COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2014) 1–51
Copyright © 2014 Cognitive Science Society, Inc. All rights reserved.
ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12146

Interactive Activation and Mutual Constraint Satisfaction in Perception and Cognition

James L. McClelland,[a] Daniel Mirman,[b] Donald J. Bolger,[c] Pranav Khaitan[d]

[a]*Department of Psychology, Stanford University*
[b]*Department of Psychology, Drexel University and Moss Rehabilitation Research Institute*
[c]*Department of Psychology, University of Maryland*
[d]*Department of Computer Science, Stanford University*

Received 25 September 2011; received in revised form 1 November 2013; accepted 2 November 2013

Abstract

In a seminal 1977 article, Rumelhart argued that perception required the simultaneous use of multiple sources of information, allowing perceivers to optimally interpret sensory information at many levels of representation in real time as information arrives. Building on Rumelhart's arguments, we present the Interactive Activation hypothesis—the idea that the mechanism used in perception and comprehension to achieve these feats exploits an interactive activation process implemented through the bidirectional propagation of activation among simple processing units. We then examine the interactive activation model of letter and word perception and the TRACE model of speech perception, as early attempts to explore this hypothesis, and review the experimental evidence relevant to their assumptions and predictions. We consider how well these models address the computational challenge posed by the problem of perception, and we consider how consistent they are with evidence from behavioral experiments. We examine empirical and theoretical controversies surrounding the idea of interactive processing, including a controversy that swirls around the relationship between interactive computation and optimal Bayesian inference. Some of the implementation details of early versions of interactive activation models caused deviation from optimality and from aspects of human performance data. More recent versions of these models, however, overcome these deficiencies. Among these is a model called the multinomial interactive activation model, which explicitly links interactive activation and Bayesian computations. We also review evidence from neurophysiological and neuroimaging studies supporting the view that interactive processing is a characteristic of the perceptual processing machinery in the brain. In sum, we argue that a computational analysis, as well as behavioral and neuroscience evidence, all support the Interactive Activation hypothesis. The evidence suggests that contemporary

Correspondence should be sent to James L. McClelland, Department of Psychology, 344 Jordan Hall, Bldg 420, 450 Serra Mall, Stanford University, Stanford, CA 94305. E-mail: mcclelland@stanford.edu (or) Daniel Mirman, Department of Psychology, Stratton Hall, 3141 Chestnut St., Drexel University, Philadelphia, PA 19104. E-mail: daniel.mirman@drexel.edu

versions of models based on the idea of interactive activation continue to provide a basis for efforts to achieve a fuller understanding of the process of perception.

Keywords: Perception; Interactive activation; Parallel distributed processing; Connectionist models; Optimal perceptual inference; Neural networks

1. Introduction

One of the foundational concepts in the parallel distributed processing (PDP) framework is interactive activation. Interactive activation is synonymous with the concept of mutual constraint satisfaction: The idea is that, as a general principle, perceptual, linguistic, and other mental representations arise through the bidirectional propagation of activation among simple, neuron-like processing units. The concept was central to the interactive activation (IA) model of letter and word perception (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981, 1982) and the TRACE model of speech perception (McClelland & Elman, 1986). In these models, the focus was on bidirectional interactions between units standing for wholes and parts, such as words and letters or phonemes; letters and letter features; and phonemes and their features. In these models, individual neuron-like processing units were assigned to represent explicitly enumerable perceptual units such as words, letters, phonemes, and features. The processing units might be viewed as standing for populations of neurons dedicated to the corresponding cognitive units (Bowers, 2009), but we hold a different view. In line with the proposal of Smolensky (1986), the processing units in the model stand for informational states encoded as alternative patterns of activity over populations of neurons each of which participates in the representation of many different items (Hinton, McClelland, & Rumelhart, 1986; Plaut & McClelland, 2010). IA models track the time evolution and content of such states, a useful projection of the full complexity of the underlying neural activity into what Smolensky called a conceptual representational space, where their relationship with overt behavior such as letter, phoneme, or word identification is easier to track.

As detailed below, the empirical motivation for interactive activation models is the observation that, in experiment after experiment, the identification or interpretation of any element or aspect of a visual, auditory, or other input is influenced by the identity and interpretation of every other element or aspect of the input. Correspondingly, there is a motivation at the level of a theory of optimal perceptual interpretation: In general, direct sensory evidence for the interpretation of an input at any level of perceptual description can be inconclusive when considered in isolation, and the most likely interpretation of each element can only be determined when the interpretation of all elements and many sources of evidence are considered together. Indeed, a single coherent interpretation of all elements may well be strongly determined by the totality of the evidence, even though all of the individual elements of evidence are highly ambiguous (Fig. 1a).

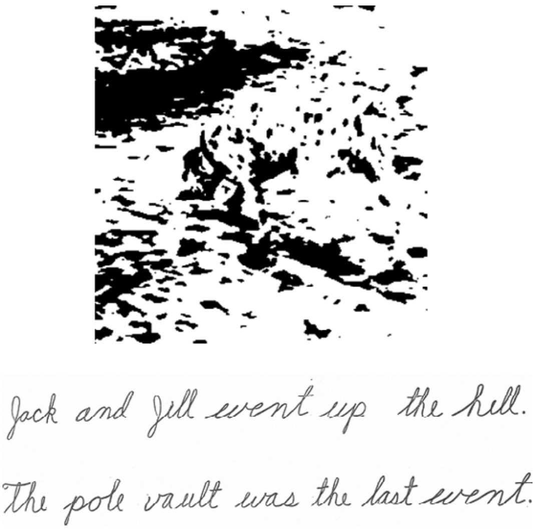
Jack and Jill went up the hill.
The pole vault was the last event.

Fig. 1. Top: A Dalmatian dog emerges from an assemblage of individually uninterpretable blotches. From James (1965). Copyright © Ronald C. James, reprinted with permission. Bottom: The hand-written words "went" in the first sentence and "event" in the second are identical, but they are perceived differently in the two different contexts. From fig. 3, p. 579 of Rumelhart (1977). Copyright © Taylor and Francis Group, reprinted with permission.

Beyond the domain of perception and comprehension, multiple simultaneous constraints apply to selection of aspects of contextually appropriate actions and reconstruction of memories, as well as many other aspects of cognition. Likewise, goals and task demands provide additional constraints that are integrated into perception, interpretation, remembering, and action, thus influencing, and often being influenced by, the outcome of processing. Chapter 1 of the PDP volumes (McClelland, Rumelhart, & Hinton, 1986) argued that these same considerations arise in all other areas of cognitive processing, including action selection, problem solving, and memory.

The idea that all aspects of perception and cognition involve parallel distributed processing in this way is an alternative to modular approaches to perception and cognition. Interactive processing allows for the possibility that specific neurons or neural populations in particular brain areas may be specialized to represent one or another type of information, so a certain kind of compartmentalization of information remains. In order, however, for all sources of information to simultaneously constrain all others, any outcome in which a particular ensemble of such neurons is active is thought to be the consequence of processing that is distributed across neural populations in multiple brain areas, including neurons that represent information of many different types. Thus, for example, while there can be brain regions dedicated to the representation of visual, semantic, auditory, and articulatory aspects of a visually presented word, the activations of neurons in all of the participating brain regions are taken to be mutually interdependent within the interactive activation/mutual constraint satisfaction framework.

1.1. Precursors to interactive activation models

The motivation for an interactive approach to perception and comprehension was laid out in a paper by Rumelhart (1977). Rumelhart reviewed existing data going back to the 19th century on the role of context in letter, phoneme, and word perception (Fig. 1b), and on the use of a range of sources of information in resolving ambiguities in syntactic and semantic interpretation of both spoken and written words and sentences. He took the goal of perception and comprehension to be to find a joint interpretation of an input at many different levels of representation, through a mutual constraint satisfaction process guided by knowledge of the prior probabilities of alternative hypotheses and of conditional probabilistic relations between these alternatives. Rumelhart went on to envision how a process of settling on such an interpretation might take place. Drawing inspiration from *Hearsay* (Reddy, Erman, Fennell, & Neely, 1973), an early artificial intelligence model of speech perception, he envisioned a data structure called a "message center" or "blackboard," where estimates of the probabilities of possible elements of the interpretation of an input could be "chalked in" for inspection and adjustment by specialized experts, each working in parallel on the contents of the blackboard. For example, for the case of written input, the estimate of the probability that the letter in a particular position in a word might be the letter A might be increased by a lexical expert that used information about a preceding C and a subsequent T along with lexical information that C, followed by A and T, spells the familiar word CAT. The lexical-level CAT hypothesis might be further strengthened if the participant has just viewed a picture containing a drawing of a cat. At the feature, letter, and word levels, the model drew on an earlier model by Rumelhart and Siple (1974) that relied on knowledge of word and letter probabilities and the conditional probabilities of letters given words to account for data on the identification of letters in displays of three-letter sequences.

2. The computational problem addressed by interactive activation models

The arguments laid out by Rumelhart (1977) support the following statement of the computational challenge faced in perception and language comprehension:

> *Search for the most probable interpretation*. Perception and language understanding are the process of seeking the most probable interpretation of a written or spoken input at many different levels of representation. An interpretation, for example, of a written or spoken linguistic expression represents the visual or auditory features present; the letters or speech sounds; the words, phrases, and sentences; and the meaning and syntactic structure of these items. The goal of the process is to find the interpretation that has the highest probability overall.

> *Exploitation of prior knowledge and context.* Because of the ubiquity of ambiguity and noise, maximizing the probability of finding the correct interpretation of any given aspect of the perceptual input depends on exploitation of prior knowledge and information from context, including adjacent elements in the expression itself, prior input, and input from other domains such as accompanying visual information.

Although Rumelhart (1977) did not stress it, we add the following important real-time constraint on a model of perception and comprehension:

> *Real-time processing constraint.* Perception and comprehension must deliver results as quickly as possible, allowing information of all different types to influence interpretation of information of all other types as it becomes available.

Our inclusion of this constraint in the formulation of the problem of perceptual inference differs from typical computational-level formulations (Feldman, Griffiths, & Mrogan, 2009; Marr, 1982), in which only inputs and outcomes are considered, without consideration of the time or processing steps required to compute the outcome. Clearly, though, time is precious, and in a dynamic world, failure to comprehend (and act) quickly can lead to missed opportunity and sometimes, catastrophe. Thus, achieving results as quickly as possible in real time is part of the computational-level challenge facing the perceptual system. Researchers coming from a computational-level starting point have begun to consider the importance of this issue (Norris, 2013; Vul, Goodman, Griffiths, & Tenenbaum, 2014).

2.1. Human perception and comprehension as an approximation to optimal perceptual inference in real time

The above statements characterize the computational problem a system of perception and comprehension must solve. Our next proposition states that human perception and comprehension mechanisms are organized to address these computational considerations:

> *Humans approximate optimal real-time perceptual inference.* Human perceivers approximate the patterns of behavior we would expect from an optimal system of perception and comprehension, exploiting context and prior knowledge to guide perception and comprehension and reflecting the influence of all sources of external input on all aspects of the interpretation as the input becomes available in real time.

There are limits on speed and accuracy that are imposed by the characteristics of neural hardware, affecting the extent to which humans can achieve a close approximation to optimality. We also note that experience is required for optimization, so that speed and accuracy both increase gradually with practice and exposure. The consequences of experience involve learning about the statistical structure of the perceptual world, tuning of perceptual and other cognitive systems to exploit this structure, and allocation of brain resources (neurons and synapses) to support performance. In the present article, we focus on perception and comprehension by skilled adults perceiving and comprehending spoken and written input from their native language, assuming that experience-dependent optimization has already occurred.

2.2 The interactive activation hypothesis

The statement of the problem and the characterization of human performance given above appear to be widely accepted, but several alternative approaches have been taken to characterizing the mechanisms that allow human perceivers to succeed in exploiting context and prior knowledge effectively. In this article, we consider the following hypothesis:

> *Interactive activation hypothesis.* Implementation of perceptual and other cognitive processes within bidirectionally connected neural networks in the brain provides the mechanism that addresses the key computational challenges facing perceptual systems, and it gives rise to the approximate conformity of human performance to optimal perceptual inference in real time.

In what follows, we discuss the history of research on interactive models in perception. We examine the early IA and TRACE models and the experimental evidence relevant to their fundamental assumptions. We consider how well they address the computational challenges specified above, and we consider how consistent they are with evidence from behavioral experiments. We examine empirical and theoretical controversies surrounding the idea of interactive processing, including a controversy that has swirled around the relationship between interactive computation and optimal Bayesian inference. We also review evidence from neurophysiological and neuroimaging studies of the neural basis of perception. To anticipate our conclusions: Computational analysis as well as behavioral and neuroscience evidence are all consistent with the Interactive Activation hypothesis. Although there have been and will likely remain those who advocate for alternative approaches, the evidence suggests to us that contemporary versions of models based on these ideas have considerable merit. At the end of the article, we revisit this conclusion and consider ways in which interactive approaches may develop in the future.

3. The interactive activation and TRACE models

Testing the IA hypothesis requires the development of explicit models that embody its assumptions, as well as the analysis of these models to understand their properties and to examine the extent of their ability to account for patterns in human behavior. The Interactive Activation model of letter and word perception (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1981, 1982), and its offspring, the TRACE Model of speech perception (McClelland & Elman, 1986), represented initial steps in such a research program, focusing primarily on modeling patterns in data.

The IA model of letter and word perception addresses the perception of letters presented in one of four display locations, either alone or together with neighboring letters in the other locations. Position-specific pools of neuron-like processing units are posited at feature and letter levels, and a word level spans the array of input positions (Fig. 2a). There are bidirectional excitatory connections between mutually consistent units in adjacent levels and bidirectional inhibitory connections among units within each pool. Before presentation of a stimulus, all units' activation values are set to a resting level slightly below 0. External input, once presented, drives feature units, which in turn activate consistent letter units and inhibit inconsistent letter units.[1] Letter units in turn activate

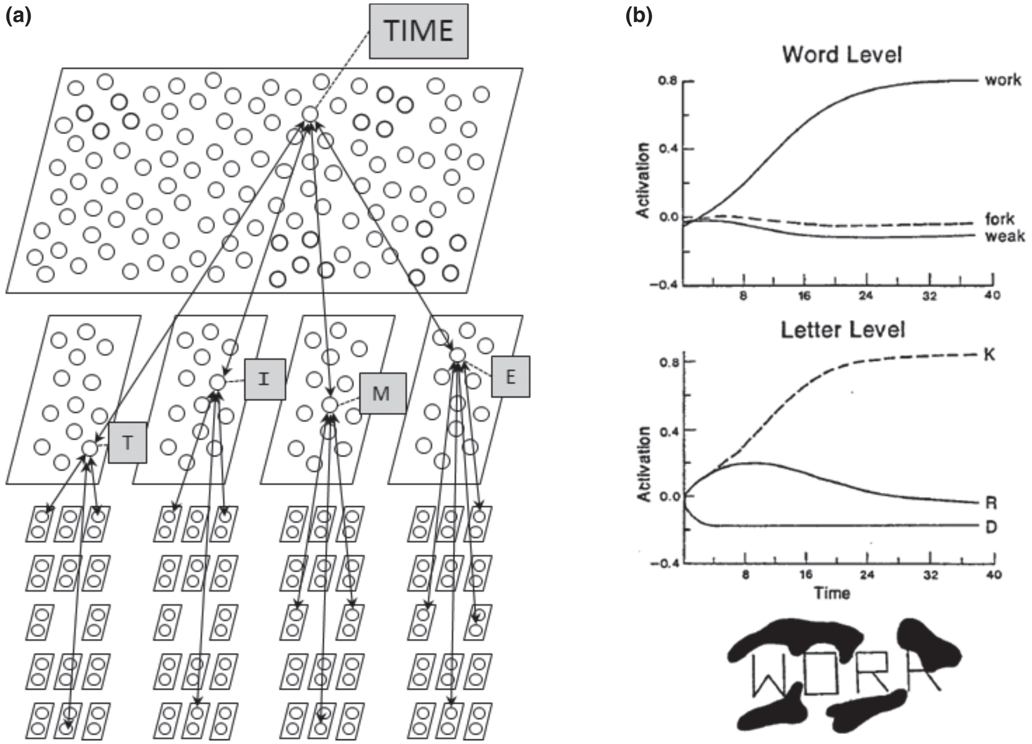

Fig. 2. (a) The interactive activation model, indicating the pools of units corresponding to words, letters in each of four positions, and features in the same positions. Excitatory connections for the word TIME and the letters and features of this word are shown. Units within each pool are mutually inhibitory, though the inhibitory connections are not drawn in. At the feature level, units are organized into pools consisting of two units, one for the presence and one the absence of each possible feature. Reprinted from fig. 6, p. 14 of McClelland (2013). Copyright ©James L. McClelland, reprinted with permission. (b) The time course of activation of letter units in the fourth position and word units in the original version of the interactive activation model, after presentation of the display shown below the figure. The visible segments in the last position are equally consistent with the letters K and R, and inconsistent with other letters. At the word level, only one known word, WORK, is consistent with the active letters in each of the four positions. This word feeds back to support the unit for K, which then dominates R in the fourth position. Reprinted from fig. 8, p. 23 of McClelland et al. (1986). Copyright © MIT Press, reprinted with permission.

consistent word units which compete with each other and also send feedback to support letters consistent with possible words. An illustration of this process, as it applies to the ambiguous input indicated in Fig. 2b, shows the time course of activation, demonstrating how it can find the contextually most likely interpretation within a few processing cycles. Although the featural input in the fourth position is equally consistent with R or K, only K makes a word (WORK) with the context letters. Due to bottom-up support from active letters in all four positions, this word becomes more active than any other word; it suppresses other competing word alternatives and provides top-down support for K, which then suppresses R via competition, leading to a state in which there is a consistent interpretation of the input at both the letter and word levels.

Details of the interactive activation process. We describe the details of the activation process as it was conceived in the IA and TRACE models. These details will be relevant later to our discussion of the relationship between interactive activation and optimal inference. The activation process, as originally formulated (adapting proposals of Grossberg, 1978), assigned continuously varying activation values to units for letters and words. The process is in principle viewed as a completely continuous process, approximated in simulations as a series of fine-grained time steps. During each time step, for each unit, its *net input* is first calculated. This is the sum, over all units projecting to it, of the activation of the sending unit times the value of the incoming connection weight from that unit, plus any direct external input to the unit:

$$net_i = \sum [a_j]^+ w_{ij} + e_i$$

The strengths of excitatory and inhibitory weights were determined by separate parameters for feature-to-letter, letter-to-word, word-to-letter, and within-layer influences. The notation $[a_j]^+$ indicates that a unit's activation value is only propagated if greater than 0.

Once the net input to each unit has been established, activations are adjusted as follows:

$$\begin{aligned} \text{If } (net_i \geq 0) &: \Delta a_i = net_i(1 - a_i) - d(a_i - r) \\ \text{otherwise} &: \Delta a_i = net_i(a_i - m) - d(a_i - r) \end{aligned}$$

These equations implement a process in which a positive net input pushes activation up toward a maximum value of 1, while a negative net input pushes activation down toward a minimum (m), usually set to $-.2$ or $-.3$. The rightmost term in each equation implements a restoring force sometimes thought of as corresponding to a decay or leakage process that tends to pull activation values toward their resting level (r); the parameter d represents the strength of this tendency.

Processing in the model is completely deterministic. To address human performance in perception experiments, where performance is probabilistic, predicted response probabilities are derived by applying the Luce choice rule to a running average of the resulting activation values, so that the probability of choosing alternative i is given by:

$$p(r_i) = e^{g\bar{a}_i} \Big/ \sum\nolimits_{i'} e^{g\bar{a}_{i'}}$$

For example, the probability of choosing the letter K as the identification response for the letter in the fourth position in the display in Fig. 2a would be calculated by setting i to be the index of the unit corresponding to the letter K in the fourth position. The index i' runs over all the letters in the same position, including the one indexed by i, and g is a scaling parameter. The quantity $\bar{a}_i$ corresponds to the running average activation of the unit for the letter in question at the time when the network is interrogated. For most of the experiments modeled in McClelland and Rumelhart (1981), this time was taken to be the time post-stimulus onset that resulted in highest possible probability of correct responding.

The TRACE model extends the ideas from the IA model to the processing of a stream of speech by postulating a much larger number of position-specific feature and letter unit arrays, as well as corresponding banks of position-aligned word units, so that there is a unit for every feature and phoneme at each position, and a unit for every word starting at every position. As spoken input arrives sequentially in real time, each successive time sample of the spoken input is directed to the next input position. In this way, the same bidirectional activation process as captured in the IA model of letter perception could be applied to the processing of spoken inputs corresponding to one or a few words. The architecture allowed phoneme-level and word-level constraints to be applied to sequences of input samples regardless of where in the input stream these samples occurred. The structure of the TRACE model should not be viewed as a literal claim about the neural mechanism. Instead, it should be seen as a higher-level characterization capturing the relative rather than absolute constraints between phoneme- and word-level information: If there is a /k/ at a particular time, it supports the word "cat" starting in the same time, and the word "ticket" starting two phonemes earlier (among many other possibilities), and these constraints are captured in the connections between units for the corresponding items in the corresponding positions.[2] Activation in this array of units formed a dynamic memory trace of the results of processing a spoken input, hence the name of the model. The architecture was inspired by the earlier concept of the blackboard as discussed by Rumelhart (1977), and a model developed at about the same time (McClelland, 1985, 1986) explored how neural hardware might implement these computations without the reduplication of units and connections.

4. Behavioral evidence

4.1. Empirical foci of the IA and TRACE models

The IA and TRACE models targeted letter and phoneme perception, addressing a large body of relevant data illustrating effects of word context on recognition of letters and speech sounds. Much of the early behavioral evidence can be summarized as explorations of *word superiority effects*: Letters are recognized more accurately when presented within words than when presented in isolation or in random sequences of letters (e.g., Reicher, 1969). The models also addressed the ubiquitous finding that ambiguous visual and speech inputs are likely to be identified as letters or phonemes consistent with surrounding lexical context (e.g., Ganong, 1980; Massaro, 1979). For example, Ganong (1980) showed that an ambiguous sound between /k/ and /g/ was more likely to be identified as /k/ in an "_iss" context (where it fits to form the word "kiss") and as /g/ in an "_ift" context (where it fits to form the word "gift"). The advantage for letters in words also extended to letters in pronounceable, word-like pseudowords (such as LEAT or TOVE, McClelland & Johnston, 1977). The IA model of word perception provided a novel account of the mechanism by which letters in pseudowords like LEAT were perceived more accurately than letters in unword-like non-words (e.g. LTAE) or single letters presented without context; in the model, the pseudoword advantage occurred through the partial activation of units for words sharing several letters with the presented input. Such items are called *neighbors* of the given input. These word units then fed back support to the units for the constituent letters, many of which are partially supported by activations of several different words (Fig. 3). Newman, Sawusch, and Luce (1997) demonstrated neighborhood effects in identification of ambiguous speech segments, consistent with this account. The IA model predicted that letters in unpronounceable strings that nevertheless had many word "neighbors" (e.g., the "L" in SLNT) would show as much facilitation as letters in comparable pronounceable strings (SLET), and an experiment reported in Rumelhart and McClelland (1982) confirmed this prediction.

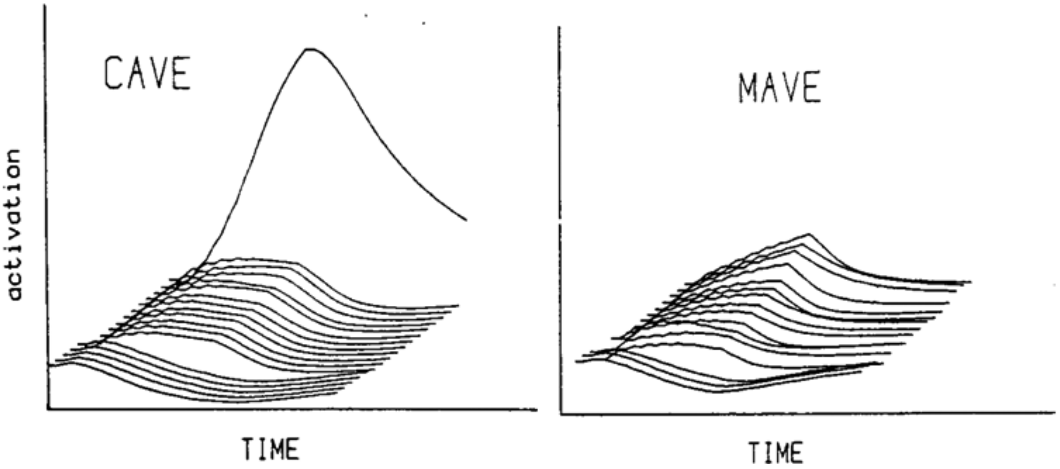

Fig. 3. Activations at the word level produced by CAVE and MAVE in the interactive activation model. Activations of all units whose activation exceeds 0 at any time during processing are shown. Activation traces are offset spatially with those reaching higher maximal activations starting behind and to the right. In the case of MAVE, several words contribute top-down support to the presented letter in each of the four-letter positions. From fig. 13, p. 396 and fig. 9, p. 393 of McClelland and Rumelhart (1981). Copyright © American Psychological Association. Reprinted with permission.

The bidirectional interactive processing in the IA and TRACE models predicts that context effects can occur for contextual elements that come after a target input element, as well as for elements that come before the target. This prediction was tested and confirmed in experiments that separately manipulated the duration of each context letter and examined its effect on the recognition of target letters in each letter position (Rumelhart & McClelland, 1982). In general, *all* context letters influence accuracy of perception of *each* target letter. Similarly, lexical effects on phoneme recognition occur for word-initial as well as embedded or word-final phonemes (Ganong, 1980; Warren, 1970), and the effects extend to contextual information in subsequent words in some studies (Sherman, 1971; Warren & Warren, 1971). Of course, if perceivers in a phoneme identification task are required to respond too soon after an ambiguous segment, subsequent context has little effect (Fox, 1984), and this was captured in simulations using the TRACE model. A wide range of additional phenomena in speech perception, including lexically based segmentation of a stream of spoken sounds into words and the perceptual magnet effect (Kuhl, 1991), were also addressed by the TRACE model.

Evidence of human conformity to the real-time processing constraint. One of the motivating phenomena leading to the development of the TRACE model was evidence supporting the view that word identification occurs in real time during speech perception. Marslen-Wilson and colleagues were the first to focus on this point, showing that identification occurs very shortly after a spoken input becomes uniquely consistent with a single possible word (Marslen-Wilson & Welsh, 1978). A large body of subsequent work examining eye movements during spoken word-to-picture matching tasks further supported the general principle that context and stimulus information mutually constrain processing in real time. Several of these studies include both non-linguistic visual input as well as spoken auditory input, as envisioned in the 1977 paper by Rumelhart. The initial experiments using this method showed that visual context influenced the immediate interpretation of a syntactically ambiguous prepositional phrase (Chambers, Tanenhaus, & Magnuson, 2004; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Subsequent studies also showed that syntactic and semantic expectations can also constrain which lexical candidates are considered. For example, Dahan and Tanenhaus (2004) showed that participants rule out possible target objects upon hearing a verb (such as “climb”) that rules out some of the objects as potential objects of the action named by the verb (e.g., a watch). Critically, these contextual influences became evident very soon after the constraining information was presented (Dahan & Tanenhaus, 2004; Magnuson, Tanenhaus, & Aslin, 2008) and were continuously updated as new information became available. This was demonstrated particularly clearly by Allopenna, Magnuson, and Tanenhaus (1998) in a study that showed that about 200 ms after word onset—the minimum required to plan and execute an eye movement—listeners were already more likely to fixate objects whose names matched the initial consonant and vowel of the word. Furthermore, their results showed that word candidates that did not match an input’s onset could still become activated if supported by enough subsequent phonological input, consistent with the idea of a set of candidates whose activations are continuously updated in light of ongoing input. Many of these papers simulated their findings using the TRACE model or simplified models based on similar assumptions (Spivey & Tanenhaus, 1998).

4.2. Evidence of the generality of context effects

Word context effects on recognition of letters and phonemes have served as a major focus for research on interactive processing, but the principle is very general and recurs across many different domains of perception and cognition. For example, just as in word recognition, there is a tendency for phonological errors in speech production to result in existing words rather than non-words, and such effects are well explained by interactive models of speech production (Dell, 1986; see also Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Rapp & Goldrick, 2000).

Interactive processing also plays an important role in visual object perception. Just as in the word advantage effects, perception of an ambiguous color can be biased by object context (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Kubat, Mirman, & Roy, 2009). For example, an ambiguous color halfway between yellow and orange is perceived as more yellow in the context of a school bus and as more orange in the context of a carrot. Furthermore, paralleling a result from Elman and McClelland (1988) discussed below, Mitterer and de Ruiter (2008) showed that object-context feedback can recalibrate color categories. The well-known illusory contours phenomenon in Kanizsa figures (Kanizsa, 1979; Fig. 4) demonstrates that a simple figure context can even induce the perception of contours that are completely absent from the input, as expected from interactive activation.

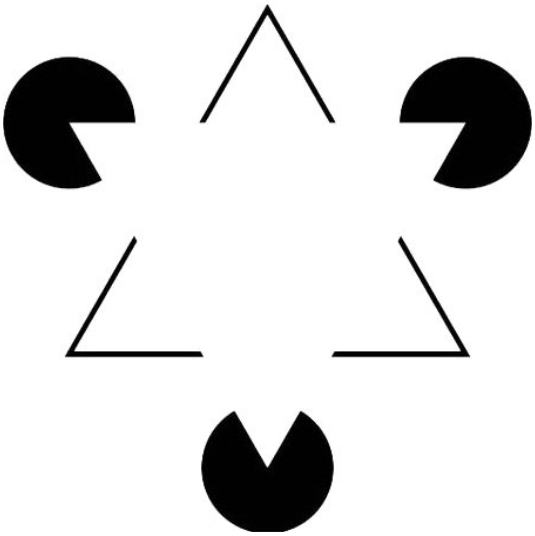

Fig. 4. Illusory contours in the Kanizsa triangle. Image source: Kanizsa triangle, Wikimedia Commons, http://en.wikipedia.org/wiki/File:Kanizsa_triangle.svg. Copyright © Wikipedia Commons. Reprinted under the GNU Free Documentation License.

Moving to higher level phenomena, it has been clear for many years that context affects the resolution of lexical ambiguity, as Rumelhart (1977) predicted. There are models of such effects that restrict context effects to a post-access selection process (Swinney, 1979), but interactive models predict that if the context is sufficiently constraining, then it could constrain which meanings of an ambiguous word are initially activated (e.g., Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982) and even cause pre-activation before the input is presented (McClelland, 1987). Eye-tracking studies have revealed such anticipatory effects in language processing in adults (e.g., Altmann & Kamide, 1999; Magnuson et al., 2008; see also the Dahan & Tanenhaus, 2004 study mentioned above) as well as infants (e.g., McMurray & Aslin, 2004). Electrophysiological scalp recordings (ERP) also suggest that words can be pre-activated by sentence contexts (van den Brink, Brown, & Hagoort, 2001; DeLong, Urbach, & Kutas, 2005).

The study of context effects has focused on how perception of elements such as letters or phonemes (or edges or colors) is affected by their immediate context (e.g., words or objects). However, processing is also affected by other contextual factors, including task instructions and relative probability of different types of stimuli. For example, lexical context effects are reduced when the proportion of non-words in a block of trials in a perceptual experiment is relatively high. Specifically, if the non-word proportion is high, the speed advantage for recognition of phonemes in words compared to non-words is reduced (Mirman, McClelland, Holt, & Magnuson, 2008), the word bias in speech errors is reduced (Hartsuiker, Corley, & Martensen, 2005), the short-term memory advantage for words over non-words is reduced (Jefferies, Frankish, & Ralph, 2006), and there is an increase in regularization errors in reading words that have inconsistent letter-sound mappings (e.g., reading “pint” to rhyme with “mint”; Monsell, Patterson, Graham, Hughes, & Milroy, 1992). These results can be interpreted as reflecting reduced activation of lexical (or possibly semantic) representations so that representations of words are less active and consequently have a smaller feedback effect (for implementations of these effects within TRACE, see Mirman et al., 2008).

The modulation of processing through attention can be implemented in networks of bidirectionally connected processing units—that is, interactive activation networks. One example of such a model is the model of attentional modulation of processing in the Ericksen flanker task (Cohen, Servan-Schreiber, & McClelland, 1992). In this model, units standing for different spatial locations are bidirectionally connected with units for features in these locations, and these units are, in turn, bidirectionally connected with position-independent units of the alternative possible target letter identities. Directing attention to a location is thought to arise from top-down activation of the unit standing for that spatial location; this enhances the activation of units for features in the corresponding position, giving them an eventual upper hand in subsequent processing, but allowing activations from inconsistent flankers nevertheless to retard identification of the item in the target location (as is observed in experiments, for example, Gratton, Coles, Sirevaag, Eriksen, & Donchin, 1988). Although some implementations of the Cohen et al. model have simplified the architecture such that not all connections are bidirectional, we take it as a given that attention to locations and stimulus features involves a bidirectional propagation of activation such that salient inputs, as well as goals and task demands, can participate in determining the focus of attention (Phaf, Van der Heijden, & Hudson, 1990). Furthermore, a recent model of interactive engagement between dorsal (action) and ventral (object) processing systems illustrates how interactive processing can facilitate the simultaneous identification of two or more objects present in a display at the same time (Henderson & McClelland, 2011).

Finally, we note that interactive activation processes may also play an important role in memory (Kumaran & McClelland, 2012; McClelland, 1981). A cue (such as an individual's name) can activate a representation of the item in memory, and this in turn can activate known features of the item, which then, through recurrence, activate other similar items. These items then in their turn can fill in additional features that are then attributed to the cued item. This use of interactivity extends similarity-based generalization models to cases in which relevant items in memory do not overlap with the cue (the individual's name may be unique) but do overlap on other dimensions that are brought into the computation via recurrent, interactive computation.

5. Interactive processing and optimal perceptual inference

While the above indicates some of the empirical support for the IA and TRACE models and demonstrates that the applicability of the principle of interactive activation extends beyond the domain of perception, it does not explicitly address the question of the relationship between the IA model and optimal perceptual inference. The topic has been the source of a heated critique in the literature on visual and auditory context effects (Massaro, 1989; Massaro & Cohen, 1991; Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000). The papers just cited argue that interactive processing will distort perception away from the pattern that is both seen in behavioral data and expected if information integration is consistent with principles of Bayesian inference, and that interactive activation causes undue contextual influence, producing, for example, inappropriate "hallucination" of lexically consistent phonemes.

It is ironic that the IA hypothesis would face such critiques, given that Rumelhart's early ideas about context effects on perception (Rumelhart, 1977; Rumelhart & Siple, 1974) were explicitly formulated in terms of probabilistic, Bayesian inference. Furthermore, the "hallucinatory" perception of contextually consistent phonemes observed in the models is, for us, exactly what the model *should* produce, both from the point of view of optimal performance in natural contexts and from the point of view of accounting for the findings in human perception. Consider what happens when a brief noise burst occurs coincident with the production of a phoneme in a spoken sentence. Listeners are likely to perceive (perceptually restore) the correct speech sound in such cases, even when the noise replaces the sound rather than being played over it (Samuel, 1981; Warren, 1970). The perception of the phoneme is in some sense a hallucination, but in a natural context, the inference that the speaker has produced the contextually appropriate sound is far more likely to be correct that the inference that he suspended his speech for the exact duration of the noise burst. In general, exploiting context to determine what we hear is more likely to lead us to hear what was really said, except in experiments where natural probabilistic contingencies can be broken.

It is true, however, that in developing the interactive activation model, Rumelhart and McClelland (1981, 1982) and McClelland and Rumelhart (1981) gave no explicit consideration to a probabilistic formulation of the problem of perception per se, drawing instead on the non-probabilistic, neurally inspired processing models proposed by Grossberg (1978, 1980) without considering whether this formulation corresponded exactly to optimal probabilistic inference. In retrospect, this appears to have led to unfortunate misunderstandings and needless controversies that we hope to put to rest in the present article. Specifically, subsequent research on interactive activation models supports two key points:

1. The IA and TRACE models, in their original formulation, did not provide an exact implementation of a principled Bayesian computation; indeed, the initial formulation of these models did distort these computations, in ways that deviate both from optimality and from human data.
2. However, variants of the models that retain their essential interactive character are consistent with Bayesian principles and can capture data that were problematic for the original formulation.

Regarding point (1), flaws in the original IA and TRACE models are discussed in McClelland (1991). There, it was observed that the activation assumptions of the model together with the assumptions about the translation of these activations into response probabilities produced patterns of choice responses that deviated from Bayesian probabilistic models and from human choice responding. These deviations occurred even in the absence of any interactivity in processing: That is, they occurred even when two sources of bottom-up information were combined to determine the activation of units standing for possible choice alternatives. Thus, the shortcomings of the original model may not have been a consequence of interactivity per se.

Here, we consider point (2) above in more detail. Specifically, we describe how a variant of the IA model called *the multinomial interactive activation* (MIA) model (Khaitan & McClelland, 2010) operates according to Bayesian principles of perceptual inference, considering the case of a display containing a sequence of four letters, as in most of the experiments modeled by the original IA model. A fuller treatment of the probabilistic principles and their relationship with computations in artificial networks is provided in McClelland (2013), and that article should be consulted by those interested in the details behind the briefer presentation here.

The MIA model draws heavily on insights brought into research on artificial neural networks by Hinton and Sejnowski in the form of the *Boltzman Machine*, first presented in a conference proceedings paper describing how such a machine could perform optimal perceptual inference (Hinton & Sejnowski, 1983), and subsequently described in the PDP volumes (Hinton & Sejnowski, 1986). We begin by describing the relevant ideas from the original Boltzmann machine.

5.1. States, their goodness, and their probability in the Boltzmann machine

In Boltzmann machines, units take on binary activation values (0 or 1). Units (which we index with the subscripts i and j) are thought of as corresponding to perceptual predicates about a sensory input (e.g., the input contains a particular line segment at a particular location, or it signals the presence of a particular object at some location). A consequence of using binary activation values is that it makes it relatively easy to consider, not only unit-by-unit probabilities but also the probability of different overall states of the network. Each state S_π corresponds to a specific pattern of [0, 1] values over all of the units, and each state has a *Goodness* G_π, corresponding to how well the state satisfies the graded constraints encoded in the connection weights (w_{ij}) among active units (a_i and a_j) and the bias terms associated with the units (b_i). Weights can be thought of as encoding probabilistic constraints between pairs of predicates, and biases can be thought of as encoding prior probabilities of individual predicates, in ways we will make precise for the MIA model below. The goodness of a state is defined as

$$G_\pi = \sum\nolimits_{i>j} w_{ij} a_i a_j + \sum_i a_i b_i$$

The subscripts i and j run over all units in the network, and the notation $i > j$ simply indicates that the connection between a pair of units, which is assumed to be symmetric ($w_{ij} = w_{ji}$) is only counted once in measuring goodness. The goodness is greater when the bias terms on active units are more positive and when the weights between active units are more positive.

When performing perceptual inference in a Boltzmann machine, some of the units may be forced or *clamped* into specified 0 or 1 values, corresponding to a sensory input pattern, while the activation values of the remaining units are set by a probabilistic updating process. The resulting states of these unclamped units are thought of as a possible interpretation of the sensory input. In the original Boltzmann machine, this updating process consisted of a sequence of updates, each of which involved selecting an unclamped unit at random. Indexing this unit as unit i, we then set its activation depending on its net input, $net_i = \sum_j a_j w_{ij} + b_i$, where j runs over units with connections to unit i. Once the net input is computed, the units' activation is set to 1 with probability

$$p_i = \frac{1}{1 + e^{-net_i/T}}$$

or to 0 with probability 1-p_i. T is a parameter called *temperature*, determining how strongly the activation is constrained by the unit's net input.

If this process is allowed to iterate for a sufficient number of updates, the probability that the network will be in any particular state S_π is equal to the exponential function of the goodness of the state scaled by the temperature, divided by the sum of corresponding quantities for all possible states (indexed by π'), including state π:

$$p(S_\pi) = \frac{e^{G_\pi/T}}{\sum_{\pi'} e^{G_\pi\prime/T}}$$

Here, a possible state is any state in which all the clamped units have their clamped values; each such state is one of the possible patterns of binary activation values over all of the remaining, unclamped units in the network. Since the sum over all the states in the denominator is the same no matter which state we are considering, we can express this relationship by saying that the probability of a state is proportional to the exponential function of the goodness of the state scaled by the temperature*:*

$$p(S_\pi) \propto e^{G_\pi/T}.$$

5.2. *Generative model of the knowledge embodied in the IA model*

The multinominal interactive activation model encodes specific probabilistic constraints in the biases and connections among units in a slight variant of the Boltzmann machine. Our next step is to define the probabilistic knowledge that we will be encoding in the network. We adopt a specific hypothetical formulation of the probabilistic knowledge that might underlie a perceiver's (implicit) beliefs about the process that might produce the arrays of visual input features in a letter perception experiment. This knowledge has the form of a *probabilistic generative model*. The concept of a generative model is a useful tool for characterizing the probabilistic structure of an environment and of the information reaching the sensory surface from the environment, and also as a hypothetical abstract characterization of the knowledge a perceiver uses in performing perceptual inference. Although the phrase was not used to describe it, a simple generative model lies at the heart of signal detection theory (Green & Swets, 1966): According to this theory, perceivers are thought to receive signals selected from either a signal plus noise distribution or a noise alone distribution. The parameters of the model are the probabilities of signal plus noise versus noise alone, and the means and standard deviations of each of the two distributions. Signal detection theory provides a theory of optimal perceptual inference in this situation. The generative model we offer here for letter displays is a bit more elaborate, but similar in spirit. It is very similar to the formulation of the beliefs about the probabilistic structure of letter displays used in the model of Rumelhart and Siple (1974), although these authors did not use this terminology.

According to our generative model, the feature array that reaches a perceiver's eye is generated by first selecting a word w_i at random from the possible words in a target lexicon (here, a set of English words that are all four letters long), with a probability $p(w_i)$ monotonically related to the word's language frequency. Once a word is selected, a sequence of letters is generated probabilistically based on the word. The probability of generating letter j in position k given that word i was selected is represented $p(l_{jk}|w_i)$. With high probability (assumed to be .9 in our simulations), the letter in each position is the correct letter for the given word, but there is a small probability that one of the other letters of the alphabet may be generated instead (given that the correct letter's probability is .9, the probability of each of the other letters is .1/25, or .004). Letters, in turn, give rise to a specification of presence or absence values for each of a set of possible letter features treated (following Rumelhart & Siple, 1974) as line segments (Fig. 5). For example, the letter T specifies that line segments should be *present* across the top of the corresponding feature array and down the middle of the array, and that other possible line segments that could occur in a feature array should be *absent*. Generation of feature values from letters and/or their registration by the perceptual system is also treated as probabilistic. Specifically, for a given letter position k, the probability of generating value v (which can be *present* or *absent*) for feature dimension f given letter j is represented $p(v_{fk}|l_{jk})$. The probability of generating the correct value of a given feature is relatively high (.9 in our simulations), and the probability of generating the incorrect value is equal to one minus this high value (.1).

Given the generative model above, it is possible to calculate the probability of every possible *path* through the generative model, where a path consists of a choice of one word, a choice of one letter in each position in the word, and a choice of one value (*present* or *absent*) for each feature in each letter position. We use the notation P_π to represent a particular path, using the same subscript π that we used previously for the states of a Boltzmann machine. This usage is appropriate, since patterns of activation in the MIA model will correspond to paths through the generative model.

The probability of a particular path P_π, represented $p(P_\pi)$, is simply the product of the probabilities of each of the individual probabilistic events assumed to underlie the creation of the path according to the generative model:

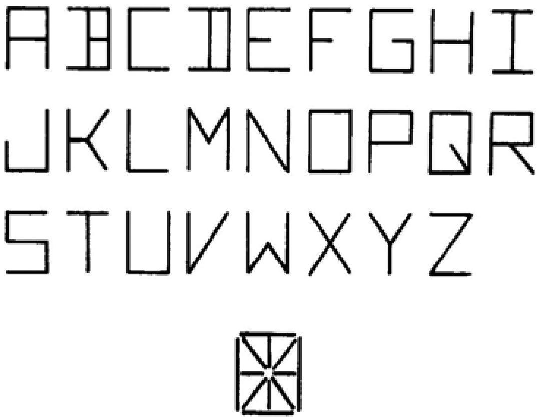

Fig. 5. The letters A–Z as they are represented in the Rumelhart & Siple font, with the full set of features shown in a single block below the letters. From fig. 2, p. 101 in Rumelhart and Siple (1974). Copyright © American Psychological Association. Reprinted with permission.

$$p(P_\pi) = p(w_i) \prod_k \left(p(l_{jk}|w_i) \prod_f (v_{fk}|l_{jk}) \right).$$

5.3. Perceptual inference under the generative model

The problem of perceptual inference (for our case) is to take a set of specified feature values $\{V\}$ and infer which of the possible paths consistent with this set of feature values gave rise to it. The possible paths are all the paths that have the given set of specified feature values. There is one such path for each combination of one word and one letter in each position (in the model, there are 1,179 possible words, and 26 possible letters per position, for $1{,}179 \times 26^4$, or approximately 540 million such paths). The probability of path π given the specified feature values, represented as $p(P_\pi|\{V\})$, is called the posterior probability of the path. The posterior probability of path P_π is given by

$$p(P_\pi|\{V\}) = p(P_\pi) / \sum\nolimits_{\pi'} p(P_{\pi'}),$$

where the summation in the denominator runs over all possible paths consistent with the specified feature values $\{V\}$.

In principle, we could calculate the probability of each such path, given the set of observed features, and choose the one that is most likely to have generated the observed features under the generative model. The multinomial IA model does not carry out this explicit calculation. Instead, the model *samples* from the set of possible activation states S_π, corresponding to possible paths through the generative model. While the model does not always sample the most probable state, it has the following property: The more probable a state is under the generative model, the more likely the state is to be sampled. We shall make this statement more precise below.

5.4. The MIA model: Using interactive activation to sample from the posterior distribution of the generative model

We now describe the MIA model and explain how it can sample from the correct posterior probability distribution over alternative possible interpretations of the set of specified feature values produced by the generative process above, where an interpretation corresponds to a path, specifying one word and one letter in each position.

As in the original IA model, the model (shown in Fig. 2) contains a unit for each possible word; a unit for each possible letter in each of four positions; and a unit for each possible value (*present* or *absent*) of each feature (e.g., horizontal across the top) of each of the four input feature arrays.[3] Units are organized into pools corresponding to sets of mutually exclusive alternatives. One pool consists of the set of units corresponding to the possible words and four other pools correspond to the sets of units for each of the possible letters in each of the four-letter positions. There are also four sets of 14 pools of units at the feature level: Each of these pools contains a "present" and an "absent" unit for a specific feature in a specific letter position.

The MIA model replaces the original model's pair-wise inhibitory connections between units in the same pool with the constraint that only one unit in a pool can be active at one time. Under this constraint, each pool now corresponds to a multinomial random variable—a variable that can take one of n alternative values, where n corresponds to the number of units in the pool. This is the feature of the model that gives rise to the word "multinomial" in its name. (Dean, 2005 proposed such a scheme in his computational model of neocortex; see also Lee & Mumford, 2003). Like the mutual inhibition assumption in the original model, the mutual exclusivity assumption in the MIA model is considered to be an idealized, conceptual-level consequence of the local inhibitory circuitry found throughout the brain; it plays a role similar to the role of the mutual inhibition between units in the same pool in the original model. This way of treating inhibition is similar to the divisive normalization model proposed by many modelers (e.g., Grossberg, 1978) and used by neuroscientists to model neural responses in visual cortex (Heeger, 1992).

In the MIA model, the probabilistic information that characterizes the generative model described above is used explicitly to set the bias terms and connection weights of the network. For reasons discussed below, the biases and weights correspond to logarithms of the relevant probabilistic quantities. Specifically, bias weights are assigned to each word unit. The value of the bias weight b_i on the unit for word i is set equal to $\ln(p(w_i))$, that is, the natural logarithm of the probability that word i would be sampled by the generative process described above (in what follows, the word "logarithm" always refers to the natural logarithm). The connection weight between each word unit w_i and each letter unit l_{jk} for letter j in position k is set to $\ln(p(l_{jk}|w_i))$, the logarithm of the probability that the letter would be generated given that word i was the word selected by the generative process. Similarly, the connection weight between the unit for letter j in position k and the feature unit for each of the two possible values of feature f in that position is set to $\ln(p(v_{fk}|l_{jk}))$, the logarithm of the probability that the feature would be generated under the generative model, given that the letter had been generated.

In summary, the MIA model embodies in its connection weights a logarithmic transformation of the probabilistic information in the generative model described above. If the model's knowledge exactly corresponded to the logs of the probabilities in a generative model that actually produced the displays used in a particular experiment, its outputs could be related to the true probabilities of events in the world that generated these inputs. Alternatively, we can think of the model as representing subjective estimates of these probabilities as they are employed by perceivers. In that case, to the extent that there are differences between the knowledge embedded in perceivers' perceptual systems and the true statistics of the world, perception that would be optimal with respect to the estimates might be non-optimal with respect to the statistics of the real world.

For the sake of our present goal of demonstrating that the multinomial IA model can sample from the posterior of the probability distribution defined by the generative model, we consider a case in which the *present* or the *absent* values of a subset of the features of a presented letter string are specified by an external input. For the example in Fig. 6, none of the features in the first position were specified, whereas the features in the second, third, and fourth positions were the features of the letters O, O, and D, respectively. According to the generative model (bars labeled calculated probability in the figure), the letters that form words with the context (F, G, H, M, and W) are all fairly likely; differences among them mostly reflect differences in the values of $p(w)$ for the associated words (FOOD, GOOD, HOOD, MOOD and WOOD).[4]

5.5. Processing in the MIA model

As in the Boltzmann machine, feature specifications are presented to the model by turning on the unit corresponding to the value of each specified feature. Processing begins with feature units clamped as specified above, and with no units active in any of the letter pools or in the word pool. Processing takes place over a number of cycles, similar to the random updating process in the Boltzmann machine. However, in our case the cycle is

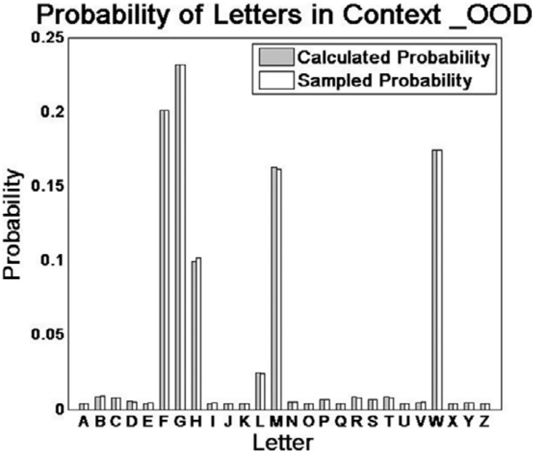

Fig. 6. Comparison of directly computed posterior probabilities and the results of the Gibbs sampling process in the multinomial interactive activation (IA) model, for letters in the first position of a four-letter display. The figure shows the calculated posterior probability of each possible letter in the first position of a four-letter array, following the presentation of a display in which no feature values are specified in the first position followed by full specification of features of the letters O, O, and D in the second, third, and fourth positions, respectively. The gray bars represent the calculated Bayesian posterior probabilities for the first letter position. These probabilities reflect the lexical knowledge embodied in the generative model. For this calculation, $p(l|w)$ was set to .9 for the correct letter in each position of the word, and .1/25 for each of the other possible letters, and $p(f|l)$ was assumed to be .9 for the correct value of each feature of each letter, and .1 for the incorrect value. The white bars represent the sampled probability that each of the letters in the first position was active after 50 iterations of the multinomial IA model. The weights in the model were set to correspond to the logarithms of the probabilities used for the Bayesian calculation, as described in the text. A total of 10,000 simulated trials were run for 100 iterations. Results are mean probabilities averaged over iterations 51–100. Slight differences between sampled and calculated probabilities are within the range of sampling error.

not random (although this detail is not critical for the functioning of the model, it makes discussion of the meaning of the computation somewhat simpler). Within a cycle, activations are first determined for each of the four-letter pools, using the existing activations at the feature and word levels; then activation is determined for the word pool, using the activations in each of the four-letter positions as well as the bias weights associated with each of the word units. Determination of activation in each pool begins by calculating each unit's net input, based on the weights, biases, and activations of other units as usual.

As previously stated, the model differs from the original IA model and indeed the original Boltzmann machine in that, at each time step, only one letter unit in each position and only one word unit may be active; the active unit is chosen probabilistically using the *softmax* function, so that, for each unit within the pool, the probability that a unit is chosen depends on the exponential function of its own net input divided by the corresponding quantities for all the units in the pool (itself included):

$$p(a_i = 1) = \frac{e^{net_i/T}}{\sum\limits_{i'} e^{net_{i'}/T}}.$$

Here, i and i' index the units in the pool being updated and T corresponds to temperature as in the Boltzmann machine. The softmax function can be viewed as an extension of the logistic function used in the Boltzmann machine, where the logistic function sets the activation of a single unit into either the *on* or the *off* state, while the softmax function sets a multinomial random variable into one of its n alternative states, in which exactly one of the units in the pool is active.

Let us now consider the relationship between this computation and sampling from the posterior probabilities of possible letters, given a set of observed features. For specificity, consider the computation of the activation for the pool of units corresponding to the letter in the second position of a four-letter display, on the first cycle of activation, when no units are active at the word level. In this case, the sending units are units corresponding to values of features in the second letter position, the receiving units are the units for possible letters in the second position of the string, and the weights are the connection weights between the letter and feature units, each of which corresponds to the log of the probability of the particular value of the feature (present or absent) given the letter. Noting that the activation of a sending unit is equal to 1 for the unit corresponding to the specified value v of feature f, and that there are no bias terms specified at the letter level in the model, the expression for the net input to letter unit j in position k can be rewritten as

$$net_{jk} = \sum_f \ln(p(v_{fk}|l_{jk})).$$

Now, when we compute $e^{net_{jk}}$ for use in the softmax function to compute the probability of activating the unit, this expression turns into $\prod_f p(v_{fk}|l_{jk})$, the probability that we would have generated the observed values of the features from the given letter, under the generative model.[5] Plugging these values into the softmax function, we see that it is equivalent to:

$$p(a_{jk}=1)=\frac{\left(\prod_f p(v_{fk}|l_{jk})\right)^{1/T}}{\sum_{j'}\left(\prod_f p(v_{fk}|l_{j'k})\right)^{1/T}}$$

For the case where $T = 1$, this equation corresponds to Bayes' formula for the posterior probability of letter j, given the values of the features (McClelland, 2013).[6] In that case, the softmax function will choose a letter to activate with a probability equal to the posterior probability of the letter given the specified features. If T is unequal to 1, these probabilities will be taken to the $1/T$ power, then renormalized. As stated before, we can express this more compactly as

$$p(a_{jk}=1)\propto\left(\prod_f p(v_{fk}|l_{jk})\right)^{1/T}$$

5.5.1. The roles of logs and exponentials in linking neural and probabilistic computation

The reader may be tempted to ask at this point why we have bothered with using the logarithms of probabilistic quantities in defining the strengths of the connection weights in the MIA model network, since we then proceed to undo this logarithmic transformation when we exponentiate the net input to a unit for use in the softmax function (see note 5) or the closely related logistic function. Indeed, it would be possible to reformulate the MIA model, directly using the prior probabilities of words and the conditional probabilities of letters given words and of features given letters, and then redefining the activation function accordingly. The reason for using the logs of these probabilistic quantities is based ultimately on the inspiration from neuroscience that continues to lie behind the MIA model and other neural network models, and on the previous history of models linking neurons to computation. The MIA model traces its lineage through a marriage of the original IA model, a descendent of an earlier model of Grossberg (1978), with the Boltzmann machine, a descendent of the earlier model of Hopfield (1982). Ultimately, these models can in turn be traced back through the Perceptron (Rosenblatt, 1958) to the McCullough-Pitts neuron (Pitts & McCullough, 1947), a device that added up weighted signals and compared them to a threshold. The idea that neurons additively combine excitatory and inhibitory signals, and then fire when a threshold is reached, is, or course, the standard intuitive simplification of a model neuron relied on by neuroscientists. In the presence of a source of additive Gaussian noise in the inputs to such a simplified model neuron, the probability of firing will closely match the logistic function of the summed or net input. Thus, the McCullough-Pitts neuron with noise added to its input turns out to be a closely approximate implementation of the logistic neuron used in Boltzmann machines, which in turn implements Bayes' rule if the weights and bias terms are set to the logs of the appropriate probabilistic quantities, as Hinton and Sejnowski (1983) were the first to point out (see McClelland, 2013, for further discussion of a possible neural basis for the softmax function).

Returning to the main thread of our development, we now consider the net input to each unit at the word level. In this case, the net input consists of the bias term representing the log of the subjective probability of the word, plus the sum of terms corresponding to the product of the activation of each letter level unit, times the weight between the word unit and the letter unit. From the first step in the computation described above, one letter unit in each position has an activation value of 1, while all other letter units' activation values are 0, so the net input to word unit i becomes

$$net_i = \ln(p(w_i)) + \sum_k \ln(p(l_{jk}|w_i))$$

where l_{jk} represents the active letter unit in position k. Now, computing e^{net_i}, we obtain the probability, under the generative model, that the word would be chosen for presentation, times the probability that the active letters would have been generated, given that the word had been chosen. Putting this into the softmax function, we obtain

$$p\,(a_i = 1) \propto \left(p(w_i)\prod\nolimits_k p(l_{jk}|w_i)\right)^{1/T}$$

Expressing this in words, the probability that a given word unit is chosen to be the only one active is proportional to the prior probability of occurrence of the word, times the probability that the word would have generated the set of active letters. Again, this implements the basic logic of Bayes rule for calculating a posterior probability that a particular word was presented, in this case given prior information (represented by $p(w_i)$) and the likelihood of evidence (in this case the active letters) given the word.

Finally, let us consider the activation of a unit j in any one of the letter pools on the next cycle, when there is a single-word unit active at the word level. The net input to each letter level unit is the same as before, but with an extra term corresponding to the log of the probability of the letter, given the active word. Once this expression is exponentiated, it corresponds to the probability of the letter given the active word, times the probability of the set of specified features, given the letters. The expression for the probability that a given letter j will be activated in position k is

$$p(a_{jk} = 1) \propto \left(p(l_{jk}|w_i)\prod\nolimits_f p(v_{fk}|l_{jk})\right)^{1/T}$$

Thus, after the second update of letter level activations, the probability that a given letter unit in each position is chosen to be the active unit in that position is proportional to the probability of the letter, given the active word, times the probability of the set of features in the given position, given the letter, scaled by $1/T$.

Note that the weights between word and letter units and between letter and feature units were defined in terms of the top-down, generative process that is treated as underlying the creation of the displays. The letter-to-feature weights are used in computing bottom-up input from feature to letter units and the word-to-letter weights are used in computing the bottom-up input from the letter to the word units. The word-to-letter weights are also used to compute the top-down influences from the word units to the letter units, and, although we do not consider it here, the letter-to-feature weights could be used to fill in missing feature-level activations. Thus, the same connection weight values are used symmetrically, in both directions, even though their values are those specified in the top-down generative model. Because the weights are used symmetrically, the model shares an essential characteristic with the Boltzmann machine: The activation updates tend to move the states of the network in the direction of states of higher overall goodness.

In summary, given the order of processing specified above, and running with $T = 1$, the probability that a given letter unit will be active in a given position will correspond to the probability of the letter given the features under the generative model. When the word level is first updated, a single word will be chosen with a probability proportional to the probability of the word given the chosen letters. Thus, our calculation will produce a sample from the possible states of the underlying generative model that could have produced the observed features. However, our estimates of the probabilities of the letters have not yet taken the word-level information into account. The next update at the letter level does take the word-level information into account, so that, for each letter position, the probability that a letter unit will be active is equal to the probability of the letter, given both the active word and the given array of features.

It might seem that the computation is complete at this point, but the probabilities of letter activations after the second update at the letter level do not exactly match their correct posterior probabilities. However, as the sampling progresses through additional cycles alternating between updates and the word and letter levels, the activation probabilities converge toward the correct posterior probabilities. The sampling procedure is a generalization to the multinomial case of the procedure used in the Boltzmann machine to set activation states. Like the Boltzmann machine sampling procedure, our procedure is an instance of *Gibbs sampling* (Geman & Geman, 1984), a widely used procedure that originated in statistical physics, where it has been shown to provide unbiased samples from the posterior of a probability distribution by making local updates of individual variables consistent with the conditional distribution of these individual variables given the current values of other variables (see McClelland, 2013, for details). This is exactly what we are doing in the MIA: We are sampling states of the letter units, conditional on states of the word and feature units; and we are sampling states of the word units, conditional on states of the letter units and feature units (although the feature units only affect the word units indirectly, via the states of the letter units).

5.6. Probabilities of states of the MIA model and pathways through the generative model

If we sample states of the MIA model at some temperature T, the probability that we will be in a given state after an initial "burn-in" period is equal to $e^{G(s_\pi)/T}$, where the goodness is defined as it was above. For the specific case of the MIA model, the goodness becomes

$$G(S_\pi) = \ln(p(w_i)) + \sum_k \left(\ln p(l_{jk}|w_i) + \sum_f \ln p(v_{fk}|l_{jk}) \right)$$

exponentiating this expression, we obtain:

$$e^{G(s_\pi)} = p(w_i) \prod_k \left(p(l_{jk}|w_i) \prod_f (v_{fk}|l_{jk}) \right)$$

The expression on the right is the probability, under the generative model, that the path through the generative model underlying the observed set of features is the one that correspond to state S_π. Plugging this into the probability-goodness equation, we see that the model visits such states with probability proportional to the temperature-scaled probability that they actually generated the observed features:

$$p(s_\pi) \propto \left(p(w_i) \prod\nolimits_k \left(p(l_{jk}|w_i) \prod\nolimits_f p(v_{fk}|l_{jk}) \right) \right)^{1/T}$$

or more simply

$$p(S_\pi) \propto p(P_\pi|\{V\})^{1/T}$$

The temperature parameter T has both an overt and a covert role in the behavior of the model. Overtly, when T is very high, all states become equiprobable, whereas when T becomes very low, only the states with the highest posterior probability have any appreciable chance of being sampled by the network after the "burn-in" period. However, if the network is run at a very low temperature, the burn-in period becomes exceedingly long. The approach initially suggested for the Boltzmann machine was to use simulated annealing, whereby T starts high and is gradually reduced. Instead of this, in the simulations we have conducted with the MIA model, we have run the model at a fixed temperature $T = 1$. In this case, we have found that the network achieves the correct posterior probability distribution in less than 20 cycles, and the approximation is quite good within about 10 cycles.

5.7. Making overt responses based on the state of the model

The development thus far shows how an interactive neural network can sample from the posterior of the probability distribution over entire states of a neural network. These states are samples from the joint distribution of assignments of both letter and word identities that could have given rise to the actual features present in the network's input. Should we be interested in determining the identity of a particular item—say, the letter in a given position, as in many visual word recognition studies, or the whole word, as in many other studies—we can observe that the probability of being in a state where the unit in question is active (regardless of the activations of other units) corresponds to the correct posterior probability of the item. In other words, the network's states are simultaneously samples from the marginal distribution of each of the multinomial variables and the joint distribution of all of these variables. This is exactly what Rumelhart (1977) envisioned as the outcome of interactive processing in perception.

To generate a response that is a sample from this distribution, say about the identity of the letter in the first position of a word, a perceiver would only need to report the identity of the letter that had been selected through the iterative settling process. Simulations of the model verify this mathematical fact; one example illustrating this is shown in Fig. 6 (see caption for further explanation).

5.8. Sampling as an approximation to optimality

We have described a model in which perception involves sampling from the posterior of the generative model characterizing the stimuli presented to the perceptual system. It should be noted here that the truly optimal policy would be to choose the alternative with the highest posterior probability, rather than sample alternatives in proportion to their relative probability, the policy we follow in the model by setting the temperature parameter T to 1[7]. Alternatively, however, we can see the temperature parameter as reflecting intrinsic processing noise in the perceptual system. In that case, we can see each trial in a perceptual experiment as an attempt to find the single best interpretation subject to the prevailing level of noise. In either case, the higher the temperature, the more random behavior will be. The advantage of higher temperature is that it allows fuller exploration of the range of possible perceptual interpretations and avoids premature commitment early in a computation.

In Boltzmann machines, optimal perceptual interpretation is made possible by gradually reducing temperature, but this policy is only guaranteed to find a global optimum after an infinite time. In view of the real-time constraint, sampling at a fixed intermediate temperature may be the compromise the brain adopts as its approximation to optimal perceptual inference in real time.

5.9. Perceptual facilitation in non-words in the MIA model

As we noted earlier, an important feature of the original IA model was the fact that it accounted for the facilitation of perception of letters in pseudowords, such as MAVE, as well as for facilitation of perception of letters in words. In the original model, this occurred because a non-word could partially activate several words that shared letters in common with the string presented. At first glance, it might be supposed that the MIA model would not show the same pattern, since only one word is active at a given time. To explore this, we considered the ambiguous displays in Fig. 7, where the letter in the second position is partially occluded but occurs either in a word, in a pseudoword, or by itself. The available features are equally consistent with the letters A and H in the Rumelhart-Siple font used to represent letters in the simulation. Can the model successfully use context to resolve the ambiguity, selecting A as the more likely alternative, even if the ambiguous segment occurs in a pseudoword context?

To address this question, simulations with each of the three displays shown in the figure were conducted. For the single letter alone case, the word level was switched completely off, as a baseline for assessing the influence of the word level in the other two contexts. The results are shown in Fig. 7. As the figure indicates, in the absence of context (white bars), the alternatives A and H are both chosen about half of the time, since the feature values specified are maximally consistent with both of these alternatives. With either context, the letter A becomes far more likely than the letter H. This occurs to a greater extent when the first position contains a C than when it contains an M, but it occurs to a considerable extent in both cases.

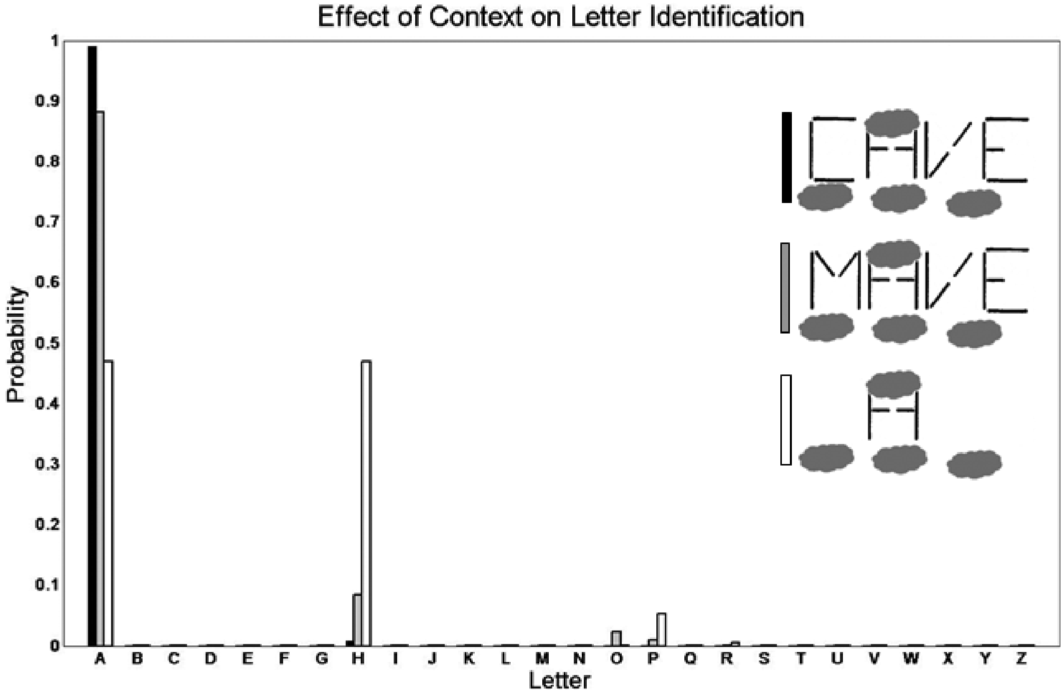

Fig. 7. Probability that different letters are activated in the second letter position when an ambiguous character equally consistent with A or H is presented in different contexts (black bar: C_VE; gray bar: M_VE; white bar: no context). In the generative model based on the Rumelhart-Siple font, both A and H are equally likely to generate the features shown, and the letter P is next most likely. But when the context is C_VE or M_VE, A is far more likely. The M_VE context supports the letter O to some degree, but the feature information is unlikely under the hypothesis that the letter is O, so overall O is much less likely than A.

Why does the model tend to choose the letter A in both contexts? When the rest of the letters form the word CAVE, the entire display is far more likely to have been generated by CAVE than any other word, and thus the letter A is far more likely to have been the letter in the second position than the letter H. When the first letter is M, no single word is highly likely to have generated all of the observed features. In fact, the word MOVE is overall more likely than any other single word (although it is inconsistent with some of the features in position 2, it is consistent will all of the features in all of the other positions). However, many other words, including CAVE, GAVE, HAVE, SAVE, and WAVE, as well as MADE, MAKE, MALE, MARE, and MATE, are all partially consistent with the full set of features. Each of the words listed is sometimes sampled at the word level; when MOVE is sampled, the model can choose O as the letter in the second position, but it can also choose A or H, since these letters can occasionally be generated according to the generative model when the underlying word was MOVE. When one of the words containing A in the second position is sampled, it almost always chooses A as the corresponding letter.[8]

5.10. The MIA model exhibits logistic additivity, addressing a limitation of the original IA model

We have seen that, in the multinomial IA model, if settling occurs at a fixed temperature $T = 1$, exact matching of posterior probabilities according to our generative model can be obtained. Do human perceivers also match these posterior probabilities? Since it is hard to obtain independent evidence of the subjective probabilities involved, the tendency has been to determine whether or not perceivers are combining context and stimulus information according to the functional form we would expect if they were performing optimally. Interestingly, there is a simple functional form that arises in the multinomial IA model and other stochastic variants of the IA model for the way in which a factorial manipulation of stimulus and context information should affect the probability of choosing a particular alternative (McClelland, 2013; Movellan & McClelland, 2001): It is easy to show (for a subset of these models, including the multinomial IA model) that the *logit* of the probability of making a particular response (where logit(p) is defined as ln (p/(1−p)) a quantity also known as the log-odds) should correspond to a sum of two quantities, one due only to the stimulus (corresponding to the relative probability of the sampled features given the item) and another due only to the context (corresponding to the relative probability of the item given the context).

$$\text{logit}(p_i) = s_i + c_i$$

An additional term b_i can be included to incorporate a bias associated with the alternative's prior probability. This relationship (which Movellan and McClelland called *logistic additivity*) holds at least approximately in the data from many studies investigating the joint effects of context and stimulus information (see Movellan & McClelland, 2001 for a review; see Pitt, 1995 for one exception). The multinomial IA model exhibits logistic additivity, and its tendency to do so is unaffected by the value of the temperature parameter (T): T can be thought of as scaling the magnitudes of the stimulus and context terms in the model's predictions, but it is not in general separately identifiable from the data.

As Massaro (1989) noted in his early critique of the original IA and TRACE models, these models did not capture the logistic additivity seen in the data from many experiments, and this failure was the basis for his conclusion that interactivity fundamentally distorts perception; similar concerns have contributed to the criticisms offered by Norris et al. (2000) and Norris and McQueen (2008). While the original model's assumptions did distort this relationship, the problem was not in fact due to interactivity: As mentioned above, the influence of multiple sources of input failed to exhibit logistic additivity under the activation functions used in the original models even when propagation of activation was strictly feed forward (McClelland, 1991). In any case, logistic additivity is observed in the MIA model, overcoming this limitation of the original model.

It is important to note that logistic additivity is observed in a number of other variants of the IA model (McClelland, 1991, 1998; Movellan & McClelland, 2001); in particular, it is not necessary to assume the unit activations are binary. Although the result is harder to prove mathematically for such cases, it has been demonstrated to hold in simulations. The variants that exhibit logistic additivity incorporate variability in the input to the model and/or intrinsic to processing within the model.

5.11. Interim summary

It is hoped that the exposition of the MIA model makes clear that interactive activation produces a good approximation to optimal perceptual interpretation in real time, in accordance with the IA hypothesis, and that the MIA model (along with other variants of the IA model) can capture the logistic additivity pattern seen in data. This does not mean, of course, that the MIA model is the best possible model of human perceptual processing or even that interactivity is a part of the process of perception. Indeed, critics have argued that interactivity is not necessary to achieve a good approximation to optimality, leading them to argue for models in which processing is unidirectional. We now turn to consider this issue.

6. Is it advantageous for influences to feed back into the perceptual system?

A number of authors have proposed that context effects on letter or phoneme identification can be adequately explained by relying only on feed-forward processing, with integration of stimulus and contextual information occurring at a subsequent, decision stage (e.g., Massaro, 1989; Norris & McQueen, 2008; Norris et al., 2000; Paap, Newsome, McDonald, & Schvaneveldt, 1982). A post-perceptual decision level that integrates perceptual and contextual information can explain how stimulus and lexical information affect letter or phoneme identification (Fig. 8a). Thus, these authors have argued, interactive activation is of no benefit, and it need not be incorporated into models of perception.

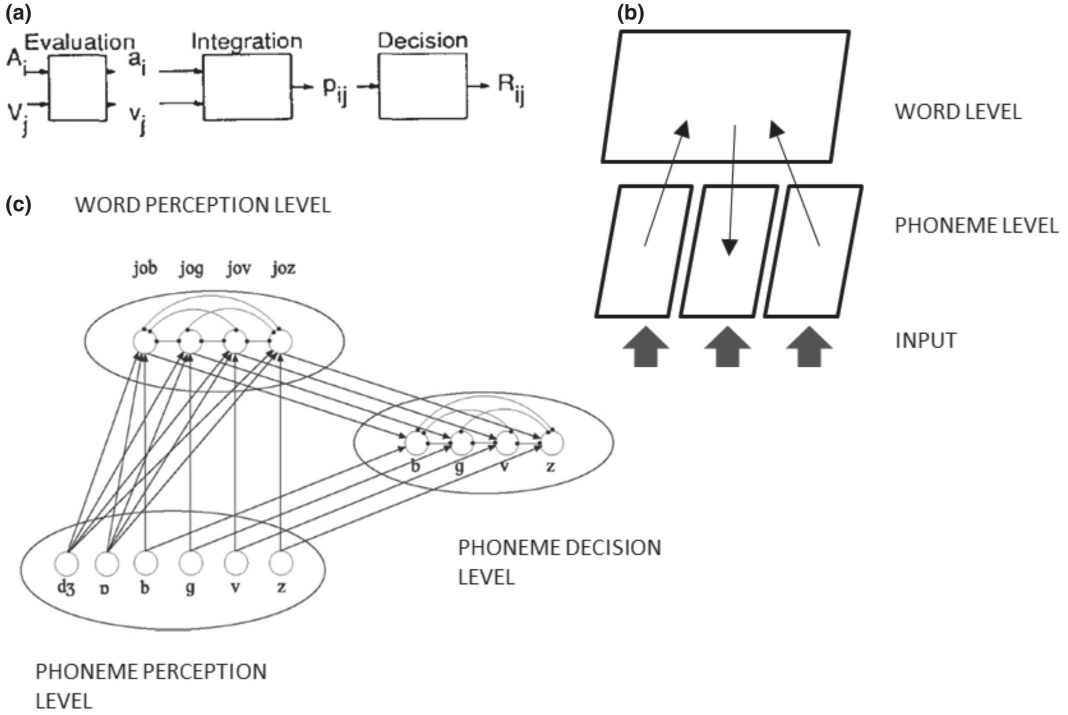

Fig. 8. (a) Massaro's schematic representation of the integration of stimulus and context information according to his Fuzzy Logical Model of Perception, reprinted from fig. 1, p. 401 of Massaro (1989). Copyright © Elsevier Ltd, reprinted with permission. The A and V variables in Massaro's figure correspond to the stimulus and context variables presented in the text. (b) Schematic diagram indicating a unidirectional propagation of information for computing the contextual and stimulus factors used in Massaro's model for the identification of the segment in the middle position of a three-phoneme syllable. (c) The architecture of the MERGE model of speech perception (Norris et al., 2000), reprinted from fig. 11, p. 384 of Norris and McQueen (2008). Copyright © American Psychological Association, reprinted with permission.

We argue that there are two important ways in which interactive activation can be beneficial:

1. It implements optimal perceptual identification over many representational levels and at many positions within a level at the same time.
2. It allows the consequences of these processes to be available inside the perceptual system itself, thereby allowing for the possibility of knock-on consequences for processing of other inputs or for processing the same item on later occasions.

We consider these points in the next two sections.

6.1. Implementing optimal inference over many levels and positions simultaneously

To underscore the first advantage of an interactive approach, we contrast it with the approach proposed by Massaro (1989), who has advocated strictly feed-forward computation for the integration of context and stimulus information in perception. Similar points apply to the approach favored by Norris and collaborators (Norris & McQueen, 2008; Norris et al., 2000), as we shall discuss below.

Massaro's model focuses on the perceptual identification of a single specified target item. For example, in one experiment (Massaro & Cohen, 1983) of the type considered by Massaro, the target item was the second speech sound in a monosyllable beginning with either /t/, /s/, /p/, or /v/, and ending in the vowel /i/ ('ee'). Seven different sounds were presented in each context, between the initial consonant and the vowel, organized on a continuum from /l/-like to /r/-like, for a total of 28 distinct stimuli in all. Each stimulus was presented many times to each participant, with the task of identifying the second segment as either /l/ or /r/.

From a Bayesian point of view, one could propose that perception depends on calculating an estimate of the posterior probability that a given input is /r/ or /l/, using both stimulus and context as sources of constraining information. This can be done by calculating, for each context *c,* the quantities $p(r|c)$ and $p(l|c)$; and also by calculating for each stimulus *s*, the quantities $p(s|r)$ and $p(s|l)$. The correct posterior for $p(r|s,c)$ is then given by:

$$p(r|s,c) = \frac{p(s|r)p(r|c)}{p(s|r)p(r|c) + p(s|l)p(l|c)}$$

Massaro's model (Fig. 8a) assumes that participants calculate quantities corresponding to normalized estimates of the probabilistic quantities in the above formulation.[9] Notably, the representation of context used in the calculation described above excludes the stimulus information from the second segment; the first and last segments specify the context, while the second provides the stimulus information used in the calculation.

The information encoded in the connection weights in a three-phoneme-slot processing system could be used to calculate the terms needed for Massaro's model, although we would then be using this knowledge in a feed-forward, rather than an interactive fashion (Fig. 8b; although arrows go up and down in this figure, each arrow goes in only one direction, and there are no feedback connections). The featural information in the first and last positions would be used to calculate $p(p|f)$ for each possible phoneme in the first and last positional slots; then, at the word level, one can calculate $p(w|\{p_1\},\{p_3\})$ for each word in the lexicon, relying as before on the assumptions of our generative model (the expression $\{p_1\}$ denotes the vector of $p(p|f)$ values for all possible phonemes in position 1, and similarly for $\{p_3\}$). The quantity $p(r|c)$ can now be calculated as the sum over all words of the probability of the word given the input in the first and last position, times the probability of r in the second position given each word, and similarly for $p(l|c)$. This corresponds to using the connection weights between the phoneme and word units in one direction in the first and last position, and in the opposite direction in the second position, as illustrated in the figure. The desired quantity $p(r|s,c)$ is then calculated by combining the lexical input with the bottom-up stimulus support calculated for the phonemes in the second position and then using the above equation. This calculated probability is then used to generate the *r* response with probability $p(r|c,s)$ or the *l* response with probability

$p(l|c,s) = 1 - p(r|c,s)$. An alternative, sampling-based approach that would produce *r* responses with the same probability would proceed by selecting a single phoneme in positions 1 and 3, based on the feature input to these positions; then selecting a single word based only on these phonemes; then selecting between *r* and *l* for the middle position based on the selected word and the feature input in the second position. In either case, the calculations are unidirectional, and contextual and stimulus support for the target item are calculated separately, as Massaro's model proscribes.

We can now contrast Massaro's feed-forward proposal with the interactive activation approach, in which a bidirectional computation is applied across all positions, as previously described. In Massaro's model, the computations outlined above are only valid for calculating the posterior probability of the phoneme in the second position. This may not seem problematic when considering the experimental paradigm used by Massaro and Cohen (1983), where the target was always the phoneme in the second position (See Fig. 8b). However, in most experiments on the perception of letters in words, including the experiments of Reicher (1969), Massaro and Klitzke (1979), and nearly all of the experiments addressed by the IA model, participants are not cued prior to the trial on which letter will be the target letter. For these cases, the multinomial IA model simultaneously samples from the correct Bayesian posterior in all four positions. Furthermore, the MIA model uses the same representation at the word level both as its sample from the distribution of possible words and as the basis for constraining perception of each possible letter. For Massaro's model, the context representation for each position excludes the bottom-up information from that position, and thus is an incomplete representation of the information relevant to the identification of the word. In short, for an input containing three letters, four different word-level quantities are needed, one for word level, and one for each letter position.[10]

Feed-forward computation in MERGE and related models. An approach very similar to Massaro's is advocated by Norris and colleagues in their models of perceptual processing of words and letters or phonemes (Norris & McQueen, 2008; Norris et al., 2000). Just as in Massaro's model, the correct feed-forward calculation of the necessary top-down constraints for each letter or phoneme is different for each item at a lower level (e.g., for phonemes in each position, the lexical context must be based on the phonemes in all other positions). In particular, when considering the role of context on identification of a target segment (e.g., the effect of the first two segments in *job* on the identification of the final segment, see Fig. 8c), bottom-up information about the target segment is not allowed to affect values at the word perceptual level until after the top-down influence from the first two segments has been combined with the target segment information in the phoneme decision layer (D. Norris, personal communication, July 2011). This would be difficult to implement, since information about speech segments overlaps in the spoken input. The difficulty is compounded when we consider the effects of subsequent context, as in the classic experiment of Ganong (1980), where the target segment is the first segment in a word context – a /g/ or /k/ followed by "iss" or "ift," or in experiments where disambiguating context occurs in a subsequent word (Warren & Warren, 1971). To explain this effect, segments subsequent to the target segment must be allowed to affect the word level, but the target segment must be prevented from doing so. In interactive models, this complication is unnecessary. Context phonemes in all positions can affect processing of each phoneme in each position simultaneously, with decisions about each being updated as information becomes available, either about prior or subsequent elements of the input.

In summary, non-interactive models in the psychological literature have not addressed the simultaneous use of context and stimulus information at multiple levels and multiple positions within a level. They have tended to focus on joint use of context and stimulus information in identifying a specified target item at one level of processing, without dealing with the fact that in natural perceptual situations, the goal is to simultaneously interpret multiple items at many different levels of processing. In contrast, interactive models allow representations of alternatives at different levels and different positions within a level to mutually constrain each other in an integrated parallel, distributed, and interactive computation.

6.2. *Knock-on consequences of interactive processing*

We now consider the second advantage of interactive models over feed-forward models: Interactivity allows effects of context to affect subsequent processing within the perceptual system. Such effects include effects on processing of neighboring items present in the immediate context of a presented item, and effects on processing of similar inputs on subsequent occasions.

Knock-on consequences for neighboring input items. A case of the first type was considered by Elman and McClelland (1988). They focused on a phenomenon in speech perception known as compensation for coarticulation (Mann & Repp, 1981; Stephens & Holt, 2003): The perceptual system seems to compensate for the effects that articulation of one phoneme has on the acoustic realization of neighboring phonemes. For example, the lip formations associated with /s/ and /ʃ/ ("sh") persist into the articulation of subsequent stop consonants like /t/ and /k/, shifting the frequency content of the successor. Perceivers compensate for this, allowing more accurate recognition of the successor. Thus, when an ambiguous sound between /t/ and /k/ is preceded by /s/, it will tend to be heard as /k/; when preceded by /ʃ/, it will tend to be heard as /t/. In this situation, the presence of background noise or articulatory variability could obscure the identity of the preceding fricative sound, robbing a strictly feed-forward system of information to allow compensation. But if that fricative sound occurred in a lexically constraining context, and feedback were allowed to influence the activation of the contextually more likely fricative, compensation could nevertheless occur, improving identification of subsequent phonemes. Elman and McClelland (1984) included a mechanism for producing such compensatory effects in one version of the TRACE model, simulating the lexically mediated compensation for coarticulation effect.

Elman and McClelland (1988) subsequently designed an experiment to determine whether lexical context could trigger compensation for coarticulation, as the TRACE model predicted. They presented ambiguous /t/ or /k/ sounds preceded by an ambiguous fricative sound halfway between /s/ and /ʃ/. In turn, the ambiguous fricative was preceded by one of two different lexical contexts, one consistent with /s/ (e.g., "Christma_") and one consistent with /ʃ/ (e.g., "fooli_"). If lexical information feeds back to influence phoneme processing, then the ambiguous fricative in "Christma_" should behave like an acoustic /s/ and cause a shift in the perception of the following phoneme toward /k/. Conversely, the same ambiguous fricative in "fooli_" should behave like an acoustic /ʃ/ and cause a shift in the perception of the following phoneme toward /t/. This is precisely what Elman and McClelland found. Although this result has been questioned (Pitt & McQueen, 1998), it has been replicated in multiple different laboratories, and with different sets of materials (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; Samuel & Pitt, 2003). Those who favor non-interactive approaches have, however, presented recent evidence further contesting the source of the effect (McQueen, Jesse, & Norris, 2009), and research on the topic continues.

Knock-on consequences for processing similar inputs on subsequent occasions. Other researchers have explored other knock-on effects of lexical context on phoneme identification that are also predicted by the interactive account. One such effect has been demonstrated using selective adaptation, a domain-general phenomenon in which repeated presentation of a particular stimulus causes a perceptual shift such that neutral stimuli are perceived as being less like the repeatedly presented stimulus. In the case of speech perception, after repeated presentation of a phoneme (e.g., /s/), perception of an ambiguous phoneme (e.g., halfway between /s/ and /ʃ/) is shifted toward the alternative interpretation (in this case, /ʃ/; e.g., Samuel, 1986; Samuel & Kat, 1996). To demonstrate lexically mediated selective adaptation, a neutral sound (an ambiguous phoneme or a noise burst) was repeatedly presented in lexical contexts that were consistent with only one interpretation. If the neutral sound was presented in /s/-biased contexts such as "bronchiti_", "arthriti_", etc., the selectively adapted representation was /s/; if it was presented in /ʃ/-biased contexts such as "aboli_", "demoli_", etc., the selectively adapted representation was /ʃ/ (Samuel, 1997, 2001). Thus, the lexical information determined which sublexical representation was selectively adapted, influencing subsequent phoneme and word identification.

A third example of knock-on consequences of lexical feedback—one that was predicted in the McClelland and Elman (1986) TRACE model paper—is lexically guided tuning of speech sound categories. Such tuning is essential for listeners to be able to correctly identify different speakers' productions, since phoneme category boundaries vary between individuals. For example, speakers of English and Spanish center their /b/ and /p/ categories at different points along a dimension called voice onset time. Furthermore, regional dialects are often distinguished by differences in details of both consonant and vowel production. Lexical information provides a ready source of information for tuning speech perception in response to such shifts in speech sounds, and several studies beginning with Norris, McQueen, and Cutler (2003) now indicate that such tuning does in fact occur in speech perception (van Linden & Vroomen, 2007 showed an analogous shift in use of visual cues from the lips; for a review see Samuel & Kraljic, 2009). The pre-lexical locus of this effect is supported by evidence that the tuning effect generalizes to influence perception of words not used in the induction of the effect (McQueen, Cutler, & Norris, 2006). In TRACE, lexical information feeds back to influence pre-lexical phoneme unit activations, and Mirman, McClelland, and Holt (2006) augmented TRACE with a simple Hebbian learning rule to adjust the feature to phoneme connections, allowing it to simulate the relevant experimental findings.

More generally, Friston (2003; see also Spratling & Johnson, 2004) has argued that top-down feedback is necessary to learn the hierarchical representations that are found throughout perceptual and cognitive systems, and indeed some form of feedback is used in many different neural network learning algorithms. Proponents of autonomous/feed-forward accounts of perception acknowledge the necessity of feedback for learning but insist that this feedback is not equivalent to the "online" feedback that influences processing in interactive activation models (e.g., Norris et al., 2003). We argue that a system in which feedback can guide learning as well as perception provides a parsimonious account. Furthermore, if feedback guides learning, then the learned representations will necessarily reflect a combination of bottom-up and top-down information, making the representations themselves both consequences of and intrinsic to their roles in interactive processing.

In sum, feedback not only allows contextual constraints to determine the identity of elements (such as letters and phonemes) of larger units (such as words) but also allows the results of this contextually determined identification process to influence processing of neighboring elements (compensation for coarticulation) and subsequent occurrences of the same elements (adaptation, retuning). Knock-on consequences of feedback provide both motivation for and evidence of direct top-down feedback in perception.

7. Neural basis of interactive processing

7.1. Basic neuroscience findings

Evidence from research on the neural basis of perception supports the presence of interactive processing in the brain. Interactive processing is supported by a basic feature of brain architecture: Wherever in the neocortex there is a "forward" path from area A to area B there tends to be a strong (sometimes much stronger) return pathway (Felleman & van Essen, 1991). Many studies correspondingly show that reversible inactivation of putatively higher level or downstream cortical areas (e.g., higher level visual or auditory cortex) affects stimulus-driven activity in primary areas (e.g., Hupé et al., 1998; Carrasco & Lomber, 2010), implicating reciprocal interactions in cortical processing. Neural recordings in rhesus monkeys indicate that the same "edge detectors" in V1 that respond to physically present edges also respond to illusory edges in Kanizsa figures. The illusory contour response in V1 was found to occur later than the response in V2, suggesting that the response in V1 was due to feedback from higher level visual processing (Lee & Nguyen, 2001). Similarly, binocular rivalry appears to be a mutual constraint satisfaction/ interactive activation process with neurons in many different visual areas, from V1/V2 to inferotemporal cortical areas, showing consistency with the global percept (Leopold & Logothetis, 1999). Evidence of bidirectional propagation of activity between occipito-temporal and pre-frontal brain areas is also seen in human magneto-encephalography (MEG) studies of visual object recognition (e.g., Bar, 2004).

In addition to top-down feedback from higher levels within a processing modality, neurophysiological studies have shown cross-modal interactions between primary regions of perceptual processing (see Ghazanfar & Schroeder, 2006 for review). To us such mutual constraints between modalities are just as much examples of the fundamental principle of mutual constraint satisfaction as the bidirectional interactions between levels in a hierarchical perceptual system. Although several studies have argued that sensory integration occurs in secondary sensory or association cortex (Bavelier & Neville, 2002; Jones & Powell, 1970) or in frontal cortex (Rizzolatti, Riggio, Dascola, & Umlita, 1980), recent evidence has pointed to the presence of top-down inputs from these association regions to primary sensory cortices in audition (Cappe & Barone, 2005; Schroeder et al., 2001) and vision (Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003) as well as direct input from auditory cortex to primary visual cortex (Falchier et al., 2002; Hall & Lomber, 2008) and vice versa (Bizley & King, 2009). Physical projections from auditory cortex terminating in area V1 have also been observed in the monkey (Falchier et al., 2002; Rockland & Ojima, 2003) and in the adult cat (Hall & Lomber, 2008), suggesting that these connections are not limited to early developmental stages. In addition, evidence from multiunit recordings in the ferret has shown that roughly 20% of the neurons in area A1 respond to visual stimulation (Bizley & King, 2009).

Overall, a growing body of evidence is challenging the idea that there is encapsulation of sensory processing at the neural level (see Ghazanfar & Schroeder, 2006). Instead, the evidence suggests that a highly interactive biological system enables the simultaneous use of information across hierarchical levels from multiple modalities for spatial localization, communication, and various social behaviors (Lewkowicz & Ghazanfar, 2009). This interactive neural system implements cognitive processing that relies on the simultaneous, coherent engagement of representations at many levels and within many modalities at the same time—that is, processing that is distributed, parallel, and interactive.

7.2. Interactivity in the brain mechanisms of human language processing

Interactive processing has also been a key theme in research on human language processing and reading. Much of this work has been conducted within the framework of the "Triangle model" of single-word reading (Seidenberg & McClelland, 1989; and subsequent extensions), which can be viewed as a version of the interactive activation model that relies on learned distributed representations rather than localist representations of units at the orthographic, phonological, and semantic level. Here, we highlight the interactive processing aspects of the framework as illustrated in Fig. 9, focusing on the timing and locus of mutual influences of phonology and orthography and of lexical effects on phonological and orthographic processing. Note that in the triangle model framework, bidirectional connections throughout the model are sensitive to lexical knowledge as well as knowledge of the patterns of covariation between orthographic and phonological representations. Specifically, the presentation of a visual or spoken word form would induce bidirectional interactions among orthographic, phonological, and semantic representations, leading to the prediction that lexical knowledge and spelling-sound consistency would affect orthographic and phonological representations, at least in skilled readers, once the relevant connections had become strengthened through experience.

Discussions of the neural basis of visual word recognition have focused heavily on the role of a region of the left occipito-temporal cortex known as the Visual Word-Form Area (VWFA; McCandliss, Cohen, & Dehaene, 2003; Dehaene, Cohen, Sigman, & Vinckier, 2005). Some have argued that VWFA functions as an orthographic "input" lexicon, a repository for visual forms of words (Kronbichler et al., 2004, 2007), while others have contended that this region is prelexical in nature (Dehaene et al., 2005), with some possible hierarchical organization of orthographic representations in or near the VWFA. In an interactive framework, a representation can be structured orthographically and still be sensitive to lexical constraints and influences from other input modalities. That is, we can consider the VWFA to be the approximate neural analog of the pool of units labeled "orthography" in the triangle model, which primarily represent orthographic structure but are also sensitive to interactive influences from other representations. A considerable body of evidence supports the view that processing in this region is susceptible to influences from other input modalities, including influences arising from tactile (Braille)

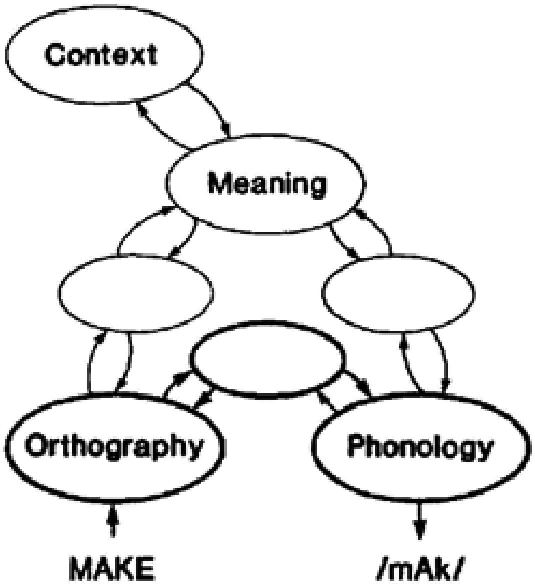

Fig. 9. The triangle model framework for single-word reading. Reprinted from fig. 1, p. 526, of Seidenberg and McClelland (1989). Copyright © American Psychological Association, reprinted with permission.

input for congenitally blind patients (Buchel, Price, Frackowiak, & Friston, 1998; Cohen et al., 1997) for handwriting (Barton, Fox, Sekunova, & Iaria, 2010) and for auditory word processing (Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Cone, Burman, Bitan, Bolger, & Booth, 2008; Desroches et al., 2010). As important, studies that have looked at the influence of consistency between a word's spelling and its sound have revealed graded effects of consistency and frequency at the item level mirroring the behavioral findings of consistency effects in naming (Bolger, Hornickel, Cone, Burman, & Booth, 2008; Bolger, Minas, Burman, & Booth, 2008; Graves, Desai, Humphries, Seidenberg, & Binder, 2010). Consistent with predictions from the Triangle model (Harm, McCandliss, & Seidenberg, 2003), Bolger, Hornickel, et al. (2008) and Bolger, Minas, et al. (2008) found that response to grapheme–phoneme consistency in the VWFA increased with reading skill. These findings support the view that interactive processing becomes established as reading skill becomes more and more automatic; this is captured in the triangle model framework in terms of the strengthening of bidirectional connections between the neurons participating in each of the three different types of representations with experience.

Neuroimaging studies of speech perception have also addressed the predictions of interactive models. Whereas accuracy of phonological perception is associated with the superior temporal cortex, decision time is associated with inferior frontal/insula cortex (Binder, Liebenthal, Possing, Medler, & Ward, 2004) and anterior cingulate/medial frontal regions of cortex (Grinband, Hirsch, & Ferrera, 2006; Grinband et al., 2011). Interactive models predict that brain regions involved in phonological processing (e.g., posterior superior temporal gyrus and Heschl's gyrus in the superior temporal sulcus) should show effects of lexical bias. In contrast, autonomous decision-level integration models predict that these lexical bias effects should be limited to brain regions involved in decision-making and response selection (e.g., inferior frontal gyrus and anterior cingulate gyrus). An fMRI study (Myers & Blumstein, 2008; see also Guediche, Salvata, & Blumstein, 2013) found that the lexical bias on categorization of ambiguous phonemes was associated with increased activation in the superior temporal gyrus. This region is also activated during auditory hallucinations of voices in patient populations (Dierks et al., 1999) and imagined speech of others in healthy individuals (McGuire, Silbersweig, & Frith, 1996).

Electrophysiological measures have provided key evidence that lexical and consistency effects occur early, during perceptual and/or lexical processing, rather than during a post-perceptual decision stage, in both visual and auditory modalities. For example, rhyming effects on visual processing of orthographically dissimilar words have been detected around 260 ms after stimulus onset (Kramer & Donchin, 1987), and syllable effects in visual word processing have been shown at around 250–350 ms (Ashby & Martin, 2008; Carreiras, Ferrand, Grainger, & Perea, 2005). Consistency effects in auditory lexical decision tasks (Perre, Midgley, & Ziegler, 2009; Perre & Ziegler, 2008) and semantic categorization tasks (Pattamadilok, Perre, Dufau, & Ziegler, 2008;Pattamadilok, Morais, De Vylder, Ventura, & Kolinsky, 2009) have been shown to occur in ERP roughly 300–350 ms post-stimulus and time locked to the point of inconsistency. Findings from MEG imaging, which provides greater spatial resolution, have localized the early rhyming effects in visual tasks to the left occipito-temporal region (Wilson, Leuthold, Lewis, Georgopoulos, & Pardo, 2005). In related work, van Linden and colleagues (van Linden, Stekelenburg, Tuomainen, & Vroomen, 2007) found that lexical context induced an early, perceptually based mismatch negativity effect, suggesting that lexical information directly affected perceptual processing stages.

Although neuron-level neuroanatomical precision is difficult to achieve in the domain of human language processing, recent studies combining multiple imaging modalities show promise for increasing both spatial and temporal precision. A study combining MEG and electro-encephalography (EEG) with anatomical MRI (Gow, Segawa, Ahlfors, & Lin, 2008) found reactivation of posterior superior temporal gyrus following activation of a region associated with lexical processing (supramarginal gyrus). An ERP study (Molinaro, Duñabeitia, Marìn-Gutièrrez, & Carreiras, 2010) found that during an early period (180–220 ms after onset) letter-like numbers in word contexts (e.g., M4T3R14L) were processed more like numbers than letters, but only slightly later (250–300 ms after onset) this pattern reversed and letter-like numbers were processed more like letters than numbers. A combined ERP-MEG study (Sohoglu, Peelle, Carlyon, & Davis, 2012) replicated the facilitative effect of prior knowledge (written text) on perceptual clarity of degraded speech and found that this effect was reflected in inferior frontal gyrus activity before superior temporal gyrus activity, consistent with top-down feedback from higher level processing in the inferior frontal gyrus modulating perceptual processing in the superior temporal gyrus.

The exact nature, timing, and location of lexical and consistency effects in visual and auditory word perception remains subject to a range of interpretations, and a considerable body of ongoing work is addressing these issues. One very general open question is whether top-down and between-modality influences should be viewed as an additional sources of constraint on interpretation, as in the interactive activation framework, or whether, instead, top-down signals should be viewed as predictions that are compared with bottom-up signals, generating error signals that then drive learning mechanisms (Friston, 2008; Mumford, 1992; Rao & Ballard, 1999). A further question is the interplay between such influences and synchronization of neural activity within and across brain regions (see Gotts, Chow, & Martin, 2012 and commentaries therein for a recent discussion).

There appears to be little doubt that top-down influences affect relatively early, modality-specific processing areas, both in language processing and in other tasks. Brain regions tend to be connected bidirectionally and there is strong neurophysiological evidence that these bidirectional connections implement interactive activation in perceptual and conceptual processes (Ghuman, Bar, Dobbins, & Schnyer, 2008; Gotts et al., 2012). Specifically within the domain of language processing, the neural evidence indicates that feedback and audio-visual interactions directly influence perceptual processing, consistent with interactive models.

8. Summary and future directions

Over the course of this article, we have laid out the case for interactive activation and mutual constraint satisfaction in perception and cognition. We have focused primarily on visual and spoken word recognition, the target phenomena first addressed by IA models, but we have also considered other applications of interactive approaches. We have explored computational theory-level considerations and neuroscience evidence as well as evidence on the role of context in perception as revealed by behavioral studies.

We have argued that interactive activation addresses key computational challenges facing perceptual systems and is consistent with a wide range of evidence, including behavioral and neuroscience evidence on the mechanisms of perception and language processing in the brain. Overall, it appears that both computational analyses and the behavioral and neuroscience evidence are consistent with the IA hypothesis.

While the computational and empirical considerations seem strongly to support an interactive perspective, there are several important challenges requiring future investigation within an interactive activation framework.

Dynamics of perception in probabilistically grounded interactive activation models. The IA hypothesis states that processing approaches the ideal of achieving optimal results in real time as information becomes available. A good deal of experimental work has been carried out showing that participants in perceptual and language-processing tasks use all of the available information and start to show sensitivity to it within a third of a second of its arrival at the sensory surface. Simulations of such findings have been provided using the original TRACE model and related, simple Luce-ratio-based models (Spivey & Tanenhaus, 1998). Future work should explore these issues in more detail, relying on probabilistically grounded models like the multinomial IA model.

We have begun to explore a related issue in the multinomial IA model (Khaitan & McClelland, 2010), namely, the build-up of performance—and of contextual influence on performance—as participants are given increasing amounts of exposure to target information (Massaro & Klitzke, 1979). This issue is important because Massaro and Cohen (1991) specifically posed it as a challenge to the interactive activation model that was not fully addressed by the stochastic version of the model presented in McClelland (1991). Specifically, if input feature information builds up over time according to the empirical function proposed by Massaro and Cohen (1991), would the processing machinery provided by the multinomial IA model show the right pattern of enhancement for perception of letters in words compared to letters in random sequences? The simulation reported in Khaitan and McClelland (2010) suggests that the answer to this question may be yes, but the simulation is preliminary, and more work is needed.

Adaptive optimization to task and instructional constraints. An important topic for further research is the adaptive optimization of processing in interactive activation models in response to task and instructional constraints. There are a number of important open issues here. First, as we have noted, participants do adjust the extent to which they show lexical influences on processing as a function of changes in the probability that stimulus items will be words or non-words. Such influences are easily incorporated into Bayesian models (Rumelhart & Siple, 1974 consider this issue extensively) and have also been incorporated into the original IA and TRACE models (Mirman et al., 2008). It appears, however, that there are limits on the extent to which participants can actually suspend the use of their knowledge of lexical constraints on speech sound identity. For example, in one recent study (Hawthorne, 2011) participants showed lexical influences on perception of speech sounds whether or not they were informed that each sound occurred equally often in each of two possible contexts, as one might expect if the knowledge of lexical constraints were hard wired into connections among the neurons involved in naturalistic language processing, and these same neurons and connections were relied upon independent of the instructional manipulation. There are empirical and theoretical questions here that deserve further consideration.

Incorporating learning and distributed representations in interactive models of perception. Research on interactive activation models of perception pre-dated the development of powerful learning models for parallel distributed processing systems that were developed in the mid-1980s. Models using learned distributed representations have been successful in addressing a wide range of aspects of linguistic and semantic processing, and we look forward to full integration of learning and distributed representation in further explorations of perceptual processing tasks. Recent developments of fast and powerful learning methods for deep belief networks (Hinton, 2014; this issue) should facilitate these explorations.

Meeting the computational challenges facing perceptual and cognitive systems in naturalistic perceptual contexts. The IA and TRACE models that have been the focus of our investigations here finesse many challenges facing the development of models that will be robust and efficient enough to succeed in matching human capabilities in naturalistic perceptual situations. These challenges are the focus of intense research among a wide range of researchers in the fields of AI, machine vision, and machine learning. Much of this work builds on neural network ideas with origins in IA models and precursors of such models, and of course a great deal of this work incorporates explicit probabilistic inferencing mechanisms. In turn, much of this work should feed back into the effort to understand human perceptual processing mechanisms, as they are instantiated in the neural mechanisms provided by the brain. The further development of interactive activation models of perception will benefit greatly from these developments.

Fully grounding IA models in the neural mechanisms provided by the brain. The final challenge we will mention is the goal of understanding exactly how the IA process is implemented in the neural machinery in the brain. Neurons and their properties have been a source of inspiration in the development of these models, and evidence from neuroscience supports the view that perceptual processing in the brain is an IA-like process, as we have reviewed. Building an integrated understanding of the way in which neural mechanisms give rise to perception is a goal that many researchers strive for; if the IA hypothesis is correct, such an integrated understanding will rely on principles of interactive activation.

Acknowledgments

This research was supported in part by Air Force Research Laboratory Grant FA9550-07-1-0537. The authors would like to acknowledge the useful comments of the members of the Parallel Distributed Processing laboratory at Stanford University.

Notes

1. Although the original IA model employed between-level inhibition as well as excitation, the TRACE model and other subsequent models used excitatory-only connections between levels with inhibition restricted to within-level interactions. The primary reason for eliminating between-level inhibition was to allow even a poorly fitting interpretation to become active when there is no better interpretation. We will return to this issue in discussing the multinomial interactive activation model below.
2. For simplicity, the IA and TRACE models assumed discrete slots for letters and phonemes, although TRACE assumed some spread of phonological features producing overlap between adjacent slots. Recent evidence reviewed in Norris (2013) suggests that both models should allow for positional uncertainty, so that letters near the appropriate position can still activate the corresponding word-level unit (e.g., TRCK should activate the word TRUCK much more than TRXY does).
3. Presentations of the original IA model did not stress that it contained separate units for the presence and for the absence of each possible feature. Fig. 2 makes this feature of the model more explicit than in earlier diagrams of the model.
4. Note that the $p(w)$ values used in the model are not raw word frequencies; instead, as in the original IA model, these probabilities are compressed (McClelland & Rumelhart, 1981). Without this compression, there would be a much larger range of variation in the posterior probabilities shown in Fig. 6. The compression of the $p(w)$ values amounts to an (implicit) "assumption" about stimulus frequency incorporated in the model. The bias terms on the word units are the natural logarithms of these already-compressed $p(w)$ values.
5. This result follows from the fact that the sum of the logarithms of a set of quantities is equal to the logarithm of the product of the quantities, for example, $\ln(a)+\ln(b) = \ln(ab)$, and the fact that e to the log of a quantity is simply the quantity itself, that is, $e^{\ln x} = x$. We also rely on the fact that $e^{x/T} = (e^x)^{1/T}$.
6. The complete Bayes' formula would contain factors for the prior probabilities of letters. However, in the generative model, letters do not have independent prior probabilities; instead, letter probabilities depend on the word level, whose influence on the letter level is incorporated on the second and subsequent updates of the units at the letter level. On the first update, letters are treated as equally probable. Such a constant factor would cancel out and is therefore not expressed in the equation.

7. It should be noted here that changing the temperature parameter is equivalent to scaling the weights and biases in the model, and these in turn represent relative probabilities and relative conditional probabilities in the generative model. Thus, a lower temperature corresponds to assuming less randomness in the generative model.
8. If the model was required to read out from the word level, it would always produce a word response, but the same would have been true of the original IA model. When asked to report all four letters, human observers do not always report words when pseudowords are presented (McClelland & Johnston, 1977). Further research is needed to determine if the pattern of whole report responses obtained with pseudowords can be explained by the MIA model, assuming readout from the four-letter positions.
9. In Massaro's model (Massaro, 1989), the relative stimulus support for *r*, called s_r, corresponds to $p(s|r)/(p(s|r)+p(s|l))$; and the relative context support c_r corresponds to $p(r|c)/(p(r|c)+p(l|c))$. The stimulus and context support for *l* are defined similarly. Since $s_r + s_l = 1$, s_l can be replaced by $1 - s_r$; similarly, c_l can be replaced by $1 - c_r$. Thus, for the two alternative case his model then becomes $p(r|s,c) = s_r\ c_r\ /\ (s_r\ c_r + (1 - s_r)(1 - c_r))$. Participants then choose the *r* response with a probability equal to the resulting estimate of $p(r|s,c)$.
10. As Pearl (1982) showed, it is possible to keep a record of the information passed up from each position to a higher level, and then cancel this back out of the top-down signal broadcast down to all lower levels from above, and a precursor of this idea was described in Rumelhart (1977). We view Pearl's proposal as an alternative implementation of an interactive model of perception; a comparison of this approach to the MIA model is provided in McClelland (2013).

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language*, *38*(4), 419–439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(1), 224–236. doi:10.1037/0096-1523.34.1.224.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629.

Barton, J. J. S., Fox, C. J., Sekunova, A., & Iaria, G. (2010). Encoding in the visual word form area: An fMRI adaptation study of words versus handwriting. *Journal of Cognitive Neuroscience*, *22*(8), 1649–1661. doi:10.1162/jocn.2009.21286.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nature Reviews Neuroscience*, *3*, 443–452.

Binder, J. R., Liebenthal, E., Possing, E. T., Medler, D. A., & Ward, B. D. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, *7*(3), 295–301.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage*, *33*(2), 739–748.

Bizley, J., & King, A. (2009). Visual influences on ferret auditory cortex. *Hearing Research*, *258*, 55–63.

Bolger, D. J., Hornickel, J., Cone, N. E., Burman, D. D., & Booth, J. R. (2008). Neural correlates of orthographic and phonological consistency effects in children. *Human Brain Mapping*, *29*(12), 1416–1429.

Bolger, D. J., Minas, J., Burman, D. D., & Booth, J. R. (2008). Differential effects of orthographic and phonological consistency in cortex for children with and without reading impairment. *Neuropsychologia*, *46*(14), 3210–3224.

Bowers, J. S. (2009). On the biological plausibility of grandmother cells: Implications for neural network theories in psychology and neuroscience. *Psychological Review*, *116*, 220–251.

van den Brink, D. l., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

Buchel, C., Price, C., Frackowiak, R. S. J., & Friston, K. (1998). Different activation patterns in the visual cortex of late and congenitally blind subjects. *Brain*, *121*, 409–419.

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*, *22*, 2886–2902.

Carrasco, A., & Lomber, S. G. (2010). Reciprocal modulatory influences between tonotopic and nontonotopic cortical fields in the cat. *The Journal of Neuroscience*, *30*(4), 1476–1487.

Carreiras, M., Ferrand, L., Grainger, J., & Perea, M. (2005). Sequential effects of phonological priming in visual word recognition. *Psychological Science*, *16*(8), 585–589. doi:10.1111/j.1467-9280.2005.01579.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(3), 687–696.

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Falz, L., Dambrosia, J., Honda, M., Sadato, N., Gerloff, C., Catala, M. D., & Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, *389*(6647), 180–183. doi:10.1038/38278

Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology*, *105*, 239–269.

Cone, N. E., Burman, D. D., Bitan, T., Bolger, D. J., & Booth, J. R. (2008). Developmental changes in brain regions involved in phonological and orthographic processing during spoken language processing. *Neuroimage*, *41*(2), 623–635.

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498.

Dean, T. (2005). A computational model of the cerebral cortex. In *Proceedings of AAAI-05* (pp. 938–943). Cambridge, MA: MIT Press.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Sciences*, *9*(7), 335–341.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117.

Desroches, A. S., Cone, N. E., Bolger, D. J., Bitan, T., Burman, D. D., & Booth, J. R. (2010). Children with reading difficulties show differences in brain regions associated with orthographic processing during spoken language processing. *Brain Research*, *1356*, 73–84.

Dierks, T., Linden, D. E. J., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W., et al. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Psychiatry: Interpersonal and Biological Processes*, *22*, 615–621.

Elman, J. L., & McClelland, J. L. (1984). Speech perception as a cognitive process: The interactive activation model. In Norman Lass (Ed.), *Speech and Language*. Vol. *10*. New York: Academic Press.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language*, *27*(2), 143–165.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, G. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, *22*, 5749–5759.

Feldman, N. H., Griffiths, T. L., & Mrogan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review*, *116*, 752–782.

Felleman, D.J. & van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 526–540.

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*, 1325–1352.

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, *6*(1), 110–125.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721741.

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Science*, *10*, 278–285.

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 Sciences of the United States of America*, *105*(24), 8405–8409.

Gotts, S. J., Chow, C. C., & Martin, A. (2012). Repetition priming and repetition suppression: A case for enhanced efficiency through neural synchronization. *Cognitive Neuroscience*, *3*(3–4), 227–237.

Gow, D. W., Segawa, J. A., Ahlfors, S. P., & Lin, F.-H. (2008). Lexical influences on speech perception: A Granger causality analysis of meg and eeg source estimates. *NeuroImage*, *43*, 614–623.

Gratton, G., Coles, M. H., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre- and post stimulus activation of response channels: A psychophysiological analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 331–344.

Graves, W. W., Desai, R., Humphries, C., Seidenberg, Mark S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex*, *20*(8), 1799–1815. doi:10.1093/cercor/bhp245.

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

Grinband, J., Hirsch, J., & Ferrera, V. P. (2006). A neural representation of categorization uncertainty in the human brain. *Neuron*, *49*(5), 757–763.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage*, *57*(2), 303–311.

Grossberg, S. A. (1978). A theory of coding, memory, and development. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception* (p. 1978). New York: Wiley.

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

Guediche, S., Salvata, C., & Blumstein, S. E. (2013). Temporal cortex reflects effects of sentence context on phonetic processing. *Journal of Cognitive Neuroscience*, *25*(5), 706–718.

Hall, A. J., & Lomber, S. G. (2008). Auditory cortex projections target the peripheral field representation of primary visual cortex. *Experimental Brain Research*, *190*(4), 413–430.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, *7*(2), 155–182.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al (1975). *Journal of Memory & Language*, *52*(1), 58–70.

Hawthorne, D. J. (2011). Can instructions diminish the influence of phonetic categories on the perception of speech sounds? Unpublished research paper, Department of Psychology, Stanford University.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*(2), 181–197. doi:10.1017/S0952523800009640.

Henderson, C. M., & McClelland, J. L. (2011). A PDP model of the simultaneous perception of multiple objects. *Connection Science*, *23*, 161–172.

Hinton, G. E. (2014). Where do features come from? *Cognitive Science*. DOI: 10.1111/cogs.12047 [Epub ahead of print]

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 1: Foundations* (pp. 77–109). Cambridge, MA: MIT Press.

Hinton, G. E., & Sejnowski, T. J. (1983). Optimal perceptual inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, DC.

Hinton, G. E., & Sejnowski, T. J. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume I. (Ch. 7, pp 282–317). Cambridge, MA: MIT Press.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558.

Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, *394*, 784–787.

James, R. C. (1965). Photo of a dalmation dog. *LIFE Magazine*, *58*(7), 120.

Jefferies, E., Frankish, C. R., & Ralph, M. A. L. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54*(1), 81–98.

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *93*, 793–820.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Khaitan, P., & McClelland, J. L. (2010). Matching exact posterior probabilities in the Multinomial Interactive Activation Model. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society* (p. 623). Austin, TX: Cognitive Science Society.

Kramer, A. F., & Donchin, E. (1987). Brain potentials as indices of orthographic and phonological interaction during word matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(1), 76.

Kronbichler, M., Bergmann, J., Hutzler, F., Staffen, W., Mair, A., Ladurner, G., & Wimmer, H. (2007). Taxi vs. Taksi: On orthographic word recognition in the left visual occipitotemporal cortex. *Journal of Cognitive Neuroscience*, *19*(10), 1584–1594.

Kronbichler, M., Hutz, F., Wimmer, H., Mair, A., Staffen, W., & Ladurner, G. (2004). The visual word form area and the frequency with which words are encountered: Evidence from a parametric fMRI study. *Neuroimage*, *21*(3), 946–953.

Kubat, R., Mirman, D., & Roy, D. K. (2009). Semantic context effects on color categorization. In N. A. Taatgen & H. V. Rijn (Eds.), *Proceedings of the 31st Annual Cognitive Science Society Meeting* (pp. 491–495). Austin, TX: Cognitive Science.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107.

Kumaran, D., & McClelland, J. L. (2011). Beyond Episodic memory: A complementary learning systems account of the hippocampal contribution to generalization. *Psychological Review*, *119*, 573–616.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–1448.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in early visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(4), 1907–1977.

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences*, *3*(7), 254–264.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*(11), 470–478.

van Linden, S., Stekelenburg, J. J., Tuomainen, J., & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience letters*, *420*(1), 49–52. doi:10.1016/j.neulet.2007.04.006.

van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1483–1494.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science*, *27*(2), 285–298.

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, *108*(3), 866–873.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, *69*, 546–558.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63.

Massaro, D. W. (1979). Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 595–609.

Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, *21*, 398–421.

Massaro, D. W., Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*, 338–348.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, *23*, 558–614.

Massaro, D. W., & Klitzke, D. (1979). The role of lateral masking and orthographic structure in letter and word perception. *Acta Psychologica*, *43*, 413–426.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, *7*(7), 293–299.

McClelland, J. L. (1981). Retrieving general and specific information from stored knowledge of specifics. In *Proceedings of the Third Annual Conference of the Cognitive Science Society* (pp. 170–172).

McClelland, J. L. (1985). Putting knowledge in its place: A scheme for programming parallel processing structures on the fly. *Cognitive Science*, *9*, 113–146.

McClelland, J. L. (1986). The programmable blackboard model of reading. In J. L. McClelland, D. E. Rumelhart, & the PDP research group. *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume II (pp. 122–169). Cambridge, MA: MIT Press.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 1–36). London: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology*, *23*, 1–44.

McClelland, J. L. (1998). Connectionist models and Bayesian inference. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 21–53). Oxford, England: Oxford University Press.

McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology*, *4*, 503. doi:10.3389/fpsyg.2013.00503.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., & Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Perception & Psychophysics*, *22*, 249–261.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, I: An account of basic findings. *Psychological Review*, *88*(5), 375–407.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The Appeal of Parallel Distributed Processing. In D. E. Rumelhart, J. L. McClelland, & the PDP research group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume I. (pp. 3–44). Cambridge, MA: MIT Press.

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*, 907–917.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. *Infancy*, *6*(2), 203–229.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language*, *61*(1), 1–18.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, *13*(6), 958–965.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science*, *32*(2), 398–417.

Mitterer, H., & de Ruiter, J. P. (2008). Recalibrating color categories using world knowledge. *Psychological Science*, *19*(7), 629–634.

Molinaro, N., Duñabeitia, J. A., Marìn-Gutièrrez, A., & Carreiras, M. (2010). From numbers to letters: Feedback regularization in visual word recognition. *Neuropsychologia*, *48*(5), 1343–1355.

Monsell, S., Patterson, K. E., Graham, A., Hughes, C. H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 452–467.

Movellan, J. R., & McClelland, J. L. (2001). The Morton-massaro law of information integration: Implications for models of perception. *Psychological Review*, *108*, 113–148.

Mumford, D. (1992). On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biological Cybernetics*, *66*, 241–251.

Myers, E. B., & Blumstein, S. E. (2008). The neural bases of the lexical effect: An fMRI investigation. *Cerebral Cortex*, *18*(2), 278–288.

Newman, R. S., Sawusch, J. R., & Luce, R. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 873–889.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *The Behavioral and Brain Sciences*, *23*(3), 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation–verification model for letter and word recognition: The word-superiority effect. *Psychological Review*, *89*, 573–594.

Pattamadilok, C., Morais, J., De Vylder, O., Ventura, P., & Kolinsky, R. (2009). The orthographic consistency effect in the recognition of French spoken words: An early developmental shift from sublexical to lexical orthographic activation. *Applied Psycholinguistics*, *30*, 441–462.

Pattamadilok, C., Perre, L., Dufau, S., & Ziegler, J. (2008). On-line orthographic influences on spoken language in a semantic task. *Journal of Cognitive Neuroscience*, *21*(1), 169–179.

Pearl, J. (1982). Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings of AAAI-82*. (pp. 133–136).

Perre, L., Midgley, K., & Ziegler, J. C. (2009). When beef primes reef more than leaf: Orthographic information affects phonological priming in spoken word recognition. *Psychophysiology*, *46*(4), 739–746. doi:10.1111/j.1469-8986.2009.00813.

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain research*, *1188*, 132–138. doi:10.1016/j.brainres.2007.10.084.

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

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *21*(4), 1037–1052.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory & Language*, *39*(3), 347–370.

Pitts, W., & McCullough, W. S. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics*, *9*, 127–147.

Plaut, D. C., & McClelland, J. L. (2010). Locating object knowledge in the brain: A critique of Bowers' (2009) attempt to revive the grandmother cell hypothesis. *Psychological Review*, *117*, 284–288.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extraclassicalreceptive-field effects. *Nature neuroscience*, *2*(1), 79–87.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499.

Reddy, D. R., Erman, L. D., Fennell, R. O., & Neely, R. B. (1973). The hearsay speech understanding system: An example of the recognition process. In *Proceedings of the 3rd international joint conference on Artificial Intelligence* (pp. 185–194). San Francisco, CA: Morgan Kaufmann.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness ofg stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Rizzolatti, G., Riggio, L., Dascola, I., & Umlita, C. (1980). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1a), 31–40.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Journal of Psychophysiology*, *50*, 19–26.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and Performance VI*. (pp. 573–603). Hillsdale, NJ: LEA.

Rumelhart, D. E., & McClelland, J. L. (1981). Interactive processing through spreading activation. In C. Perfetti & A. Lesgold (Eds.), *Interactive processes in reading* (pp 37–60). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94.

Rumelhart, D. E., & Siple, P. (1974). The process of recognizing tachistoscopically presented words. *Psychological Review*, *81*, 99–118.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474.

Samuel, A. G. (1986). Red herring detectors and speech perception: In defense of selective adaptation. *Cognitive Psychology*, *18*(4), 452–499.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology*, *32*(2), 97–127.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, *12*(4), 348–351.

Samuel, A. G., & Kat, D. (1996). Early levels of analysis of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 676–694.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language*, *48*(2), 416–434.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology*, *85*, 1322–1327.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*(4), 489–537.

Sherman, G. (1971). The phonemic restoration effect: An insight into the mechanisms of speech perception. Unpublished master's thesis, University of Wisconsin, Milwaukee.

Smolensky, P. (1986). Neural and conceptual interpretation of PDP models. In J. L. McClelland, D. E. Rumelhart, & the PDP research group. *Parallel distributed processing: Explorations in the microstructure of cognition*. Volume II, (pp. 390–431). Cambridge, MA: MIT Press.

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *The Journal of Neuroscience*, *32*(25), 8443–8453.

Spivey, M., & Tanenhaus, M. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1521–1543.

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, *16*(2), 219–237.

Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech (L). *Journal of the Acoustical Society of America*, *114*(6,Pt1), 3036–3039.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 632–634.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*, 392–393.

Warren, R. M., & Warren, R. P. (1971). Some age differences in auditory perception. *Bulletin of the New York Academy of Medicine*, *47*(11), 1365.

Wilson, T. W., Leuthold, A. C., Lewis, S. M., Georgopoulos, A. P., & Pardo, P. J. (2005). Cognitive dimensions of orthographic stimuli affect occipitotemporal dynamics. *Experimental Brain Research.*, *167* (2), 141–147. doi:10.1007/s00221-005-0011-4.