are outlined.

Keywords Priming · Predictive coding · Plasticity · Sharpening · Synchrony · Semantic memory

The human cognitive system is made up of many constituent elements that are capable of a wide range of impressive abilities, including the rapid perception and recognition of objects,

fluid distillation of gist and meaning, sophisticated verbal communication, complex reasoning and planning, and flexible goal-based action with continual adjustment of motor behavior in response to a changing environment. All of these abilities require a complex base of long-term knowledge that has been acquired gradually over years in many different contexts and that is processed and retrieved in large part along conceptual and contextual dimensions (e.g., see Squire & Wixted, 2011, and Martin, 2007, for review). With the study of patients with amnesia and damage to the medial aspects of the temporal lobes, such as HM (see Cohen & Squire, 1980; Corkin, 1984), and patients suffering from forms of progressive dementia that attack the neocortex, such as "semantic dementia" (e.g., Hodges et al., 1992; Snowden, Goulding, & Neary, 1989; Warrington, 1975), our understanding of human learning and memory has been divided broadly into separate memory systems (e.g., Schacter, 1987; Squire, 1992; Tulving, 1972). Structures in the medial temporal lobes (e.g., hippocampus) are thought to be critical for the acquisition of "episodic" knowledge that is contextualized in time and place (Aggleton & Brown, 1999; Burgess, Maguire, & O'Keefe, 2002). In contrast, the neocortex is thought to store our long-term knowledge about the world that has been abstracted across individual experiences, with modality-specific perceptual knowledge residing within the corresponding sensory neocortical processing streams that help to mediate stimulus perception (e.g., Hasson et al., 2002; Levy et al., 2001; Martin, 2007; Martin et al., 1995; Simmons et al., 2005, 2007, 2013). Various aspects of conceptual knowledge are thought to be represented in select regions of the occipitotemporal, parietal, and frontal cortex, with an organization that is sensitive and selective to conceptual categories over multiple levels (e.g., Chao et al., 1999; Epstein, 2008; Gotts et al., 2011; Kanwisher & Yovel, 2006; Mahon et al., 2007; Mahon & Caramazza, 2009; Martin, 2007; Martin et al., 1996; Nieder & Dehaene,

✉ Stephen J. Gotts
gottss@mail.nih.gov

¹ Section on Cognitive Neuropsychology, Laboratory of Brain and Cognition, National Institute of Mental Health (NIMH), National Institutes of Health, Building 10, Rm 4C-104, Bethesda, MD 20892-1366, USA

2009; Taylor & Downing, 2011). While cognitive neuroscience and neuropsychology have helped to characterize the basic neuroanatomical and functional organization of these various sorts of long-term knowledge in the adult, the mechanisms that are responsible for the development and acquisition of this basic organization are much less clear. We do know that the large-scale brain organization of categories such as of faces, objects, number, and places continues to develop throughout childhood and into adolescence and adulthood (e.g., Cantlon et al., 2006; Cantlon et al., 2011; Golarai et al. 2007; Golarai et al., 2010; Scherf et al., 2007; Scherf et al., 2011), but what are the mechanisms and constraints that shape this organization?

In the current paper, I focus on a relatively simple but powerful notion that learning of this knowledge is incremental across experiences using Hebbian-like, activity-dependent plasticity at synaptic connections throughout the cortex, referred to hereafter as "incremental learning" (e.g., Becker et al., 1997; Friston, 2005; Friston & Kiebel, 2009; Jacobs, 1999; McClelland, McNaughton, & O'Reilly, 1995; McClelland & Rumelhart, 1985; McRae, de Sa, & Seidenberg, 1997; Oppenheim, Dell, & Schwartz, 2010; Plaut & Behrmann, 2011; Rogers & McClelland, 2004). A key feature of this idea is that different task demands can engage different cortical systems and cells within those systems, which, in turn, can qualitatively alter the nature of the neural representations that are acquired in order to support improved task performance (e.g., Farah & McClelland, 1991; Plaut, 2002; Rogers & McClelland, 2004; Tyler et al., 2000). I discuss research findings related to behavioral "repetition priming" and neural "repetition suppression," putative behavioral and neural correlates of incremental learning mechanisms. While repetition priming and repetition suppression have been reviewed a number of times previously (e.g., Gotts, Chow, & Martin, 2012a; Grill-Spector et al., 2006; Henson, 2003; Henson et al., 2014; Schacter & Buckner, 1998; Wiggs & Martin, 1998), their joint relationship to broader incremental learning proposals has not been evaluated to date.

The detailed organization of the paper is as follows: (1) A brief review of the behavioral phenomenon of repetition priming, (2) thereafter a review of models of incremental learning as an account of priming and an explanation of how different task contexts can qualitatively alter learned representations in the models, (3) review of the neural phenomenon of "repetition suppression" that commonly accompanies priming in neuroscience studies, highlighting the challenges that it poses for existing incremental learning models. (4) After providing an overview of this basic information, I then discuss neuroscience evidence that both supports and challenges a basic tenet of incremental learning models that task demands help to determine the nature of the representations that are acquired. (5) I close by outlining experiments that should help to resolve these challenges.

Repetition priming as incremental learning

In object identification tasks such as picture naming, stimulus repetition leads to faster and more accurate responses, a phenomenon referred to as "repetition priming" (see Schacter & Buckner, 1998; Tulving & Schacter, 1990, for review). Repetition priming is a robust and relatively automatic phenomenon, occurring across a wide range of tasks and sensory modalities of input and output. Priming is automatic in the sense that it is incidental to identification or other task processing and occurs even when participants' explicit recollection of the stimuli is at a chance level. These effects in object identification can be extremely long-lasting, persisting over days and months after only a small number of presentations (e.g., 1–2), although effects are strongest after little or no delay (Cave, 1997; McKone, 1995, 1998; Mitchell, 2006). Repetition priming effects are also robust to small changes in stimulus variables, such as size, position, or viewpoint in vision, and they transfer across different exemplars of the same type of object (Biederman & Cooper, 1991, 1992; Cave, Bost, & Cobb, 1996; Koutstaal et al., 2001; Srinivas, 1996). The long-lived nature of the phenomenon when using large stimulus sets clearly implicates some kind of long-term learning mechanism, which at the neural level might correspond to mechanisms such as long-term potentiation and depression (LTP/LTD; e.g., Bliss & Lomo, 1973; Mulkey & Malenka, 1992; Nabavi et al., 2014). The sparing of repetition priming in amnesia (e.g., Cave & Squire, 1992), in which patients typically have severe damage to the medial temporal cortex, also suggests a neocortical basis to the phenomenon, with perhaps some contribution of medial temporal lobe structures in neurotypical subjects (e.g., Henson & Gagnepain, 2010; Voss & Paller, 2008). With regard to the distinction between episodic versus long-term perceptual and conceptual knowledge, repetition priming is more squarely linked with changes in perceptual and conceptual knowledge. Consistent with this, repetition priming is decomposable into multiple components, with separate contributions of perceptual, conceptual, decision, and motor influences (e.g., Dobbins et al., 2004; Horner & Henson, 2008, 2012; Race, Shanker, & Wagner, 2009; Race, Badre, & Wagner, 2010; Schacter, Wig, & Stevens, 2007; Wig, Buckner, & Schacter, 2009; Wig et al., 2005). Priming is therefore maximized when stimulus and task are identical across repetitions and attenuated when these factors are changed in various ways.

But what is the functional role of repetition priming? While response times can improve with repetition on the order of 5–10 % (around an 80 ms improvement in picture naming, which has a base response time of around 700–800 ms), it is difficult to make the case that these effects are directly functionally relevant or critical to survival in some way. After all, inter-subject variability is typically quite a lot larger than the magnitudes of priming effects. Instead, priming has been

suggested to reflect tiny, incremental changes to representations with each experience, reflecting neocortical plasticity mechanisms (e.g., Becker et al., 1997; McClelland & Rumelhart, 1985; McClelland et al., 1995; Newman & Norman, 2010; Norman et al., 2006; Stark & McClelland, 2000). On this view, priming within an experiment is only a brief snapshot of the influences that shape representations over a much longer period of time. This interpretation also carries the consequence that there may be no direct "function" of priming, *per se*. Rather, it is a reflection of the ongoing operation of the brain's plasticity mechanisms that are fine-tuning existing representations each time a stimulus is encountered and/or forming new representations for novel stimuli.

Incremental learning models

Why should such learning be incremental? A forceful argument was put forward by Jay McClelland and colleagues (McClelland et al., 1995) within the framework of distributed connectionist neural network models. McClelland et al. made the case that certain kinds of learning are easily performed by artificial neural networks in single trials through the use of sparse representations. If patterns to be learned have little or no overlap (i.e., they are uncorrelated/ orthogonal to one another), a simple Hebbian plasticity rule – a local correlation of pre- and post-synaptic activity – can form such memories in a single exposure (e.g., Hopfield, 1982; Oja, 1982). Individual memories will not conflict or interfere with one another as long as they do not share cells/synapses in common. They made the case that this is precisely what is done by the hippocampus and related structures when encoding episodic memories: Neocortical inputs are bound together through sparse, non-overlapping hippocampal representations. In contrast, they argued that representations in the neocortex are more distributed and overlapping, with similarity-based generalization to related stimuli through cell and synapse overlap. Experience with a variety of neural network learning algorithms capable of acquiring these sorts of representations suggested that such learning had to be much slower and more incremental, interleaving experience across the wide range of stimuli that needed to be accommodated and interrelated. Without incremental modifications and interleaved presentation, such models suffer from "catastrophic interference" and an overall degradation of the knowledge base (see McClelland et al., 1995; McCloskey & Cohen, 1989, for further discussion). It is worth emphasizing that while these conclusions about the incremental nature of neocortical plasticity are derived from artificial and relatively simple models, the primary limitations in this case derive from the particular combination of highly distributed representations, local activity-dependent synaptic plasticity mechanisms, and the large range and number of stimuli being learned for which similarity relations are

important. These constraints would appear to apply equally well to the solution(s) adopted by the actual brain.

Incremental learning mechanisms have long been championed by proponents of connectionist models (e.g., Rumelhart, McClelland & the PDP Research Group, 1986) and more recently by proponents of "predictive coding" models (which are a form of connectionist model; e.g., Friston, 2005; Friston & Kiebel, 2009), with these mechanisms serving as the basis of accounts of behavior from a wide range of cognitive domains (see Rogers & McClelland, 2014, for a recent overview). Connectionist neural networks are composed of interconnected pools of computational elements referred to as "units," with simple rules of activity integration and propagation amongst the units via weighted connections, also referred to as "synaptic strengths." While such models have typically been abstracted away from the details of the brain, their underlying equations describing activity propagation are isomorphic to more biophysically derived population firing-rate models (e.g., Amit & Tsodyks, 1991; Gerstner, 1998; Wilson & Cowan, 1972; see Ermentrout, 1998, for review), and more recent versions of these models have explicitly included constraints from neuroscience, such as a bias for short-range connections, as well as including more anatomical constraints and constraints on mechanisms of plasticity (e.g., Braver, Barch, & Cohen, 1999; Gotts & Plaut, 2002; Hazy, Frank, & O'Reilly, 2007; Jacobs & Jordan, 1992; Ketz, Morkonda, & O'Reilly, 2013; Norman et al., 2006; O'Reilly, 2006; Plaut, 2002; Plaut & Behrmann, 2011; Usher et al., 1999).

Incremental learning models: supervised versus unsupervised learning mechanisms

Paramount to the current discussion is the distinction between "supervised" and "unsupervised" learning mechanisms. Unsupervised learning rules include the simple Hebbian rule whereby connection strengths are updated by the correlation between pre- and post-synaptic activity states (e.g., ranging from -1 to $+1$): weights are strengthened between connected units that are both active at the same time (e.g., both active at $+1$) and weakened when activities disagree (e.g., units have activities of -1 vs. $+1$). While this rule is simple and bears a strong resemblance to empirical studies of LTP and LTD (see O'Reilly, 2001, for discussion), it is known to be strongly restricted in the range of patterns that it can relate. For example, it will typically fail when mapping correlated input patterns to distinct or uncorrelated output patterns, highlighted by the well known XOR problem (i.e., when either of two individual input units is turned on, the output is on; when the input units are either both on or both off, the output unit is off; e.g., Minsky & Papert, 1969). In contrast, supervised algorithms are capable of learning this and other challenging problems, with learned representations contained in the synaptic strengths that can preserve the similarity relationships among

the patterns, conveying the ability to generalize to related patterns that have not been directly experienced before.

Perhaps the most famous of these supervised rules is the Delta (or Widrow-Hoff) rule with back-propagation (e.g., Rumelhart, Hinton, & Williams, 1986), which works by specifying the patterns to be learned across the inputs and outputs. The actual outputs produced by the model in response to the input is compared to the specified outputs (or "target" values), and weights are adjusted to reduce the error. In a two-layer network model (input and output units only), the resultant learning rule is quite similar to the simpler Hebbian rule, with the term for post-synaptic activity replaced by the difference between target and actual output. The inclusion of at least one "hidden" layer of units (that receives no direct inputs or targets), along with a non-linear activation function, allows such models to learn any arbitrary pattern set provided that the model contains a large enough pool of hidden units. However, this requires the network to "back-propagate" the error one or more steps away from the specified output activity, such that appropriate learning requires the use of non-local information. The availability of specified output activity states to the actual brain in real learning situations has also been questioned (e.g., Mazzoni, Andersen, & Jordan, 1991). Taken together, these issues have led many to conclude that this form of model is not biologically plausible (e.g., Crick, 1989; O'Reilly, 1996, 2001; see Rogers & McClelland, 2014, for a recent discussion).

The Contrastive Hebbian Rule, first proposed in the context of the Boltzmann Machine model (e.g., Ackley, Hinton, & Sejnowski, 1985), eliminates both the non-local computations and explicit target signal elements of the Delta Rule with back-propagation. This model accrues pre- and post-synaptic activity coincidence statistics during the absence of patterned stimulation (using internally generated, random activity during the "negative phase") and compares these locally to the same pre- and post-synaptic coincidences during patterned stimulation (the "positive phase," which could correspond to a simultaneous presentation of sensory and motor states). The weight adjustment is then the difference of two simpler Hebbian terms and can be conceptualized as representing the difference between the existing model and the activity states dictated by the environment (filtered through the model). In this sense it is not directly supervised but "quasi-supervised," and it is capable of learning the same range of input-output patterns as the Delta rule plus back-propagation (see Hinton, 2003; O'Reilly, 1996, for further discussion).

Incremental learning models: the role of task demands on learned representations

In contrast to unsupervised learning rules, supervised and quasi-supervised rules have the property that the task being trained has a large impact on the learned representations

throughout the entire model. For example, take the case where artificial sensory patterns take on values of 0 ("off") or 1 ("on") for each unit and are organized into two "categories," with a set of "on" units that are shared across patterns within a category, along with unique "on" units for each individual pattern. For simplicity, let's stipulate that the two categories of stimuli have no "on" units in common. If these sensory patterns were intended to represent a domain such as a high-level visual form, then patterns from the same artificial category might represent visual objects with similar form. If these input patterns are presented to a three-layer, bi-directionally interacting (i.e., recurrent) connectionist neural network model with either the Delta rule or the Contrastive Hebbian rule, the nature of the weights and hidden-unit patterns that develop will be strongly influenced by the nature of the output patterns that are assigned to the input patterns. If this model is trained in a "categorization" task, such that each of the two categories of input patterns is assigned to a single category pattern over the output units that is the same for all members of the category, then highly similar internal representations (across the hidden units due to the weights) will be learned for stimuli from the same input category during training. The model will first build strong weights from the category – "shared" input units to activate the output "category" pattern – due to their high frequency of experience across patterns (with more opportunity for weight changes), followed by stronger weights from the pattern-unique units.

In contrast, we might train the same model to map each input pattern to a unique output pattern, with little similarity between the output patterns – within or across the input categories. If we stay with the "input as visual form" analogy, this task might be similar to object naming, in which an arbitrary verbal label is assigned to each picture, with visual similarity having little to do with the phonemic relationships among the corresponding names (see Plaut & Shallice, 1993; Rogers & McClelland, 2004, for other examples along these lines). The internal representations that result from training the model in this task will be strongly different from what occurred in the categorization task. The patterns of activity at the hidden layer for each input pattern from the same category will be more distinct and less overlapping relative to the categorization task, with more stimulus-selective tuning. This occurs because the model is forced to use the pattern-unique units almost entirely to activate the correct output units. Interference effects (weight changes that reverse themselves across patterns) and slower learning will also occur as the model first attempts to use the more frequently presented category-shared units to reduce the difference between internally and externally generated activity states (i.e., error).

It follows from the properties of the incremental learning models discussed above that if priming is due to the incremental learning of internal representations in the context of a certain task with certain demands (i.e., the nature of the pattern

mappings being learned, be they systematic or arbitrary, and the sensory and motor modalities involved), then those task demands should also have a measurable impact on priming effects behaviorally. In particular, distinct effects should be observed in categorization tasks versus those observed in tasks requiring the individuation of sensory stimuli (such as in object naming). Specific predictions from any given incremental learning proposal will vary somewhat with the details of the precise model. However, a number of common factors will be expected to matter for any model utilizing a supervised or quasi-supervised learning algorithm and distributed patterns of activity. These models predict that the acquisition of long-term knowledge representations will be influenced by: (1) the nature of the task demands during stimulus processing, (2) the details of the learning algorithm in terms of the performance metric being optimized, (3) the statistics of sensory inputs and motor outputs, and finally, (4) the neuroanatomic constraints on activity propagation within the cerebral cortex (i.e., structural connectivity; see also Mahon & Caramazza, 2011; Mahon, 2015, for review and discussion of this last point). In the context of the current article and following Oppenheim et al. (2010), I will use the term *incremental learning* to refer to distributed connectionist-style models that utilize supervised or quasi-supervised learning rules to gradually minimize the difference between model and data.

While connectionist models of incremental learning are posed in terms of neuron-like and brain-like elements, they are often more targeted towards accounting for behavioral effects and neuropsychological deficits in patients than accounting directly for effects in neural activity. Indeed, the original connectionist model of repetition priming utilizing incremental learning (McClelland & Rumelhart, 1985) was proposed well before the acquisition of brain data from monkeys and humans on the effects of stimulus repetition on neural activity (see Desimone, 1996; Schacter & Buckner, 1998; Wiggs & Martin, 1998, for reviews). As discussed in the next section, the details of neural repetition effects pose a series of challenges to these earlier models.

The puzzle of repetition suppression

While behavioral performance commonly improves with stimulus repetition, neural activity measured in single cells in monkeys and BOLD functional magnetic resonance imaging (fMRI) in humans commonly decreases, a phenomenon referred to as "repetition suppression" (see Grill-Spector, Henson, & Martin, 2006; Gotts et al., 2012a, for reviews). Repetition suppression shares many of the empirical characteristics of repetition priming: it is relatively automatic, even occurring in animals under anesthesia (e.g., Miller, Gochin, &

Gross, 1991); it can be long-lasting following a small number of stimulus exposures (up to several days, e.g., van Turennout et al., 2000, 2003), although effects are strongest with little or no delay (e.g., Grill-Spector & Malach, 2001; Jiang et al., 2000; van Turennout et al., 2003); it is robust to small changes in stimulus form and/or task, with the largest effects when stimuli and tasks are identical across repetition (e.g., Horner & Henson, 2008; Koutstaal et al., 2001; Lueschow, Miller, & Desimone, 1994). Repetition suppression in humans is typically observed throughout the neocortex within regions that are engaged by the task being performed (see Schacter & Buckner, 1998; Wiggs & Martin, 1998, for reviews). Attempts to relate repetition suppression and priming phenomena quantitatively in humans using the magnitude of the BOLD signal change in fMRI and the magnitude of response time change have met with mixed success (e.g., Dobbins et al., 2004; Lustig & Buckner, 2004; Maccotta & Buckner, 2004; Sayres & Grill-Spector, 2006; Turk-Browne et al., 2006; see also McMahon & Olson, 2007). Nevertheless, the common joint observation of these phenomena has led to a variety of theoretical proposals to link them mechanistically (see Gotts et al., 2012a, for review).

Foremost among the questions raised by the observation of neural repetition suppression is: how can reduced neural activity lead to improved behavioral performance? Somehow, reduced levels of neural activity are being relayed more efficiently throughout brain regions engaged in task performance, producing better performance at a lower metabolic cost. But where does this improved efficiency come from? Most activity-based models derived from cognitive psychology, such as spreading activation (e.g., Anderson, 1983; Collins & Loftus, 1975) and most of the connectionist neural networks discussed above (e.g., McClelland & Rumelhart, 1985) have hypothesized greater activity in stimulus-selective nodes or units, along with decreased activity in units that prefer different stimuli. In many neuroscientific contexts, this is precisely what is observed at the neural level under conditions of facilitated behavioral performance, such as elevated firing rate activity during visual selective attention in cells that prefer the attended object (e.g., Luck et al., 1997) or elevated firing rate in motion-selective cells prior to a motion discrimination with the preferred motion direction (e.g., Newsome, Britten, & Movshon, 1989). However, in the case of stimulus repetition, firing rate activity in inferior temporal cortex and prefrontal cortex is reduced – often with the greatest decreases observed in the cells that respond best to the stimulus (e.g., Li, Miller, & Desimone, 1993; McMahon & Olson, 2007; Miller, Li, & Desimone, 1993). This observation of proportional "scaling" of firing rate responses has its analog in fMRI studies of repetition-related activity changes in humans, as well (e.g., Weiner et al., 2010). The details of neural repetition suppression therefore appear to pose a strong challenge to existing neural network incremental learning models.

Models that jointly address repetition suppression and priming

Facilitation model

In response, a variety of mechanisms have been proposed to link repetition suppression and repetition priming, some of which involve fundamental alterations to neural representations and neural tuning and others that do not. For example, the "facilitation" model (James et al., 2000; Henson, 2003) holds that neural responses are simply advanced in time and terminate earlier with repetition, thereby simultaneously explaining faster response times (priming) and repetition suppression as long as the earlier termination of activity is sufficiently fast to result in overall reductions in activity. This is not so much an articulation of detailed plasticity mechanisms as it is a description of activity states that could resolve the puzzle of repetition suppression. While the BOLD response in fMRI is too slow to directly examine this model, single-cell recording studies in monkeys with novel and familiar stimuli, as well as with stimulus repetition during the experiment, have notably failed to yield evidence in support of this theory (cf. Woloszyn & Sheinberg, 2012). Firing rate responses to repeated/familiar stimuli are typically reduced throughout stimulus processing in the inferior temporal and frontal cortex and fail to be advanced ahead of responses to novel stimuli (e.g., Anderson et al., 2008; Freedman et al., 2006; Li et al., 1993; McMahan & Olson, 2007; Pedreira et al., 2010; Rainer & Miller, 2000; Verhoef et al., 2008). Note that the facilitation model also has little explanation for how neural representations are formed in the first place.

Synchrony model

Another account that makes little if any prediction about fundamental changes in the content of neural representations and neural tuning preferences with repetition is the "synchrony" model (Gotts, 2003; Gilbert et al., 2010; Ghuman et al., 2008; see Gotts et al., 2012a, for discussion). On this view, cells fire at rates that are reduced overall with repetition but are more synchronized in their spike times, permitting better propagation of individual spikes throughout the entire processing pathway in both the feed-forward and feedback directions, facilitating earlier responses. There is supporting evidence for this model, both in single-cell recording studies in animals (e.g., Anderson et al., 2008; Brunet et al., 2014; Kaliukhovich & Vogels, 2012; Hansen & Dragoi, 2011; von Stein, Chiang, & Konig, 2000; Wang et al., 2011) and in humans using MEG (magnetoencephalography; Gilbert et al., 2010; Ghuman et al., 2008) and intracranial EEG (electroencephalography; Engell & McCarthy, 2014). Like the facilitation model, though, the synchrony model requires additional mechanisms to explain the basic formation of representations. Rather, it serves more as a "gain" modulator within existing networks

that permits enhanced neural processing efficiency (see Fries et al., 2001; Salinas & Sejnowski, 2001, for similar proposals). There is the possibility that neural plasticity mechanisms such as spike-timing-dependent LTP and LTD (e.g., Bi & Poo, 1998; Markram, Lubke, Frotscher, & Sakmann, 1997; Sjöström, Turrigiano, & Nelson, 2001) underlie longer-term changes in synchrony, but these relationships have yet to be specified in enough detail to make any clear predictions about neural tuning changes that could be tested in experiments.

Perceptual "sharpening" model

There are two prominent accounts of repetition suppression and priming that are posed in terms of fundamental changes in neural tuning that are directly relevant for incremental learning proposals, namely the perceptual "sharpening" and "predictive coding" models. The perceptual "sharpening" theory (Desimone, 1996; Wiggs & Martin, 1998) holds that while neural activity is decreasing overall with stimulus repetition, firing rates are becoming more selectively tuned, with the largest decreases occurring in cells that are poorly responsive and/or weakly tuned to the repeated stimuli. In contrast, cells that are most responsive and selectively tuned to the repeated stimuli may maintain their firing rate levels or decrease much less than the poorly tuned cells. With improved stimulus selectivity, bottom-up support would be removed for alternative or competing representations in downstream brain regions, allowing more rapid propagation of stimulus-selective activity throughout task-engaged neural pathways, as well as faster and more accurate behavioral responses. In the case of stimulus repetition over seconds or minutes, the evidence to date has failed to support sharpened tuning, with support instead for proportional scaling of neural responses (e.g., Li et al., 1993; McMahan & Olson, 2007; Miller et al., 1993; Weiner et al., 2010; see also De Baene & Vogels, 2010). However, over much longer durations of practice with a stimulus set over days and weeks, a number of studies have found evidence consistent with sharper tuning in the inferior temporal cortex and lateral prefrontal cortex in monkeys (e.g., Baker et al., 2002; Freedman et al., 2006; Rainer & Miller, 2000) and in lateral occipital cortex in humans using fMRI-adaptation¹ (e.g., Gillebert et al., 2009; Jiang et al., 2007). Initially conceived as occurring in more perceptual brain regions, such as extrastriate and ventral temporal cortex in vision (e.g.,

¹ "fMRI Adaptation" employs short-term stimulus repetitions in fMRI to rapidly and temporarily induce local decreases in BOLD activity thought to be specific to the cells and synapses just engaged in processing (e.g., Grill-Spector & Malach, 2001; Naccache & Dehaene, 2001). Neural tuning along a particular dimension can then be probed by presenting a new stimulus varying in that dimension from the adapting stimulus and measuring the recovered response.

Desimone, 1996; Wiggs & Martin, 1998), the basic notion of representational sharpening has since been extended more generally throughout the brain to other domains of knowledge representation, including frontal regions involved in higher level cognitive decisions (see Grill-Spector et al., 2006; Norman & O'Reilly, 2003, for discussion). However, the observation of proportional scaling – as opposed to sharpening – at delays typical of most repetition priming and repetition suppression effects calls into question how well-linked sharpening is to priming (discussed in Gotts et al., 2012a, b). Perceptual sharpening by itself also has no direct way to explain the development of non-perceptual long-term knowledge representations such as conceptual knowledge. As stated, this model should prune away information that is shared across the representations of related items, making individual stimuli more discriminable but undercutting generalization of knowledge to similar items.

Predictive coding model

The "predictive coding" model (Friston, 2005; Friston & Kiebel, 2009; Henson, 2003) casts the cortex as a form of hierarchical generative Bayesian statistical model. Perceptual inference occurs as a progressive interaction between bottom-up sensory input ("evidence") and top-down expectations ("prediction") throughout the cortical hierarchy. Top-down predictions serve to inhibit or suppress the bottom-up sensory evidence, with residual activity in the lower levels of the cortical hierarchy serving as "prediction error," which is, in turn, relayed back toward the higher levels. Top-down model predictions are improved with stimulus repetition through the application of its learning algorithm, leading to stronger suppression of cells encoding prediction error in earlier sensory regions (i.e., repetition suppression), but also to the speeding up (i.e., facilitation) of evoked neural responses via an increase in synaptic gain due to enhanced encoding precision and confidence, ultimately reflected as behavioral priming. This can produce a situation similar to that expected by the sharpening model, with a boosting of certain prediction errors regarding the best hypotheses about the cause of sensory input while suppressing poorer alternative hypotheses (see Friston, 2012, for discussion). The predictive coding model has received some preliminary experimental support from effective connectivity studies of repetition suppression in fMRI. For example, Ewbank et al. (2011) used Dynamic Causal Modeling (DCM; e.g., Friston, Harrison, & Penny, 2003) to examine causal interactions between the fusiform body area (FBA) and the extrastriate body area (EBA) while subjects viewed pictures of human bodies. They observed repetition suppression to repeated bodies of the same identity in both EBA and FBA, along with increased top-down causal interactions from FBA to EBA. While the quantitative link between top-down coupling and the magnitude of repetition

suppression in earlier sensory regions has yet to be documented, the finding of increased top-down coupling simultaneously with repetition suppression is nevertheless consistent with a core prediction of this model.

The predictive coding model is a particular form of incremental learning proposal in which changes to model parameters minimize the difference between model and observed data. Unlike the facilitation, synchrony, and sharpening models, it specifically addresses the initial formation and ongoing modification of perceptual and conceptual representations. Due to its learning algorithm (the expectation-maximization, or E-M algorithm), it shares a key experimental prediction with earlier supervised neural network models of priming: representational changes should be qualitatively and quantitatively modulated by alterations in task demands and the low- and high-level statistics of sensorimotor relationships. Furthermore, all of these models (including the McClelland & Rumelhart, 1985, model) when trained to discriminate or individuate stimuli will utilize some form of representational sharpening (as in the object-naming example in the section *Incremental learning models: the role of task demands on learned representations* above). This appears to stand in contrast with the experimental data for repetitions separated by delays of seconds and several minutes that have documented a pattern of activity change more consistent with proportional scaling rather than sharpening (e.g., Li et al., 1993; McMahon & Olson, 2007; Miller et al., 1993; Weiner et al., 2010). However, the relationships among repetition suppression, priming, and the changes in underlying knowledge representations (i.e., changes in neural tuning and selectivity) over longer delays that are clearly relevant for longer-term plasticity mechanisms (e.g., 30–60 min or longer) have been less examined. In the next section, the results of studies using fMRI-adaptation in humans to study repetition-related changes in neural tuning with an eye toward the impact of task demands on tuning changes are discussed.

Task-dependence of changes to neural representations that accompany stimulus repetition

Artificial category learning

One neuroscientific topic in the literature with obvious relevance to the current discussion is category learning, with studies that manipulate stimulus features that are either relevant or irrelevant to appropriate categorization. A number of these studies have examined training-related changes in neural tuning, including single-cell recordings in monkeys (e.g., De Baene et al., 2008; Freedman et al., 2001; Op de Beeck et al., 2001; Sigala & Logothetis, 2002) and fMRI-adaptation in humans (e.g., Gillebert, et al., 2009; Jiang et al., 2007; van der Linden et al., 2010). While not all of these studies have

been directly focused on repetition suppression and few if any have measured behavioral priming, they provide a relatively direct window into the effects of stimulus repetition and task relevance on changes in neural representations. For example, Freedman and colleagues (Freedman et al., 2001, 2002, 2003, 2006) trained monkeys over a period of many days to classify morphed pictures of cats and dogs into two or more arbitrary categories. After training, single neurons in the lateral prefrontal cortex were strongly tuned to the trained category and less sensitive to stimulus form. When monkeys were re-trained with new category boundaries using the same stimuli, cells in the prefrontal cortex became sensitive to the new boundaries and lost sensitivity to the old boundaries. In contrast, cells in the inferior temporal cortex (monkey area TE) were tuned more to stimulus form with weaker sensitivity to category. These cells further showed evidence of training-related increases in selectivity to stimulus form (i.e., perceptual "sharpening") when compared to a novel stimulus set (Freedman et al., 2006).

Jiang et al. (2007) have obtained highly compatible results with those of Freedman et al. when using fMRI-adaptation in humans. Subjects were trained to classify morphed pictures of cars into two arbitrary categories. After approximately a week of training, greater release from adaptation was observed to small differences in stimulus form irrespective of category in right lateral occipital cortex (LOC), consistent with perceptual sharpening. Strong sensitivity to trained category was instead observed in right lateral prefrontal cortex, with less sensitivity to stimulus form. Category tuning in the prefrontal cortex was also significantly associated with behavioral measures of categorization performance. Like Jiang et al. (2007), two subsequent studies of visual category learning in humans using fMRI-adaptation found increased stimulus selectivity in occipitotemporal regions, consistent with perceptual sharpening, but without additional evidence of sensitivity to category that would have aided the explicit task demands (e.g., Gillebert et al., 2009; van der Linden et al., 2010).

In contrast to these studies, two more recent category learning experiments in humans have found evidence of tuning changes to task-relevant dimensions in the occipitotemporal cortex using fMRI-adaptation (Folstein et al., 2013; van der Linden et al., 2014), as well as strong tuning to category in the prefrontal cortex (van der Linden et al., 2014). The basis of the contrasting results is not entirely clear, but it is possible that the composite nature of the form changes in earlier studies made it more difficult for participants to attend to single relevant form dimensions which could have aided earlier occipitotemporal plasticity. Task-relevant alterations may also be more visible during the performance of the categorization training task itself rather than unrelated tasks that are often used during fMRI-adaptation (e.g., van der Linden et al., 2014; see also McKee et al., 2014). Regardless, it is clear from all of the studies with coverage of the frontal lobes that task-

relevant information is represented most strongly in prefrontal cortex after category training, with some studies showing smaller task-relevant alterations in tuning in occipitotemporal cortex as well. These observations are in reasonably good accord with expectations of incremental learning models. The role of attention to relevant stimulus dimensions, which may be critical for the observation of category-sensitivity in occipitotemporal regions in the category training studies just discussed, might be implemented in these models through elevated activity in the cells encoding the relevant sensory dimensions, leading to enhanced plasticity with respect to those dimensions (e.g., Cohen, Dunbar, & McClelland, 1990).

A connectionist account of hemispheric specialization and category preferences in occipitotemporal regions for words, faces, and houses

In the artificial category-learning studies just discussed, recent evidence has demonstrated support for the training-related emergence of category distinctions in occipitotemporal regions. Along these same lines, Plaut and Behrmann (2011) have put forward a connectionist account of occipitotemporal category preferences and hemispheric specialization for more naturally occurring categories of words, faces, and houses. The model supplements incremental learning mechanisms with a bias for learning short-range connections and hemispheric constraints on anatomical connections. Simple retinotopic maps of units from each hemisphere representing early visual cortex send input to intermediate pools of units intended to represent activity in the fusiform gyrus of each hemisphere, which in turn send input to pools of units representing lexical processing ("language units") and the "identity" of each object. Critically, the language units only receive inputs from the left-hemisphere intermediate units, whereas the identity units receive inputs from both hemispheres. The model was trained to identify individual instances of faces, words, and houses, with some variation in the overall size of the stimuli in retinotopic coordinates. After training, the model learned to use more "foveal" units for representing words and faces, since the critical information for discriminating among these stimuli was conveyed through the more foveal parts of the input map. Similarly, the model used more "peripheral" portions of the intermediate units to represent the house stimuli. These aspects of the model are reminiscent of the neuroimaging results of Levy, Hasson, Malach and colleagues demonstrating an object category by retinotopic position bias in occipitotemporal cortex (e.g., Hasson et al., 2002; Levy et al., 2001). No specific division or anatomic constraint within the pools of intermediate units was required for this foveal/peripheral bias to occur. It resulted solely from the bias for short-range connections in learning and the intrinsic retinotopic biases present in the stimuli themselves. The model also produced a left-hemisphere bias for the

word stimuli and a more right-hemisphere bias for the face and house stimuli. This occurred for similar reasons in the model as for the foveal/peripheral preferences: only the left hemisphere intermediate units provided input to the language units, with the bias for short connections leading to the left hemisphere word preferences. The right hemisphere bias for the face and house stimuli resulted at least in part from interference and competition for representation between the word and face/house stimuli. While this model is highly simplified, it provides one concrete instantiation of how incremental learning mechanisms could lead to reliable conceptual category specificity without a neural architecture that is dedicated *a priori* to the representation of certain conceptual content.

Priming, repetition suppression, and tuning changes in object naming

Results from the artificial category learning studies above appear to provide support for the influence of task demands on learned representations, consistent with incremental learning models. However, training durations in these category learning studies are quite a lot longer than those used in most repetition priming and repetition suppression studies, with effects of learning being measured after hundreds or thousands of trials of highly similar visual stimuli. Repetition priming was also not assessed in any of the above studies, making it difficult to assess the inter-relationships of priming, repetition suppression, and changes in tuning, task-relevant or otherwise. A recent study in our laboratory has directly examined these inter-relationships using fMRI adaptation (Gotts, Milleville, & Martin, 2014). Rather than training subjects to categorize novel objects, we had subjects repeatedly name a large set of well-known objects (e.g., dog, hammer, etc.) as is commonly done in studies of repetition priming and repetition suppression (e.g., Cave, 1997; van Turennout et al., 2000, 2003). The pressure in this task is quite different to categorization tasks in that the focus is not on the features that are common to objects in the same category but on the distinctive features that are selectively associated with each object's identity and name. As discussed above in the section *Incremental learning models: the role of task demands on learned representations*, the task demands of picture naming should promote perceptual sharpening, attenuating shared object features that support related but incorrect object names. If perceptually sharper tuning underlies both repetition suppression in occipitotemporal regions and behavioral priming during object naming, then we should be able to observe these phenomena simultaneously and relate them quantitatively across subjects.

Subjects in our study were asked to name a set of 50 objects five times in a pseudorandom order prior to fMRI. As expected, they named objects faster and more accurately across the five repetitions, with robust effects of repetition priming

observed at the end of the experiment. Thirty minutes after the initial naming practice, the same subjects participated in an fMRI-adaptation task. They viewed adaptation sequences composed of rapidly repeated objects (3–6 repetitions over several seconds) that were either named previously or that were new for the fMRI session, followed by single "deviant" object pictures used to measure recovery from adaptation and that shared a relationship to the adapted picture (a different exemplar of the same object, a conceptual associate, or an unrelated picture).² As expected, effects of short-term adaptation and recovery were found throughout visually responsive brain regions, and occipitotemporal cortical regions displayed repetition suppression to previously named relative to new adaptors. However, these same occipitotemporal regions failed to exhibit pronounced changes in neural tuning, similar to other studies examining effects of within-session repetition on tuning (e.g., De Baene & Vogels, 2010; Li et al., 1993; McMahon & Olson, 2007; Weiner et al., 2010). Changes in neural tuning were indeed observed during fMRI in the left lateral prefrontal cortex, although they were in the opposite direction of what was predicted by the task demands. Greater residual adaptation was observed to different exemplars and conceptual associates following previously named adapting stimuli, consistent with *broader* rather than *sharper* tuning to conceptually related objects. Furthermore, this change in neural tuning was directly related to the proportion of conceptual errors made by subjects in the picture naming sessions both before and after fMRI.

In a follow-up behavioral experiment in a separate group of subjects, we found that the same pre-exposure task led to greater semantic priming (e.g., Dell' Acqua & Grainger, 1999; Meyer, Schvaneveldt, & Ruddy, 1975) when using the previously named pictures as briefly presented primes, with the magnitude of later repetition priming predicting the magnitude of this increase in semantic priming across subjects. In other words, when viewing the picture of a "cow," the representations of conceptually related concepts such as "horse" were more active if cow had been recently named – despite the fact that the task is requiring individuation of "cow" and "horse" in order for the correct naming response to be given. It would be one thing if these results were an aberration, but increases in conceptual relatedness and/or interference following object identification have been observed in a variety of other paradigms, such as blocked cycling picture naming, in which the task demands are quite similar to those in our study (e.g., Belke et al., 2005; Damian & Als, 2005; Damian et al., 2001; Hodgson et al., 2003; Howard et al., 2006; Hsiao et al., 2009; Maess et al., 2002; Schnur et al., 2006; Schnur et al.,

² The subjects' overt task during fMRI was to push a button when they saw a man-made object, which were intermixed between the adaptation sequences (all animal stimuli). Subjects were instructed to attend to all images.

2009; see also Navarrete, Mahon, & Caramazza, 2010; Schnur, 2014). Increased incidence of conceptual naming errors, including perseverations on previous naming responses, is also quite common following repeated picture naming (e.g., Gotts et al., 2002; Hsiao et al., 2009; Santo Pietro & Rigrodsky, 1982; Vitkovitch & Humphreys, 1991; Vitkovitch et al., 1993).

Challenges to incremental learning posed by Gotts et al. (2014)

Taken together, the results from our study appear to present two basic challenges to incremental learning models. The first is that after a relatively small number of stimulus repetitions, activity decreases occur in the occipitotemporal cortex without much concomitant change in neural tuning. This undermines the unitary explanation of priming, repetition suppression, and neural tuning provided by the perceptual sharpening and predictive coding models (at least as previously discussed in Friston, 2012). These results are in better accord with models of efficiency without obvious tuning changes, such as the synchrony model (see Gotts et al., 2012a, b, for discussion) or possibly the facilitation model, although the facilitation model has yet to receive clear empirical support under typical viewing conditions (identifying isolated objects) from studies that have measured brain activity with high spatial and temporal resolution. This is not to say that effects of repetition suppression are unrelated to changes in tuning over longer durations of practice. However, over the durations of practice that are typical of most repetition priming studies (2–5 stimulus repetitions within a single experimental session), strong increases in stimulus-selective tuning have not been observed. Instead, these results highlight the possibility that strong, long-lasting decreases in stimulus-evoked activity (i.e. repetition suppression) can take place, without much qualitative change in the underlying neural representations. Most incremental learning models focus solely on optimizing task performance rather than on the metabolic cost of neural information processing (see Aiello & Wheeler, 1995; Raichle & Mintun, 2006, for reviews), and incorporating this dimension may be critical in accounting for phenomena such as these.

The second basic challenge to incremental learning models is that the changes in tuning that *were* observed in our study in the left lateral prefrontal cortex (conceptual broadening), along with similar behavioral results from other studies (e.g., increased semantic errors in speeded picture naming), were counter to those expected by the task demands in picture naming. Even in the event that representational sharpening was occurring in occipitotemporal regions in this study but our methods were insufficiently sensitive to detect it, the evidence for greater activation of related object representations in the lateral prefrontal cortex directly contradicts the expected outcome of such sharpening. In other words, regions in the

prefrontal cortex that are presumably downstream of occipitotemporal regions exhibited reduced rather than enhanced stimulus-selectivity, undermining this popular joint explanation of repetition suppression and priming. These effects cannot be easily dismissed as artifactual in some way, as they were manifest behaviorally both as increased incidence of conceptual naming errors and as increased semantic priming magnitude (see also Gronau, Neta, & Bar, 2008, for an earlier demonstration of the same conceptual broadening effect in fMRI Adaptation).

Putting incremental learning to the test

Thus far, we have considered incremental learning models as accounts of how perceptual and conceptual representations form, as well as how the phenomena of repetition priming and repetition suppression may relate to incremental learning mechanisms. Core to incremental learning is the gradual optimization of task performance (minimizing the difference between model and data), which predicts that neural representations should be qualitatively shaped by task demands, as well as by anatomical connectivity and the statistics of sensorimotor contingencies. Consistent with these models, category training studies have yielded evidence that manipulating feature relevance has an impact on learned representations in prefrontal cortex, with cells showing strong preferences for the trained category boundaries. While early studies failed to find much evidence of category tuning in accord with task demands in occipitotemporal regions (e.g., Gillebert et al., 2009; Jiang et al., 2007; Op de Beeck et al., 2001), more recent evidence suggests that these kinds of effects can be observed – albeit at smaller magnitudes than observed in prefrontal cortex (e.g., De Baene et al., 2008; Folstein et al., 2013; van der Linden et al., 2014). Additionally, a number of studies have now documented increases in stimulus selectivity in ventral temporal cortex after extensive training, both in monkeys and in humans (e.g., Baker et al., 2002; Freedman et al., 2006; Jiang et al., 2007; Gillebert et al., 2009; see also Rainer & Miller, 2000).

The more challenging effects for incremental learning models to explain are ones that happen after a smaller number of within-session repetitions (2–5), which are more typical of repetition priming studies. This is not a small issue for the interpretation of repetition priming and repetition suppression effects over these durations given that these are the intervals over which such effects are largest in magnitude, gradually decaying or suffering from interference over longer intervals (e.g., Li et al., 1993; McKone, 1998; van Turennout et al., 2000, 2003). One might also expect that these contexts would have more ecological validity, since they are more in line with what people typically encounter when interacting with objects intermittently in the environment or with their corresponding names in print. As mentioned in the previous section, these

challenging effects are twofold: (1) there is a notable lack of evidence of perceptual sharpening over a small number of repetitions (e.g., De Baene & Vogels, 2010; Gotts et al., 2014; Li et al., 1993; McMahon & Olson, 2007; Miller et al., 1993; Weiner et al., 2010), and (2) there is evidence of an expansion of conceptual representations in picture naming in contrast to the apparent task demands for perceptual sharpening (e.g., Gotts et al., 2014; see Gronau, Neta & Bar, 2008, for related results). Incremental learning models might respond to the first challenge by incorporating effects such as spike synchrony (e.g., Friston, 2012), retaining their basic plasticity mechanisms that apply to average firing rate (see Gotts, 2003; Gotts et al., 2012a, for discussion), but how might the second challenge be addressed? Oppenheim et al. (2010) have put forward one suggestion for the data from blocked cyclic picture naming. They trained a connectionist model using the Delta rule to map an array of semantic features representing object meaning to corresponding object names (lexical units). Each time an object was presented to the model, the weights between the semantic features and the appropriate lexical unit were strengthened via the learning rule, and weights between those features and incorrect lexical units of semantic associates that were partially activated via shared semantic features were weakened. Over trials, lexical units were activated more rapidly and accurately (repetition priming), but this occurred faster when blocks were composed of unrelated items, since each object suffered less from weakened semantic- > lexical weights involving shared semantic features with other objects in the set (along the lines discussed for the object naming example in the section *Incremental learning models: the role of task demands on learned representations*). While this model accounts quite nicely for a variety of effects in blocked cyclic naming (e.g., Belke, 2008; Belke, 2008; Damian & Als, 2005; Howard et al., 2006; Hsiao et al., 2009; Navarrete et al., 2014; Schnur et al., 2006), it is less clear how it would address the broadened conceptual representations and enhanced semantic priming observed in Gotts et al. (2014). This model should decrease relatedness of conceptual associates by attenuating the impact of shared semantic features on lexical processing over learning. However, one can imagine how adding aspects of visual processing might change this picture. It is possible that early in learning, weight changes between visual and conceptual processing would strengthen conceptual relationships among associates to reduce error at the lexical level, producing increased conceptual similarity and semantic priming (see early training epochs of Simulations 3 and 5 in Rogers & McClelland, 2004, for related examples). However, it seems clear that over the course of additional training, this should eventually give way to the model utilizing distinctive features at the expense of shared features (as in the current Oppenheim et al. model), eliminating the activation of inappropriate related names and ultimately decreasing conceptual relatedness

effects (e.g., see later epochs in Simulations 3 and 5 in Rogers & McClelland, 2004). Perhaps this would also have been seen experimentally in the task used by Gotts et al. (2014) if greater than five repetitions had been used prior to fMRI. A related possibility is that the tuning changes accompanying picture naming are interacting with the existing structure of perceptual and conceptual representations, with changes that – while conceptual – are less strong than they would have been under a task such as categorization. In other words, it is possible that the task demands have actually shifted the changes more toward perceptual sharpening than they would have otherwise been under a different task. Under this view, the current results would already be consistent with incremental learning models with no additional modification.

Future experiments

What is needed in the future to examine these issues are experiments that manipulate the nature of the task being performed on repeated objects while measuring repetition suppression and behavioral priming, contrasting a task such as naming with various levels of conceptual categorization, or alternatively with fine visual discriminations on the pictured objects (e.g., Martin & Gotts, 2005; Wig, 2012). It would also be important to have a much more continuous manipulation of repetition, ranging from a small number of repetitions up through hundreds or even thousands as has been done in category learning studies. An important variable with regard to the nature of plasticity and representational changes that is typically conflated with number of repetitions is time. Separating time from repetition number will be critical for examining these relationships. Counterbalancing item sets with matched properties (category membership, frequency, familiarity, etc.) will also be critical, perhaps assigning different object sets to different tasks and comparing the relative changes.

Unsupervised alternatives to incremental learning

While current evidence suggests that incremental learning models show promise for addressing the acquisition of perceptual and conceptual knowledge structure in the brain, other possibilities remain that posit more dedicated functions to occipital and temporal cortex, with unsupervised learning applying throughout the visual system in a more feed-forward manner that relies on biases in low-level image statistics between stimulus categories to explain any category preferences in occipitotemporal cortex (e.g., Riesenhuber & Poggio, 2002). On this view, most of the flexibility and adaptability of knowledge representation is handled by the prefrontal cortex (see Jiang et al., 2007; Scholl et al., 2014, for discussion). Indeed, it is interesting to note that conceptual "broadening" is observed when the task is categorization (e.g., Jiang et al., 2007), as well as when the task requires individuation (e.g., Gotts

et al., 2014; see also Gronau et al., 2008), at least raising the possibility that prefrontal cortex has a more general role in automatically extracting more abstract category-like structure from individual objects presented in the same context. At a minimum, the future experiments just outlined will help to delineate the bounds on existing incremental learning proposals, allowing us to see how much of the adult knowledge organization is plausibly based in supervised versus unsupervised learning mechanisms, as well as to what extent innate anatomical constraints might contribute.

What kind of models are needed?

In the current paper, I have discussed the capabilities of existing incremental learning models to address data on the acquisition of perceptual and conceptual knowledge and the relationship between priming and repetition suppression. While many connectionist models have been proposed to address behavioral data and neuropsychological deficits in this domain, the most glaring omission in these models is the lack of detailed accounts of the related brain physiology in terms of timing and magnitude of activity in particular brain regions. To be sure, the integration of these disparate forms of data within individual models is daunting. There are two main limiting factors that, if addressed, may help with these goals. The first is that for humans, unlike monkeys, we have less detailed knowledge of delineated brain regions and their circuit-level interrelationships. This makes it quite difficult to incorporate such constraints into models. Recent studies of functional connectivity in human fMRI have made a great deal of progress parcellating the brain into large-scale functional circuits (e.g., Power et al., 2011; Yeo et al., 2011), and this should eventually help to constrain model architectures. The second main limitation is that it is unclear how one should relate unit activity and other parameters in connectionist models directly to neural measures. However, links already exist between firing-rate-based and spiking neural networks (e.g., Amit & Tsodyks, 1991; Gerstner, 1998; Wilson & Cowan, 1972), and these offer the promise of more detailed relationships and testable predictions in single-cell recording studies in monkeys and neuroimaging studies in humans (see Gotts, 2003; Gilbert et al., 2010, for one example along these lines). I view current proposals such as the Plaut and Behrmann (2011) model as steps in the right direction with regard to these goals. With improved ability to make contact between models and neural data directly, interactive theory building and direct testing of hypotheses about the neural mechanisms underlying learning and the formation of knowledge representations may be possible.

Acknowledgments The author would like to thank Alex Martin, Jay McClelland, David Plaut, Carson Chow, Gary Dell, Gary Oppenheim, and Sharon Thompson-Schill for helpful discussions, and Brad Mahon,

Ken Norman, and an anonymous reviewer for insightful comments on the manuscript. The writing of this paper was supported by the National Institute of Mental Health, NIH, Division of Intramural Research.

References

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147–169.
- Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, 22, 425–489.
- Aiello, L. C., & Wheeler, P. (1995). The expensive tissue hypothesis: The brain and digestive system in human and primate evolution. *Current Anthropology*, 36, 199–221.
- Amit, D. J., & Tsodyks, M. V. (1991). Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate - spikes, rates and neuronal gain. *Network*, 2, 259–273.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, B., Mruczek, R. E., Kawasaki, K., & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral Cortex*, 18, 2540–2552.
- Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5, 1210–1216.
- Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 23, 1059–1082.
- Belke, E. (2008). Effects of working memory load on lexical-semantic encoding in language production. *Psychonomic Bulletin & Review*, 15, 357–363.
- Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology*, 58, 667–692.
- Bi, G. Q., & Poo, M. M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18, 10464–10472.
- Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object recognition. *Perception*, 20, 585–593.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 121–133.
- Bliss, T. V., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology*, 232, 331–356.
- Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biological Psychiatry*, 46, 312–328.
- Brunet, N.M., Bosman, C.A., Vinck, M., Roberts, M., Oostenveld, R., Desimone, R. ... & Fries, P. (2014). Stimulus repetition modulates gamma-band synchronization in primate visual cortex. *Proceedings of the National Academy of Sciences - USA*, 111, 3626–3631.
- Burgess, N., Maguire, E. A., & O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35, 625–641.
- Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, 4, e125. doi:10.1371/journal.pbio.0040125

- Cantlon, J. F., Pineda, P., Dehaene, S., & Pelphrey, K. A. (2011). Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cerebral Cortex*, 21, 191–199.
- Cave, C. B. (1997). Very long-lasting priming in picture naming. *Psychological Science*, 8, 322–325.
- Cave, C. B., & Squire, L. R. (1992). Intact and long-lasting repetition priming in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 509–520.
- Cave, C. B., Bost, P. R., & Cobb, R. E. (1996). Effects of color and pattern on implicit and explicit picture memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 639–653.
- Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, 2, 913–919.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science*, 210, 207–210.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332–361.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–428.
- Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H.M. *Seminars in Neurology*, 4, 249–259.
- Crick, F. (1989). The recent excitement about neural networks. *Nature*, 337, 129–132.
- Damian, M. F., & Als, L. C. (2005). Long-lasting semantic context effects in the spoken production of object names. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 31, 1372–1384.
- Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, 81, B77–B86.
- De Baene, W., & Vogels, R. (2010). Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. *Cerebral Cortex*, 20, 2145–2165.
- De Baene, W., Ons, B., Wagemans, J., & Vogels, R. (2008). Effects of category learning on the stimulus selectivity of macaque inferior temporal neurons. *Learning & Memory*, 15, 717–727.
- Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, 73, B1–B15.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, USA*, 93, 13494–13499.
- Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature*, 428, 316–319.
- Engell, A. D., & McCarthy, G. (2014). Repetition suppression of face-selective evoked and induced EEG recorded from the human cortex. *Human Brain Mapping*, 35, 4155–4162.
- Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Science*, 12, 388–396.
- Ermentrout, G. B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, 61, 353–430.
- Ewbank, M. P., Lawson, R. P., Henson, R. N., Rowe, J. B., Passamonti, L., & Calder, A. J. (2011). Changes in "top-down" connectivity underlie repetition suppression in the ventral visual pathway. *Journal of Neuroscience*, 31, 5635–5642.
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120, 339–357.
- Folstein, J. R., Palmieri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex*, 23, 814–823.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291, 312–316.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2002). Visual categorization and the primate prefrontal cortex: Neurophysiology and behavior. *Journal of Neurophysiology*, 88, 929–941.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, 23, 5235–5246.
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral Cortex*, 16, 1631–1644.
- Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, 291, 1560–1563.
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, 360, 815–836.
- Friston, K. J. (2012). Predictive coding, precision and synchrony. *Cognitive Neuroscience*, 3, 238–239.
- Friston, K. J., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society London B: Biological Sciences*, 364, 1211–1221.
- Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modeling. *Neuroimage*, 19, 1273–1302.
- Gerstner, W. (1998). Time structure of the activity in neural network models. *Physical Review E*, 51, 738–758.
- Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Science, USA*, 105, 8405–8409.
- Gilbert, J. R., Gotts, S. J., Carver, F. W., & Martin, A. (2010). Object repetition leads to local increases in the temporal coordination of neural responses. *Frontiers in Human Neuroscience*, 4, 30. doi:10.3389/fnhum.2010.00030
- Gillebert, C. R., Op de Beeck, H. P., Panis, S., & Wagemans, J. (2009). Subordinate categorization enhances the neural selectivity in human object-selective cortex for fine shape differences. *Journal of Cognitive Neuroscience*, 21, 1054–1064.
- Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, 10, 512–522.
- Golarai, G., Liberman, A., Yoon, J. M., & Grill-Spector, K. (2010). Differential development of the ventral visual cortex extends through adolescence. *Frontiers in Human Neuroscience*, 3, 80. doi:10.3389/neuro.09.080.2009
- Gotts, S. J. (2003). *Mechanisms Underlying Enhanced Processing Efficiency in Neural Systems*. Pittsburgh: Carnegie Mellon University.
- Gotts, S. J., & Plaut, D. C. (2002). The impact of synaptic depression following brain damage: A connectionist account of "access/refractory" and "degraded-store" semantic impairments. *Cognitive, Affective, & Behavioral Neuroscience*, 2, 187–213.
- Gotts, S. J., della Rocchetta, A. I., & Cipolotti, L. (2002). Mechanisms underlying perseveration in aphasia: Evidence from a single case study. *Neuropsychologia*, 40, 1930–1947.
- Gotts, S. J., Milleville, S. C., Bellgowan, P. S., & Martin, A. (2011). Broad and narrow conceptual tuning in the human frontal lobes. *Cerebral Cortex*, 21, 477–491.
- Gotts, S. J., Chow, C. C., & Martin, A. (2012a). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*, 3, 227–237.

- Gotts, S. J., Chow, C. C., & Martin, A. (2012b). Repetition priming and repetition suppression: Multiple mechanisms in need of testing. *Cognitive Neuroscience*, 3, 250–259.
- Gotts, S. J., Milleville, S. C., & Martin, A. (2014). Object identification leads to a conceptual broadening of object representations in lateral prefrontal cortex. *Neuropsychologia*. doi:10.1016/j.neuropsychologia.2014.10.041
- Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107, 293–321.
- Grill-Spector, K., Henson, R. N., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Science*, 10, 14–23.
- Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, 20, 371–388.
- Hansen, B. J., & Dragoi, V. (2011). Adaptation-induced synchronization in laminar cortical circuits. *Proceedings of the National Academy of Sciences, USA*, 108, 10720–10725.
- Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, 34, 479–490.
- Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society B*, 362, 1601–1613.
- Henson, R. N. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, 70, 53–81.
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20, 1315–1326.
- Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus-response bindings in priming. *Trends in Cognitive Science*. doi:10.1016/j.tics.2014.03.004
- Hinton, G. E. (2003). The ups and downs of Hebb synapses. *Canadian Psychology*, 44, 10–13.
- Hodges, J. R., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain*, 115, 1783–1806.
- Hodgson, C., Schwartz, M. F., Brecher, A., & Rossi, N. (2003). Effects of relatedness, repetition, and rate: Further investigations of context-sensitive naming. *Brain and Language*, 104, 89–101.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, USA*, 79, 2554–2558.
- Horner, A. J., & Henson, R. N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia*, 46, 1979–1991.
- Horner, A. J., & Henson, R. N. (2012). Incongruent abstract stimulus-response bindings result in response interference: fMRI and EEG evidence from visual object classification priming. *Journal of Cognitive Neuroscience*, 24, 760–773.
- Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition*, 100, 464–482.
- Hsiao, E. Y., Schwartz, M. F., Schnur, T. T., & Dell, G. S. (2009). Temporal characteristics of semantic perseverations induced by blocked-cyclic picture naming. *Brain and Language*, 108, 133–144.
- Jacobs, R. A. (1999). Computational studies of the development of functionally specialized neural modules. *Trends in Cognitive Science*, 3, 31–38.
- Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias toward short connections. *Journal of Cognitive Neuroscience*, 4, 323–336.
- James, T. W., Humphreys, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). The effects of visual object priming on brain activation before and after recognition. *Current Biology*, 10, 1017–1024.
- Jiang, Y., Haxby, J. V., Martin, A., Ungerleider, L. G., & Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science*, 287, 643–646.
- Jiang, X., Bradley, E., Rini, R. A., Zeffiro, T., Vanmeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, 53, 891–903.
- Kaliukhovich, D. A., & Vogels, R. (2012). Stimulus repetition affects both strength and synchrony of macaque inferior temporal cortical activity. *Journal of Neurophysiology*, 107, 3509–3527.
- Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B - Biological Sciences*, 361, 2109–2128.
- Ketz, N., Morkonda, S. G., & O'Reilly, R. C. (2013). Theta coordinated error-driven learning in the hippocampus. *PLoS Computational Biology*, 9, e1003067. doi:10.1371/journal.pcbi.1003067
- Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia*, 39, 184–199.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, 4, 533–539.
- Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology*, 69, 1918–1929.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77, 24–42.
- Lueschow, A., Miller, E. K., & Desimone, R. (1994). Inferior temporal mechanisms for invariant object recognition. *Cerebral Cortex*, 5, 523–531.
- Lustig, C., & Buckner, R. L. (2004). Preserved neural correlates of priming in old age and dementia. *Neuron*, 42, 865–875.
- Maccotta, L., & Buckner, R. L. (2004). Evidence for neural effects of repetition that directly correlate with behavioral priming. *Journal of Cognitive Neuroscience*, 16, 1625–1632.
- Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, 14, 455–462.
- Mahon, B. Z. (2015). Missed connections: A connectivity constrained account of the representation and organization of object concepts. In E. Margolis & S. Laurence (Eds.), *Concepts: New Directions*. Cambridge: MIT Press.
- Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, 60, 27–51.
- Mahon, B. Z., & Caramazza, A. (2011). What drives the organization of object knowledge in the brain? *Trends in Cognitive Science*, 15, 97–103.
- Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiat, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, 55, 507–520.
- Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, 275, 213–215.
- Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, 58, 25–45.
- Martin, A., & Gotts, S. J. (2005). Making the causal link: Frontal cortex activity and repetition priming. *Nature Neuroscience*, 8, 1134–1135.
- Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science*, 270, 102–105.

- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, 379, 649–652.
- Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991). A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences, USA*, 88, 4433–4437.
- McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, 114, 159–188.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102, 419–457.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–165). New York: Academic Press.
- McKee, J. L., Riesenhuber, M., Miller, E. K., & Freedman, D. J. (2014). Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *Journal of Neuroscience*, 34, 16065–16075.
- McKone, E. (1995). Short-term implicit memory for words and non-words. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1108–1126.
- McKone, E. (1998). The decay of short-term implicit memory: Unpacking lag. *Memory & Cognition*, 26, 1173–1186.
- McMahon, D. B., & Olson, C. R. (2007). Repetition suppression in monkey inferotemporal cortex: Relation to behavioral priming. *Journal of Neurophysiology*, 97, 3532–3543.
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130.
- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1975). Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 98–118). London: Academic Press.
- Miller, E. K., Gochin, P. M., & Gross, C. G. (1991). Habituation-like decrease in the responses of neurons in inferior temporal cortex of the macaque. *Visual Neuroscience*, 7, 357–362.
- Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *Journal of Neuroscience*, 13, 1460–1478.
- Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. Cambridge: MIT Press.
- Mitchell, D. B. (2006). Nonconscious priming after 17 years: Invulnerable implicit memory? *Psychological Science*, 17, 925–929.
- Mulkey, R. M., & Malenka, R. C. (1992). Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. *Neuron*, 9, 967–975.
- Nabavi, S., Fox, R., Proulx, C. D., Lin, J. Y., Tsien, R. Y., & Malinow, R. (2014). Engineering a memory with LTD and LTP. *Nature*, 511, 348–352.
- Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, 11, 966–974.
- Navarrete, E., Mahon, B. Z., & Caramazza, A. (2010). The cumulative semantic cost does not reflect lexical selection by competition. *Acta Psychologica*, 134, 279–289.
- Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, 76, 253–272.
- Newman, E. L., & Norman, K. A. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral Cortex*, 20, 2760–2770.
- Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, 341, 52–54.
- Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, 32, 185–208.
- Norman, K. A., Newman, E. L., Detre, G. J., & Polyn, S. M. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*, 18, 1577–1610.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological Review*, 110, 611–646.
- Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15, 267–273.
- Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*, 4, 1244–1252.
- Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114, 227–252.
- O'Reilly, R. C. (1996). *Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm*. Pittsburgh: Carnegie Mellon University.
- O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, 13, 1199–1241.
- O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science*, 314, 91–94.
- Pedreira, C., Mormann, F., Kraskov, A., Cerf, M., Fried, I., Koch, C., & Quiroga, R. Q. (2010). Responses of human medial temporal lobe neurons are modulated by stimulus repetition. *Journal of Neurophysiology*, 103, 97–107.
- Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19, 603–639.
- Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology*, 28, 251–275.
- Plaut, D. C., & Shallice, T. (1993). Perseverative and semantic influences on visual object naming errors in optic aphasia: A connectionist account. *Journal of Cognitive Neuroscience*, 5, 89–117.
- Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, 72, 665–78.
- Race, E. A., Shanker, S., & Wagner, A. D. (2009). Neural priming in human frontal cortex: Multiple forms of learning reduce demands on the prefrontal executive system. *Journal of Cognitive Neuroscience*, 21, 1766–1781.
- Race, E. A., Badre, D., & Wagner, A. D. (2010). Multiple forms of learning yield temporally distinct electrophysiological repetition effects. *Cerebral Cortex*, 20, 1726–1738.
- Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annual Review of Neuroscience*, 29, 449–476.
- Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, 27, 179–189.
- Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12, 162–168.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge: MIT Press.
- Rogers, T. T., & McClelland, J. L. (2014). Parallel Distributed Processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, 38, 1024–1077.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536.

- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations & volume II: Psychological and biological models*. Cambridge: MIT Press.
- Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2, 539–550.
- Santo Pietro, M. J., & Rigrinsky, S. (1982). The effects of temporal and semantic conditions on the occurrence of the error response of perseveration in adult aphasics. *Journal of Speech and Hearing Research*, 25, 184–192.
- Sayres, R., & Grill-Spector, K. (2006). Object-selective cortex exhibits performance-independent repetition suppression. *Journal of Neurophysiology*, 95, 995–1007.
- Schacter, D. L. (1987). Implicit memory - History and current status. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 501–518.
- Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, 20, 185–195.
- Schacter, D. L., Wig, G. S., & Stevens, W. D. (2007). Reductions in cortical activity during priming. *Current Opinion in Neurobiology*, 17, 171–176.
- Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science*, 10, F15–F30.
- Scherf, K. S., Luna, B., Avidan, G., & Behrmann, M. (2011). "What" precedes "which": Developmental neural tuning in face- and place-related cortex. *Cerebral Cortex*, 21, 1963–1980.
- Schnur, T. T. (2014). The persistence of cumulative semantic interference during naming. *Journal of Memory & Language*, 75, 27–44.
- Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, 54, 199–227.
- Schnur, T. T., Schwartz, M. F., Kimberg, D. Y., Hirshorn, E., Coslett, H. B., & Thompson-Schill, S. L. (2009). Localizing interference during naming: Convergent neuroimaging and neuropsychological evidence for the function of Broca's area. *Proceedings of the National Academy of Sciences - USA*, 106, 322–327.
- Scholl, C. A., Jiang, X., Martin, J. G., & Riesenhuber, M. (2014). Time course of shape and category selectivity revealed by EEG rapid adaptation. *Journal of Cognitive Neuroscience*, 26, 408–421.
- Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in primate temporal cortex. *Nature*, 415, 318–320.
- Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, 15, 1602–1608.
- Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, 45, 2802–2810.
- Simmons, W.K., Rapuano, K.M., Kallman, S.J., Ingeholm, J.E., Miller, B., Gotts, S.J. ... & Martin, A. (2013). Category-specific integration of homeostatic signals in caudal but not rostral human insula. *Nature Neuroscience*, 16, 1551–1552.
- Sjöström, P. J., Turrigiano, G. G., & Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32, 1149–1164.
- Snowden, J. S., Goulding, P. J., & Neary, D. (1989). Semantic dementia: A form of circumscribed cerebral atrophy. *Behavioral Neurology*, 2, 167–182.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195–231.
- Squire, L. R., & Zola-Morgan, J. T. (2011). The cognitive neuroscience of human memory since H.M. *Annual Review of Neuroscience*, 34, 259–288.
- Srinivas, K. (1996). Size and reflection effects in priming: A test of transfer-appropriate processing. *Memory & Cognition*, 24, 441–452.
- Stark, C. E., & McClelland, J. L. (2000). Repetition priming of word, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26, 945–972.
- Taylor, J. C., & Downing, P. E. (2011). Division of labor between lateral and ventral extrastriate representations of faces, bodies, and objects. *Journal of Cognitive Neuroscience*, 23, 4122–4137.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247, 301–306.
- Turk-Browne, N. B., Yi, D. J., Leber, A. B., & Chun, M. M. (2006). Visual quality determines the direction of neural repetition effects. *Cerebral Cortex*, 17, 425–433.
- Tyler, L. K., Moss, H. E., Durrant-Peatfield, M. R., & Levy, J. P. (2000). Conceptual structure and the structure of concepts: A distributed account of category-specific deficits. *Brain and Language*, 75, 195–231.
- Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowsky, J., & Aston-Jones, G. (1999). The role of locus coeruleus in the regulation of cognitive performance. *Science*, 283, 549–554.
- van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience*, 22, 1270–1282.
- van der Linden, M., Wegman, J., & Fernandez, G. (2014). Task- and experience-dependent cortical selectivity to features informative for categorization. *Journal of Cognitive Neuroscience*, 26, 319–333.
- van Turennout, M., Ellmore, T., & Martin, A. (2000). Long-lasting cortical plasticity in the object naming system. *Nature Neuroscience*, 3, 1329–1334.
- van Turennout, M., Biemond, L., & Martin, A. (2003). Modulation of neural activity during object naming: Effects of time and practice. *Cerebral Cortex*, 13, 381–391.
- Verhoeve, B.-E., Kayaert, G., Franko, E., Vangeneugden, J., & Vogels, R. (2008). Stimulus similarity-contingent neural adaptation can be time and cortical area dependent. *Journal of Neuroscience*, 28, 10631–10640.
- Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded picture naming: It's in the links. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 17, 664–680.
- Vitkovitch, M., Humphreys, G. W., & Lloyd-Jones, T. J. (1993). On naming a giraffe a zebra: Picture naming errors across different object categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19, 243–259.
- von Stein, A., Chiang, C., & König, P. (2000). Top-down processing mediated by interareal synchronization. *Proceedings of the National Academy of Sciences, USA*, 97, 14748–14753.
- Voss, J. L., & Paller, K. A. (2008). Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. *Neuropsychologia*, 46, 3021–3029.
- Wang, Y., Iliescu, B. F., Ma, J., Josić, K., & Dragoi, V. (2011). Adaptive changes in neuronal synchronization in macaque V4. *Journal of Neuroscience*, 31, 13204–13213.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, 27, 635–657.
- Weiner, K. S., Sayres, R., Vinberg, J., & Grill-Spector, K. (2010). fMRI-adaptation and category selectivity in human ventral temporal cortex: Regional differences across time scales. *Journal of Neurophysiology*, 103, 3349–3365.
- Wig, G. S. (2012). Repetition suppression and repetition priming are processing outcomes. *Cognitive Neuroscience*, 3, 247–248.
- Wig, G. S., Grafton, S. T., Demos, K. E., & Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nature Neuroscience*, 8, 1228–1233.

- Wig, G. S., Buckner, R. L., & Schacter, D. L. (2009). Repetition priming influences distinct brain systems: Evidence from task-evoked data and resting-state correlations. *Journal of Neurophysiology*, *101*, 2632–2648.
- Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*, 227–233.
- Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.
- Woloszyn, L., & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*, 193–205.
- Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M. ... & Buckner, R.L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–65.