

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/40906580>

# A Neurocomputational Account of Taxonomic Responding and Fast Mapping in Early Word Learning

ARTICLE *in* PSYCHOLOGICAL REVIEW · JANUARY 2010

Impact Factor: 7.97 · DOI: 10.1037/a0018130 · Source: PubMed

---

CITATIONS

51

---

READS

120

## 2 AUTHORS:



**Julien Mayor**

The University of Nottingham Malaysia Campus

25 PUBLICATIONS 120 CITATIONS

[SEE PROFILE](#)



**Kim Plunkett**

University of Oxford

142 PUBLICATIONS 4,587 CITATIONS

[SEE PROFILE](#)

# A neuro-computational account of taxonomic responding and fast mapping in early word learning

Julien Mayor & Kim Plunkett

Department of Experimental Psychology, University of Oxford,  
South Parks Road, Oxford, OX1 3UD, United Kingdom  
E-mail: julien.mayor@psy.ox.ac.uk, kim.plunkett@psy.ox.ac.uk

## Abstract

We present a neuro-computational model using self-organising maps that accounts for the emergence of taxonomic responding and fast mapping in early word learning, as well as a rapid increase in the rate of acquisition of words observed in late infancy. The quality and efficiency of generalisation of word-object associations is directly related to the quality of pre-lexical, categorical representations in the model. We show how synapto-genesis supports coherent generalisation of word-object associations and that later synaptic pruning minimises metabolic costs without being detrimental to word learning. The role played by joint attentional activities is identified in the model, both at the level of selecting efficient cross-modal synapses and at the behavioural level, by accelerating and refining overall vocabulary acquisition. The model can account for the qualitative shift in the way infants use words, from an associative to a referential-like use, for the pattern of over-extension errors in production and comprehension observed during early childhood and typicality effects observed in lexical development. Interesting by-products of the model include a potential explanation of the shift from prototype to exemplar based effects reported for adult category formation, an account of mispronunciation effects in early lexical development and extendability to include accounts of individual differences in lexical development and specific disorders such as Williams syndrome. The model demonstrates how an established constraint on lexical learning which has often been regarded as domain-specific can emerge from domain-general learning principles that are simultaneously biologically, psychologically and socially plausible.

*Keywords: lexical constraints, early word learning, taxonomic responding, fast mapping, connectionist modelling, generalisation, vocabulary spurt, categorical perception, prototype-based categorisation, exemplar-based categorisation, typicality effects, over-extension errors, Williams syndrome, synapto-genesis, synaptic pruning, critical period, mispronunciation*

## Introduction

A central issue in early lexical acquisition is how infants learn the meaning of words. In a typical labelling situation, the caregiver points at an object (*Fido* the dog) and says “Look, this is a dog!”. In these circumstances, the infant has to rule out a huge number of possible meanings. For example, the word might refer to the size, to the shape, to the colour, to the individual *Fido*, etc., (Quine, 1960). However, infants reliably interpret the word ‘dog’ as a label that can be used for this dog and for all dogs.

An influential solution to this conundrum is that infants make use of linguistic constraints to narrow the hypothesis space in order to assign meaning to words (Markman, 1989). These constraints are intended to provide sufficient information to guide the infant in solving the difficult task of word learning. Three distinct constraints have been proposed to narrow down the set of possible meanings; these are the whole object, the taxonomic and the mutual exclusivity assumptions. The whole object constraint guides the infant to consider the whole object set against its background, and not part of it or a subset of its attributes. Once the infant has narrowed down the hypothesis space to the whole object, she has to determine the range of objects to which the label can be applied. In a labelling situation, infants prefer to extend labels to taxonomic matches (dog-wolf) than to thematic matches (dog-bone); this is referred to as the taxonomic constraint. In order to minimise over-extension of labels as well as to make word learning more efficient, the mutual exclusivity assumption specifies that no two words mean the same thing or that infants are biased to reject synonyms (also formulated as the Principle of Contrast, Clark, 1990). The goal of the present paper is to provide a neuro-computational account of one of these word learning constraints, namely the taxonomic constraint, and to evaluate the assumptions and implications of this account in the context of early word learning.

Markman and Hutchinson (1984) introduced a “weak” form of the taxonomic constraint as a relative preference for taxonomic over thematic extension of labels. They presented one group of young children (2-3 year olds) with an object (say, a dog) that was labelled with a novel word (e.g., ‘dax’). The children were subsequently asked to find “another dax” from a pair of stimuli consisting of a taxonomic alternative (e.g., a cat) and a thematic alternative (e.g., a bone). In another group of children, the object was not labelled. The children were subsequently asked to find “another one” from the same pair of stimuli (the taxonomic and the thematic alternatives). Children were more likely to pick the taxonomic alternative (the cat) when asked to find “another dax” than when asked for “another one”, suggesting that children take novel words to refer to taxonomic categories and not to groups of objects defined by thematic relations.

Many other studies have provided evidence for this kind of taxonomic preference in children (e.g., Mervis & Pani, 1980; Baldwin, 1989; Waxman & Kosowski, 1990; Waxman & Gelman, 1986). A strong form of the taxonomic constraint assumes that “when infants embark upon the process of lexical acquisition, they are initially biased to interpret a word applied to an object as referring to that object and to other members of its kind” (Waxman & Markow, 1995, p.257). In other words, from a *single labelling event*, the infant infers that every object that belongs to the same category is called by the same name. This capacity provides a powerful tool for identifying objects and for communication, since it enables individuals to refer to objects they have never seen before. This ability to generalise

from a single labelling event is closely related to the concept of “fast mapping” (Carey & Bartlett, 1978) whereby infants demonstrate rapid mastery of the appropriate use of labels after a limited number of learning opportunities. Infants who are taxonomic responders are therefore likely to be “fast mappers”. The neuro-computational account described in this paper focuses on the strong form of the taxonomic constraint and demonstrates how fast mapping results.

It is often assumed that these word learning constraints are available to children at the onset of lexical development (Markman, 1989). However, this assumption leaves open the question as to how children learn that lexical constraints can be applied to word learning. Some researchers have proposed that the taxonomic constraint is an adaptation of perceptually-based categorisation. For example, L. B. Smith (2000) argued that category membership can be determined just by looking at the shape of the object when considering the vocabularies of infants. Based on this observation, Samuelson and Smith (1999) provide support for the hypothesis that the objects referred to by early nouns present perceptual regularities that would create a shape-bias (Landau, Smith, & Jones, 1988; Graham & Poulin-Dubois, 1999); children at 2 and 3 years of age show a clear preference to group items according to their overall shape.

Poulin-Dubois et al. (1999) attempted to disentangle the overlap between taxonomic relations and shape similarity in 18- and 24-month-olds. They showed that both age groups extended novel labels to a taxonomically unrelated shape-match, rather than to a taxonomic match differing in shape. They concluded that taxonomic responding can be reduced to shape-biased generalisation of novel labels. In contrast, Cimpian and Markman (2005) argued that category membership can be determined by factors other than purely perceptual characteristics and that the shape bias is a task-induced artifact rather than a general word-learning strategy. They demonstrated that the shape bias disappeared when objects were more complex, presented in context or if 3- to 5-year-old children were not forced to choose between objects. Moreover, young children preferred taxonomic equivalents rather than shape matches when these cues were in conflict. Overall these findings lead us to conclude that the nature of the conceptual representations that support taxonomic responding in infants and young children are quite diverse, incorporating perceptual features such as shape and conceptual factors such as function and other attributes that are not immediately apparent in the perceptual scene (Gopnik & Meltzoff, 1987).

It is noteworthy that there is little evidence that infants generalise newly-learned word-object associations before 15 months of age (Markman & Jaswal, 2004, p.377). Markman (1992) suggested that word learning is divided into two stages; the first stage is governed by simple associative mechanisms that result in slow learning and limited vocabulary size. The second stage is characterised by the use of lexical constraints which support efficient word learning and accelerated vocabulary growth. Similarly, other authors have proposed that the transition towards a referential use of words is associated with the emergence of constraints on lexical acquisition (Clark, 1983; Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989; W. E. Merriman & Bowman, 1989). Nazzi and Bertoncini (2003) propose that the vocabulary spurt observed at around 18 months of age reflects qualitative changes in the way words are acquired and defines a qualitative transition between proto-words (pairing of an under-specified phonological representation to a specific object, see Stager & Werker, 1997) to genuine words (pairing of a specified phonological representation with

a category of objects). Nazzi and Bertoncini (2003) propose that the vocabulary spurt corresponds to a shift from an associative to a referential lexical acquisition mechanism.

The goal of this paper is to outline a mechanism that can learn to respond in a strong taxonomic fashion, i.e., robust generalisation from few labelling events. We will show that the mechanism has emergent properties that can account for the apparent qualitative shift in the way infants use words, from an associative to a referential use, for fast mapping, as well as the quantitative change in the rate of acquisition of vocabulary items that occur during the second year of life. The model exhibits these desirable properties and many others, despite the fact that it relies on continuous changes and does not require the introduction of additional learning mechanisms during the course of learning. In other words, taxonomic responding, fast mapping and non-linear vocabulary growth are emergent properties of learning. We will demonstrate within a modelling framework that an efficient pre-established, categorisation capacity is a prerequisite to successful word learning (Macnamara, 1972). Once the requisite categories are established, joint attentional activities, even a single labelling event, provide the basis for taxonomic responding, in which a label is successfully extended to members of the formed category. Once a label is learnt for the whole category, the word is considered to be part of the infants' lexicon. Before describing the model in detail, we review some previous models of word learning to highlight the unique contribution of the current model as well as the properties it shares with earlier models of word learning.

### Previous network models of word learning

#### *Multilayer neural networks*

With the introduction of the backpropagation algorithm (Werbos, 1974; Rumelhart, Hinton, & Williams, 1986), feed-forward neural networks have been very successful in modelling a wide range of experimental aspects of cognition, including word learning. Plunkett, Sinha, Møller, and Strandsby (1992) used an auto-encoder network to model the development of label-object associations in late infancy. The auto-encoder network architecture consisted of two partially merging sub-networks: a visual and a linguistic sub-network. The visual pathway was presented with random dot images (Posner & Keele, 1968) that were preprocessed using input receptors with Gaussian receptive fields. The second input pathway processed linguistic input corresponding to the names of the random dot patterns. Thirty two categories, each containing 8 objects derived from a different random dot prototype, and 32 labels were presented to the auto-encoder network in a 3-phase training cycle involving the object alone, the label alone and the object-label pair, aimed at capturing the attention switching process to the label, the object and finally the object-label pair. The performance of the network was evaluated by analysing the network's ability to produce the correct label when only an image was presented (production), and to produce the correct image when only a label was presented (comprehension). The model successfully captured the well-known vocabulary spurt, patterns of over- and under-extension errors, prototype effects and the comprehension-production asymmetry observed in late infancy. The vocabulary spurt in the model was directly attributable to the network's discovery of the structure of the visual category space. Mapping labels to objects was easier once the category membership of an object had been deciphered. However, the input labels facili-

tated the process of categorisation as demonstrated by a delay in clustering of the visual space in control simulations where the labels were not auto-encoded. The comprehension-production asymmetry in the model derived from the greater sparseness in the coding of the linguistic input compared the visual input. The pattern of over- and under-extension errors derived from transient anomalies in the clustering of random dot figures in the visual category space. Prototype effects emerged from their overall similarity advantage to the exemplars on which the model was trained.

Regier (2005) proposed a model of word learning that combined the macro-structure of the Plunkett et al. (1992) model with the selective attentional mechanisms implemented in Kruschke's (1992) model of category learning. Regier's model exhibited an acceleration in learning, mimicking the vocabulary spurt and fast mapping, a honing of word-form whereby it learnt to acquire words that differed only slightly in phonological form from existing words and honing of word meaning whereby it learnt to constrain the generalization of word meanings to specific dimensions such as shape. The model also exhibited facilitated learning of second labels later in training. All of these effects were attributed to the identification of the relevant dimensions in the stimulus for category formation and learning the attentional weights associated with them, as in the ALCOVE model (Kruschke, 1992). In Regier's model, the attention weights for shaping meaning were driven by an error signal derived from the form of the word, and attention weights for shaping word-form were driven by an error signal derived from meaning. Word learning and object categorisation were therefore inextricably linked in this error-driven, supervised learning system. An important advantage of Regier's (2005) model over Plunkett et al.'s (1992) model is that it could readily remember a single label-object association, even after learning many other intervening object-label events. This capacity was achieved through the use of local, exemplar-based representations (Kruschke, 1992) that resist catastrophic interference. In networks that use highly distributed exemplar representations, catastrophic interference is only avoided by using an interleaved training regime (Plunkett et al., 1992) or by exploiting a complementary learning systems approach (McClelland, McNaughton, & O'Reilly, 1995).

In a series of publications (Rogers et al., 2004; Rogers & McClelland, 2004; McClelland & Rogers, 2003), Rogers and colleagues modelled the emergence of conceptual representations using PDP networks trained to associate animate and inanimate objects and their attributes. The model accounted for the gradual differentiation of conceptual structure that is evident in infancy and early childhood and the privileged status of the basic category level of object naming. A central factor in determining these effects was the variable patterns of *coherent covariation* apparent in the properties associated with different object concepts. Rogers and McClelland (2004) showed that the over-extension of objects names was driven by the frequency of presentation of items and that reduction of over-extension errors was causally related to the progressive differentiation of categories in the network. Rogers et al. (2004) lesioned the model by removing connections between layers and found parallels between the network's behaviour and the pattern of errors observed in patients with semantic dementia.

Colunga and Smith (2005) described a connectionist model that learns to associate shape with solid objects and material with nonsolid things. Learning was based on the correlations between shape and category membership for solid objects, while material was varied randomly. Through multiple presentations of the objects along with their corre-

lated attributes, the network identified the determining attributes for category membership (shape for solid objects and material for nonsolid items) and discarded irrelevant attributes (material for solid objects and shape for nonsolid things). The network's representation of novel objects changed after the presentation of a small vocabulary representative of young children. Effectively, the presentation of objects modified the similarity space so that novel solid objects sharing the same shape were more similar from the network's perspective than novel objects sharing the same material. The model provides a clear illustration of how the presentation of a structured environment to the network can create a word-learning bias, sensitivity to shape, changing the way novel objects are represented.

Landauer and Dumais (1997) described a mathematical method for gaining global knowledge indirectly from local co-occurrence data. The model highlights the fact that the development of high level structure in semantics facilitates the assimilation of new words that fit the structure. Because semantic structure takes time to develop, non-linearity in the rate of label assimilation is an inevitable property of the model. Landauer and Dumais (1997) suggest that this non-linearity could form the basis of the vocabulary spurt. However, it is important to note that the latent semantic analysis obtained from the Landauer and Dumais (1997) model is based entirely on processing of very large quantities of textual information. It is unclear whether infants and young children would have experienced the auditory equivalent of the quantity and range of the textual co-occurrence relations available to the model. Nor does Landauer and Dumais' model *know* the objects to which its lexicon refers. It is therefore unable to respond in a taxonomic fashion. Nevertheless, Landauer and Dumais (1997) highlighted the importance of initial structure for efficient lexical acquisition.

*Learning in the absence of (constant) supervision.* None of the models described above exhibited taxonomic responding and fast mapping since they involved training the networks on repeated and multiple object-label associations in order to extract the statistical structure of the label-category relationships. Classical backpropagation networks need constant supervision provided by an error-correction learning algorithm in order to generate meaningful patterns of generalisation. Presentation of a representative set of objects and their corresponding targets, in order to sample the training space, permits discovery of the extensions and limits of categories (Plunkett & Nakisa, 1997). Training on a single input-target pair (a single supervision event) leads to the memorisation of the pair through application of the error-correction learning algorithm. Through multiple supervision events, multi-layered backpropagation networks are able to learn the set of relevant combinations of attributes, discarding those that do not contribute to a meaningful decision for category membership. As a consequence, these models sustain little coherent generalisation of object-label associations following a single labelling event. In contrast, infants are able to respond taxonomically after a single labelling event, i.e., they can generalise the label to other members of the category. Furthermore, it is generally assumed that infants are able to learn about the structure of categories and sounds in an unsupervised manner during early development, e.g., infants do not need to be provided with the label 'dog' to learn the category 'dog', nor do they need the category to learn about the label.

Although we shall argue that word learning is essentially a *supervised* activity, since the mappings between labels and objects are arbitrary, the preconditions that enable efficient word learning, namely object categorisation and word identification, are essentially

unsupervised activities. The type of supervised learning used in homogeneous classical back-propagation networks is insufficient to explain these abilities. Therefore, a different type of network architecture is required to handle taxonomic responding and fast mapping in a manner that honours the mixture of supervised and unsupervised activities that underpin word learning.

Since word learning is essentially a supervised activity whereas object category formation and label identification are unsupervised, we require a network model that incorporates both types of learning processes. There is a class of models that can extract categories out of a complex input space without supervision, using biologically-grounded learning mechanisms. They rely on Self-Organising Maps (Kohonen, 1984). We will review the most important models of this kind that are relevant to word learning and then show how an adaptation of these models can successfully account for taxonomic responding, following a single labelling event, i.e., fast mapping.

### *Self-organising models*

Most models of word learning that use self-organising maps exploit activity-dependent learning algorithms. Self-organising maps (SOMs) are known to be an efficient computational method for forming categories in an unsupervised way in a complex input space. SOMs extract statistical regularities from the input and form categories of similar objects without explicit supervision. They are efficient at forming categories from high-dimensional input data through dimensionality reduction and self-organisation around topological maps. SOMs are thereby able to capture the *natural clustering* of objects which share properties with each other. At the end of the process of self-organisation, similar objects activate neighbouring neurones in the map.

SOMs are not only a very powerful method for extracting information from complex data-sets, they can also be implemented as a biologically grounded algorithm (Kohonen, 1993). The product of self-organisation is very similar to the organisation of a wide range of mammalian cortical areas. In particular, SOMs have been very successful at modelling the architecture of the primary visual cortex (Miikkulainen, Bednar, Choe, & Sirosh, 2005) where neighbouring neurones are responsive to similar orientations of the visual scene (Hubel & Wiesel, 1962). Topologically organised maps have also been found in the bat auditory cortex (Suga & O'Neill, 1979). This kind of topographical organisation in the brain has been shown to be optimised in terms of total length of neural wiring (Durbin & Mitchinson, 1990). Furthermore, connections between SOMs can be modulated by the activity of individual neurones in each map. There is an extensive literature identifying such synapses in the cortex (Kelso, Ganong, & Brown, 1986; Malinow & Miller, 1986; T. Brown, Kairiss, & Keenan, 1990). These types of learning mechanisms are known as Hebbian synapses (Hebb, 1949), where a synapse connecting two neurones that are active simultaneously should be strengthened.

Previous models have used self-organising maps to simulate aspects of word learning.<sup>1</sup> Schyns (1991) was the first attempt to use SOMs in order to model conceptual acquisition. He proposed a neural network consisting of two modules. The first module used a SOM to categorise simple geometrical forms into 3 distinct categories. A second supervised module

---

<sup>1</sup>Other types of map architectures have been used to model aspects of early infant babbling and speech perception (Westermann & Miranda, 2004).



learnt the names for the 3 categories. The model demonstrated how supervised and unsupervised learning techniques could be combined to achieve simple word-learning capabilities. Miikkulainen (1997) introduced a model of dyslexic and aphasic disorders called DISLEX based on a SOM architecture. The model assumed that the lexical system consists of separate topologically ordered feature maps for different modalities and lexical semantics. The system was able to link sounds and meanings based on their co-occurrence through Hebbian learning. Through selective lesioning procedures, the model was able to mimic language dysfunction, semantic slips, category-specific aphasic impairments and dyslexic behaviour.

Li and colleagues (2004; 2007) proposed two variants of a self-organising neural network model of early lexical development. The first version, DevLex, consisted of two self-organising maps: a growing semantic map and a growing phonological map that were connected via associative links trained by Hebbian learning (Li et al., 2004). It allowed for the representation of a dynamically changing linguistic environment in language learning. The model simulated the acquisition of linguistic categories, such as nouns, verbs and adjectives, as well as lexical confusions as a function of word density and semantic similarity. DevLex also accounted for age-of-acquisition effects in the course of learning a lexicon. The model also predicted relations between lexical density and lexical confusion, and between lexical density and resistance to damage, in a manner that resembled the behaviour of the Rogers et al. (2004) simulation.

In a second variant called DevLex-II (Li et al., 2007), an additional phonemic sequence output layer was included in the network. The model simulated the challenge infants face when they need to articulate phonemic sequences of words. DevLex-II highlighted the role played by input characteristics, such as word frequency or word length, on the rate of acquisition and size of the infant’s vocabulary. The model underwent a rapid increase of lexical growth similar to a vocabulary spurt. Individual differences in word learning trajectories in the model were associated with differences in phonological short-term memory. Finally, the network’s recovery from damage at different points in development was shown to be non-monotonic, suggesting critical period effects. DevLex-II, together with DevLex, provide insights into the individual variations in the rate and size of early vocabulary development with respect to input characteristics (e.g., word frequency and word length effects) and in terms of the consolidation of word-meaning structures.

These models demonstrate the utility of SOMs in understanding lexical development in infants and young children. They show how non-linear lexical growth, differentiated linguistic categories, lexical confusion, and many other properties can emerge through a process of learning using biologically plausible mechanisms which are capable of scaling up to a level that is consistent with human vocabulary sizes. However, none of these models have attempted to simulate the generalisation properties of word-object associations. Nor have they dealt with the lack of invariance in word pronunciation and the diversity of objects that confront infants in the word learning situation. In the next section, we introduce a self-organising model of early word learning that builds upon these earlier applications of SOMs and identify the conditions under which taxonomic responding can emerge in a training environment where object categories and word meanings have yet to be discovered.

## Taxonomic responding as an emergent property

*Overview*

We identify two main contributing factors to successful word learning: pre-lexical categorisation abilities and joint attentional activities. We briefly review the time course of the developmental trajectory of the infants' perceptual systems and how speech perception abilities correlate with later language skills. We then describe the evolution of joint attentional skills in the first two years of life and review evidence for the correlation between joint attentional activities and later word learning capacity. Finally, we outline how both contributing effects are included in our model of early lexical acquisition.

*Early infancy: development of perceptual systems.* Early in development, the infant extracts statistical regularities from the visual scene and speech, enabling her to refine her perceptual systems. The visual system adapts from a basic structure in the visual cortex to an organisation involving complex orientation selectivity of its cells. The organisation of the mammalian visual cortex has been shown to be shaped by the statistics of the visual environment (Crair, Gillespie, & Stryker, 1998); manipulation of the statistics of visual scenes in cats (Blakemore & Cooper, 1970) and environmental deprivation in ferrets (L. White, Coppola, & Fitzpatrick, 2001) have emphasised the role of experience in fine-tuning the organisation of the visual cortex. Similarly, during the first weeks of human life, the development of cortical channels with specialised bandpass spatial-frequency filtering and orientation tuning takes place (Suter et al., 1994). By the end of the second year, brain structures are considered to have the appearance of those of adults (Johnson, 2001).

A similar perceptual refinement occurs at the level of speech processing. From a coarse sensitivity to speech, as shown in prenatal exposure experiments (Mehler et al., 2002) and fetal voice recognition (Kisilevsky et al., 2003), to identifying language-specific sound patterns at 3 months, as shown in forward versus backward speech experiments (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002), infants can segment words from a continuous stream of speech by about 7 months of age (Jusczyk & Aslin, 1995). Infants show a sensitivity to native vs. non-native phonetic contrasts by 10–12 months of age (Werker & Tees, 1984). It is noteworthy that the difference in brain activity between known vs. unknown words is bilaterally distributed in 13 month-olds but is lateralised to the left hemisphere in 20 month-olds (Mills, Coffey-Corina, & Neville, 1997). This suggests that brain organisation is still adapting to the native language late in the second year of life even though infants have become sophisticated speech processors.

Two studies have examined the degree to which early measures of speech perception predict infants' later language skills. Individual differences in speech perception at 6 months of age predict infants' later language abilities at 13, 16 and 24 months of age (F. Tsao, Liu, & Kuhl, 2004). General auditory or cognitive abilities are not the only explanation of individual differences in language abilities. Kuhl et al. (2005) showed both a positive correlation of native contrast discrimination and a negative correlation of non-native contrast discrimination to later abilities, suggesting a reorganisation of the resources dedicated to native language processing. Similarly, deficits in speech perception predict language learning impairments (Ziegler, Pech-Georgel, George, & Lorenzi, 2005) and more general specific language impairments are associated with both poor auditory perception

(Elliott, Hammer, & Scholl, 1989) and impairment of visual imagery (Johnston & Weismer, 1983). Hart and Risley (1995) as well as Huttenlocher (1991) also reported a correlation between the amount of raw exposure to speech and later vocabulary size. The link between early speech sensitivities and later language skills contributes to the idea that perceptual refinement “bootstraps” word learning (Werker & Yeung, 2005). A successful word learning model should incorporate the idea that word learning benefits from a well-established pre-lexical categorisation capacity.

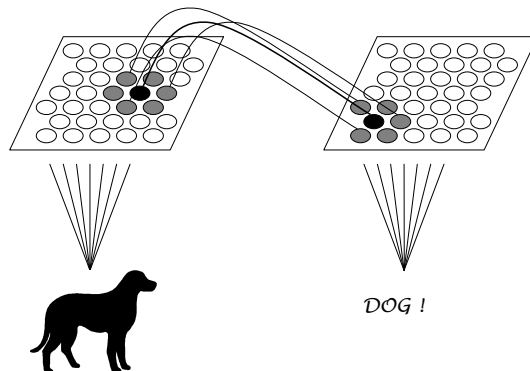
*Integration of joint attentional activities.* During the first year of life, infants develop joint attentional skills that are important for promoting language development. Most 2 month-olds fail to turn their head to follow an adult’s line of regard (Bruner & Scaife, 1975). Butterworth and colleagues (Butterworth & Cochran, 1980; Butterworth & Grover, 1990; Butterworth & Jarrett, 1991) have shown that 6-month-old infants reliably turn their heads to the correct side when following an adults’ direction of gaze, but did not identify which object on that side was the target. By 12 months, infants successfully turn their heads and identified objects, provided the objects are already in their visual fields. By 18 months, infants engage in active search and can locate the target even if it is located behind them. Houston-Price et al. (2006) have shown that 15-month-old infants can disambiguate the referent of a novel label purely on the basis of head-turn information. In summary, early in the second year of life, infants are able to engage in fine-tuned visual joint attentional activities.

The amount of joint attentional activities between infants and their caregivers is a predictor of word learning abilities (Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). Shared attentional experience correlates with novel word learning skills at 17 months (Tomasello & Farrar, 1986), at 18 months (Dunham, 1993) and 16 to 19 months (Baldwin, 1991). Pointing behaviour also has been shown to correlate positively with later language skills (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979; Camaioni, Castelli, Longobardi, & Volterra, 1991). It has been shown that infants’ vocabulary acquisition proceeds at a faster pace when parents are sensitive to their infant’s own focus of attention when providing language (M. Harris, Jones, Brookes, & Grant, 1986; Dunham & Dunham, 1990). Joint attention plays a crucial role in early word learning since it enables the infant to match words and their referents through gaze following or pointing behaviour.

#### *Overview of the model.*

The model consists of two separate SOMs, visual and auditory, that receive visual input and input from acoustic tokens, respectively. We assume that infants have already developed the ability to segment objects out of complex visual scenes (Kellman, Spelke, & Short, 1986; Kaufmann-Hayoz, Kaufmann, & Stucki, 1986) and labels from continuous speech (Jusczyk & Aslin, 1995; Saffran, Aslin, & Newport, 1996) by the time they start forming categories of objects and word types. Through the separate presentation of multiple object and label tokens, both SOMs form categories based on the similarity of the complex set of input patterns. The organisation of the stimuli on the maps mimics the infant’s perceptual refinement of her sensory cortices, based on the unsupervised experience of seeing different objects and hearing different speech fragments. By the time infants are able to engage in joint attentional activities, their perceptual systems are already well-organised. Both the visual and auditory maps have undergone self-organisation, so that when a joint

attentional event occurs, effective associations between the pre-established categories can be formed.



*Figure 1.* Sketch of the network in a joint attentional event. When a dog is presented to the visual map, a coherent activity pattern emerges. Similarly, when an acoustic token of the label ‘dog’ is presented to the auditory map, a selection of neurones will be activated. Synapses connecting the two maps are modulated according to the Hebb rule. The reinforcement of synapses originating from neurones neighbouring the maximally active neurone is a key element in generalising single associations and therefore taxonomic responding.

Joint attentional events are mimicked through the simultaneous presentation of objects and their labels and constitute the supervised component of word learning that is essential for learning the arbitrary mappings between labels and objects.<sup>2</sup> Synapses connecting active neurones on both maps are reinforced through Hebbian learning (Hebb, 1949), as seen in Fig. 1. Due to the topographical organisation of the maps that takes place in early development, many neighbouring neurones on each map will be activated by the presentation of an object and its corresponding sound pattern. Cross-modal Hebbian learning will then take place for neighbouring neurones on each map. Therefore, the association between the paired object and its corresponding sound pattern will be generalised, automatically building associations between all objects in its category to all sound patterns of the appropriate type. A single labelling event is thereby able to induce a taxonomic response with the label extended to all objects of like type: the novel word is learnt.

Note that the organisation of the model does not permit stimuli for modalities outside the separate pathways to influence processing within a pathway. This is another simplifying assumption of the model. There is ample evidence that labels play a role in object category formation from infancy (Plunkett, Hu, & Cohen, 2008) to adulthood (Lupyan, Rakison, &

<sup>2</sup>Note that our definition of a joint attentional event assumes that an object and its appropriate label are presented simultaneously to the visual and auditory maps, rather than spurious combinations of objects and random labels. In principle, the model would learn to ignore such spurious combinations because they are infrequent compared to the correct combinations targeted by the supervised joint attentional events. For simplicity of exposition in the model we highlight only joint attentional events which involve simultaneous activations of objects and their appropriate labels.

McClelland, 2007). The exact role for labels remains, however, subject to debate. For example, an adaptation of the architecture (Glozzi, Mayor, Hu, & Plunkett, 2009) introduced here was shown to capture the experimental data described in Plunkett et al. (2008) to a fine-grained level of detail, even though labels were treated as additional features rather than names by the model. In the present model, we take the approach that names are mapped onto pre-existing non-linguistic concepts, following the view that “this process is considered to be the primary mechanism of lexical development in early childhood as well as an important one thereafter (Clark, 1983; Johnston, 1985; Mervis, 1987; K. Nelson, 1985)” (W. Merriman, Schuster, & Hager, 1991, p.288). In the discussion, we will comment on the limitations of the assumption that labels map onto pre-existing categories in greater detail.

*Potential location of the model in the brain.*

Although the model we propose is an extremely simplified and abstract version of any final, detailed account of the neural mechanisms underlying word learning, we suggest that there is good evidence that the macro-organisation of the model fits the integration of auditory and visual pathways in the human brain. A particularly important assumption of the model is that neural activity produced independently by visual and acoustic stimuli correspond to the activation of mental representations in the brain that are sufficiently proximal to each other that they can readily form Hebbian connections under supervised joint attentional events. We also assume that the individual auditory and visual maps in the model are the product of a long chain of neural computations performed in auditory and visual cortex. Here, we highlight what we take to be some of the essential neural mechanisms for a neuro-anatomically plausible account of word learning.

It is well-established that early visual areas are organised retinotopically (Holmes, 1945). Deeper into the visual processing stream, the visual association cortex receives input from areas of secondary visual cortex as well as from the secondary areas of other sensory systems (Pinel, 2008). The largest single area of association cortex that receives visual input is the posterior parietal cortex (for the dorsal stream, involved in object recognition and form representation), where neurones have larger receptive fields and the stimuli to which the neurones respond are more specific and more complex (Zeki, 1993). Different areas of the association cortex are specialised for particular types of visual analysis. A number of areas of visual association cortex have been identified and in each functional area, neurones have been found to respond most vigorously to different aspects of visual stimuli, each of these areas being laid out retinotopically (Grill-Spector & Malach, 2004). We suppose that the topographical organisation of the SOM in the model corresponds to an abstract template for these retinotopically-organised areas of the visual cortex.

At the end of the dorsal stream, in the posterior infero-temporal and anterior infero-temporal cortices, it has been shown that neurones are selective to parameters relevant to analysis of form and that lesions to infero-temporal cortex cause deficits in discriminations of shape (Gross, Rocha-Miranda, & Bender, 1972; Desimone, Albright, Gross, & Bruce, 1984). In these regions, even though 40% of neurones were found to be nonspecific to visual stimuli, a similar proportion were sensitive to complex objects, such as patterns, hands or faces (e.g., see Maunsell & Newsome, 1987, p.388). Up to a third of neurones near the fundus of the superior temporal sulcus are considered to be face selective (Desimone et al., 1984; Rolls, 1984; D. Tsao, Freiwald, Tootell, & Livingstone, 2006). This is also indicative

of a topographical organisation in association cortex. Moreover, lesions to infero-temporal cortex can induce category specific agnosia such as difficulty in recognizing animals or man-made objects (Schwarzlose, Swisher, Dang, & Kanwisher, 2008), providing further evidence for structure in the representation of higher-order categories.

Adjacent areas in the superior, posterior and lateral parts of the temporal lobes are involved in high-level auditory processing. In humans, the left temporal lobe in particular is specialised for speech processing; including Wernicke’s area, which spans the region between temporal and parietal lobes (along with Broca’s area in the frontal lobe). The functions of the left temporal lobe extend to comprehension, naming, verbal memory and other language functions. Furthermore, it has been suggested that the involvement of Broca’s area in lexical semantics emerges from non-linguistic functions, which are prerequisites for language acquisition (Müller & Basho, 2004). The model therefore provides a simplified account of the interaction between adjacent visual and auditory areas of the temporal cortex. Note that spurious connections between these adjacent regions would lead to an unusual set of cross-modal associations, as seen for example in adult synesthesia (Rouw & Scholte, 2007). Furthermore, the presence of exuberant, cross-modal connections between proximal maps in association cortex prior to synaptic pruning have led some researchers to regard infants as functional synesthetes (Maurer & Mondloch, 2006).

Interestingly, infero-temporal neurons are usually selective for shape (Schwartz, Desimone, Albright, & Gross, 1983; Janssen, Vogels, & Orban, 2000). Early cross-modal mappings could involve neurones from the infero-temporal cortex to high-level auditory processing, thereby highlighting a prominent role for shapes in early categories (Landau et al., 1988).

It should be reiterated that the model presented here is a highly simplified and abstract account of the neural mechanisms that may be underlying word learning. We do not exclude the possibility that other learning mechanisms are involved, such as the complementary learning system approach developed by McClelland et al. (1995), or that a much broader range of brain areas may be involved in word learning, e.g., areas of motor cortex in verb learning. Our purpose here is to point out that the macro-architecture of the model is consistent with our current knowledge of brain organisation and learning.

### *Formal presentation*

#### *Self-organisation of the uni-modal maps: pre-lexical categorisation.*

The uni-modal self-organising maps use the standard Kohonen learning algorithm (Kohonen, 1984). Each map (auditory and visual) consists of an hexagonal grid of 625 neurones receiving acoustic and visual inputs, respectively. Each neurone  $k$  is associated with a vector  $m_k$ . For the presentation of each input pattern  $x$ , the vectors  $m_k$  are modified according to the following procedure: We find the Best Matching Unit (BMU)  $c$ , defined by the condition

$$\|m_c(x) - x\| \leq \|m_j - x\| \quad \forall j$$

where  $\|\cdot\|$  measures the standard Euclidean distance. By extension, we can identify the second best matching unit, the third, and so on. We apply the standard weight update rule with a learning rate that decays over time,  $\alpha(t) = \frac{0.05}{1+t/2000}$  and a Gaussian neighbourhood function of the distance  $d_{ik}$  between neurones  $i$  and  $k$  on the map,  $N(i, k)_t = e^{-d_{ik}^2/2\sigma^2(t)}$ ,

that shrinks linearly over time, from  $\sigma(0) = 4$  to  $\sigma(T_{max}) = 1$ . We define an averaged quantisation error, so that the Euclidian distance between input patterns and their respective BMU is:

$$E = \langle \|x - m_c(x)\| \rangle_x$$

where  $m_c(x)$  is the best matching unit for input pattern  $x$  and  $\langle \cdot \rangle_x$  indicates an averaging over all input patterns. We should emphasise that the quantisation error  $E$  is not a traditional error teacher signal. It is a global measure of weight alignment to the input in the map. In order to shorten simulation time, we used a batch version of the algorithm (Kohonen, 1984) except in those cases where we are interested in evaluating network responses to individual training events. Extensive comparisons between the standard pattern update algorithm and the batch algorithm revealed no qualitative difference in the topographical organisation of the maps. In all experiments, map sizes were fixed to a 25x25 hexagonal grid of neurones.

#### *Coding the inputs.*

Objects are represented as distorted dot patterns. Random dot patterns have been studied extensively by Posner and colleagues (1962, 1964; 1967; 1968, 1970) and have been shown to provide a good approach to the study of concept learning through the psychophysics of form similarity. Moreover, this abstract coding scheme allows us to remain agnostic to the types of attributes involved in category formation, since our goal is to provide a mechanism for taxonomic responding independent of the features involved in category formation itself. For example, functional aspects or other non-perceptual or non-obvious object attributes can contribute to category formation. The only assumption we make is that categories possess a “family resemblance structure” (Rosch & Mervis, 1975), a structure that has been shown to reflect the structure of basic-level categories of objects which constitute the main part of the infant’s early lexicon (Anglin, 1977; Horton & Markman, 1980). Basic-level categories have been shown to minimise intra-category variance while maximising inter-category distances (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Mervis & Rosch, 1981; Murphy & Brownell, 1985). Both at the perceptual level (basic categories have a common shape, Rosch et al., 1976; Landau et al., 1988) and at a functional level (basic categories have more movements in common than superordinates, Rosch et al., 1976), the family resemblance structure offers a suitable approximation to the organisation of categories in early infancy. The random dot categories used in the model reflect the sparseness of representation that is thought to underly the organisation of real world categories (Murphy, 2004). As we shall see, the sparseness of these input representations are particularly well-suited to the capacity of SOMs to identify the *natural clustering* of objects in the input space. Note, however, that the categories of random dot patterns do not make up a hierarchy of categories such as one might find in the animal world (labrador  $\rightarrow$  dog  $\rightarrow$  mammal  $\rightarrow$  animal). This will have implications for our interpretations of over-extensions when we come to consider the errors in the performance of the model. We created 100 patterns, or prototypes (Rosch, 1973), by randomly distributing 9 dots over a 30x30 square. We then generated 24 tokens in each category, consisting of 8 tokens at each of 3 levels of distortions. Distortions are achieved by moving each dot by an amount drawn from a normal distribution with a standard deviation of 1, 2 and 3 respectively. Prototypes were not included in the training set.

There is little consensus in the field as to what acoustic information infants use when identifying words. A series of studies emphasise the fact that infants pay attention to much more than simple phonological features. In particular, it has been shown that at 2 months they are able to make allophonic distinctions (Hohne & Jusczyk, 1994), at 9 months of age, infants are sensitive both to stress and phonetic information (Jusczyk, 1993) and by 17 months, they pay attention to co-articulation (Plunkett, 2006). All of these sensitivities to the speech signal may have an important impact on early lexical development. Therefore, we exploit the whole acoustic signature of tokens in order to avoid discarding relevant acoustic information. In a preliminary study, we measured the variance in the sound patterns both across different speakers (referred to as “high variance”) and within repetitions of words by the same speaker (“low variance”). We then extracted the acoustic signature from raw speech waveforms for 100 acoustic categories (words) produced by a female native speaker. We generated 24 acoustic signatures for each acoustic category, centred around the recorded speech signals. Each category consisted of 80% low variance modifications of the prototype, mimicking the high prevalence of the mother’s speech, and of 20% high variance modifications, thereby providing the model with a variance in word pronunciation introduced by different speakers. Word tokens were then normalised in length and sampled at regular intervals, 4 times per syllable<sup>3</sup>. To a first approximation, this procedure provides a stable sound representation even when different speakers pronounce words at a different speed. After sampling, the sounds were filtered using the Mel Scale in order to approximate the sensitivity of the human ear. We decided to sample seven frequency bands to maintain a diverse source of information about the speech signal. Input vectors were concatenations of these four 7-dimensional mel-cepstrum vectors, derived from FFT-based log spectra<sup>4</sup>.

*Training the cross-modal associations; joint attentional activities.* We mimic joint attentional activities between the caregiver and the infant by presenting simultaneously to both maps a randomly picked object from the data-set and an acoustic token randomly picked from within the matching category (see Fig. 1). We build cross-modal associations by learning Hebbian connections between both maps. As a further simplification of the model, we use bidirectional synapses whose amplitudes are modulated by the activity of the connecting neurones. We define the neural activity of a neurone  $k$  to be  $a_k = e^{-q_k/\tau}$  where  $q_k$  is the quantisation error associated with neurone  $k$  and  $\tau = 0.5$  is a normalisation constant.

It is known that cortical synaptic density is not constant during life. There is first a process of synapse genesis, a maturational process in which neurones make contact with other neurones. There are essentially two mechanisms for synapto-genesis in the human brain; one is random formation of synapses followed by functional stabilisation and the other is the formation of neuronal circuits by intrinsic mechanisms such as bidirectional signalling (P. Huttenlocher, 2002). Whereas the latter mechanism is observed in brain

<sup>3</sup>We found that for monosyllabic words, having 3 samples per syllable is sufficient from the point of view of word-object generalisation performance as described in the results section. We found no statistically significant improvement when increasing the number of time-slices beyond  $N = 3$ .

<sup>4</sup>The mel-cepstrum vectors are obtained by applying the following procedure: take the Fourier transform of a windowed excerpt of the signal, map the log amplitudes of the spectrum obtained onto the Mel scale, using triangular overlapping windows and finally take the Discrete Cosine Transform of the list of Mel log-amplitudes, see Davis and Mermelstein (1980) for detailed information.



regions where environmental input is less central, such as for controlling walking or breathing, the former is thought to be the main mechanism of synapto-genesis in systems where learning is important. “Randomly formed synapses appear to form a substrate for the development of circuits that are dependent on environmental input. The ready availability of billions of unspecified synaptic contacts in the immature cerebral cortex may be important for the formation of the synaptic circuits underlying the development of higher cortical functions, including mathematical skills, musical ability, and language functions.” (P. Huttenlocher, 2002, p.41). We will therefore model “blind” synapto-genesis, the process of forming synapses at random locations (Bourgeois & Rakic, 1993) by linking together only a percentage of randomly picked neurones on both maps, a percentage that increases linearly with age, from 1% connectivity to full connectivity after 500 epochs of training.

After the end of the synapto-genesis phase, the number of synapses peaks at an age that varies from one brain structure to another. For lower level brain structures, synapto-genesis takes place much earlier than for higher cortical areas. The emergence of visual functions occurs in the first months of life, at the same time as synapto-genesis in the visual cortex. For example, the ability to see objects in three dimensions, stereopsis, emerges at the exact time of rapid synapto-genesis in the primary visual cortex, between the ages of 3 and 5 months (Wilson, 1988). Synapto-genesis in language-related areas takes place much later in development. P. Huttenlocher (2000) describes the developmental trajectories of synaptic densities in the auditory cortex (Herschl’s gyrus), in receptive language areas (Wernicke’s area) and in productive speech areas (Broca’s area). Interestingly, the timing at which synapto-genesis takes place reflects the sequence of functional development in these regions. Synaptic densities peak at around 2-3 years of age in Wernicke’s and Broca’s areas, much later than the age at which synapto-genesis occurs in early visual and auditory areas.

We mimic these developmental differences in synapto-genesis between cortical areas by implementing synapto-genesis between the maps but not within the maps, reflecting the correspondence between early and late synapto-genesis for lower and higher cortical organisation, respectively. At the time word learning takes place, primary cortices have already undergone functional stabilisation. A second reason for implementing changes in synaptic densities across the maps but not within the maps themselves comes from neuro-anatomical studies. The number of synapses between two neurones has been found to decrease systematically with increasing distances in many different brain areas. For example, in the rat visual cortex the average number of contacts between pre- and post-synaptic neurones was found to be about 6 times larger when separated by a few microns compared to a distance of 200 microns (Hellwig, 2000). In the cat visual cortex, the number of somatic connections peaks at around 60 microns (with an average of 40 post-synaptic somata contacted by an axon) and decreases to about 5 connections when neurones are located at 300 microns from each other. Likewise, in the rat somato-sensory cortex, the average number of contacts between neurones was found to decrease as their separation increases (Kalisman, Silberberg, & Markram, 2003). This progressive reduction in the number of synaptic contacts between two neurones as their distance increases is well captured by theoretical models (Amirikian & Friston, 2005) and is assumed to occur across all brain areas and in all species. As a consequence, synapto-genesis and synaptic pruning will have a greater impact on two neurones that are distant from each and therefore have a small number of synapses connecting each other. The addition/removal of one synapse would effectively add/remove a direct

connection for such neurones. In contrast, locally-connected neurones are densely connected and even strong synapto-genesis and pruning would only modulate the magnitude of their interaction. In our model, long-distance connections correspond to those between the maps whereas local connections are those within the maps. Synapto-genesis and pruning is therefore limited to the connections between the maps.

Later in development, the number of synapses decreases by a process of elimination of *weak* synapses. This synaptic pruning mechanism is driven by experience (Stryker & Harris, 1986; Roe, Pallas, Hahm, & Sur, 1990) and is thought to minimise energy consumption<sup>5</sup> (Roland, 1993; Feinberg, Thode Jr, Chugani, & March, 1990) as well as optimise network properties (Miller, Keller, & Stryker, 1989; Chechik, 1998). This full process of “blind” synapto-genesis and experience-dependent pruning is summarised by Bourgeois (1993): “the initial production of synapses is controlled intrinsically while their stabilization or selective elimination after birth is determined by complex epigenetic mechanisms”. The development of the human cortex is largely shaped by input to the system and a natural by-product of this process is the large number of functionally unspecified synapses, that may disappear later in development. Based on differences in synapto-genesis and pruning in human and other primate cerebral cortex, some researchers conjecture that the unique human mastery of higher cognitive functions may lie in the unique developmental trajectories of synaptic connectivity patterns (P. Huttenlocher, 2002). Quartz and Sejnowski (1997) have also emphasised the importance of the balance between synapto-genesis and pruning mechanisms for the development of higher level cognitive functions. It is also known that the total number of neurones decreases over time (Abitz et al., 2007). However we do not address this issue in the model. Nevertheless, the neuro-anatomical data points strongly to a role for synapto-genesis and pruning in higher cognitive functions and therefore we choose to implement it as an important feature of the model.<sup>6</sup>

All  $S$  synapses between the maps were first randomly initialised with a normal distribution centred on  $1/\sqrt{S}$  and with a standard deviation of  $\frac{1}{\sqrt{1000*S}}$ . Synapse amplitudes are modulated according to a standard Hebb rule with saturation. Therefore, synapse weights stay in a physiological range even for high neural activities. The synapse connecting neurone  $i$  from the visual map to neurone  $j$  of the auditory map is computed as follows:

$$w_{ij}(n+1) = w_{ij}(n) + 1 - e^{-\lambda a_i a_j}$$

where  $n$  refers to the index of the word-object pairing and  $\lambda = 10$  is the learning rate. The free parameter  $\lambda$  can be chosen over a wide range of values with qualitatively similar

---

<sup>5</sup>Brain metabolism has two main energy expenses; the signalling part, associated with an *active brain* state and the maintenance of neural circuitry, including for example protein trafficking at synapses, as they are renewed all the time (Attwell & Laughlin, 2001). The exact energy costs associated with each part is difficult to estimate. However, the analysis of brain consumption in coma was estimated to be half of its normal consumption (Siesjö, 1978; Ames, 2000), indicating that the maintenance part is non-negligible. Even the suppression of silent, or non-functional, synapses would therefore lead to a reduction of brain metabolic costs.

<sup>6</sup>Recently, Mareschal et al. (2007) have advocated a neuro-constructivist approach to cognitive development that emphasises the impact of neuro-genesis on the emergence of cognitive structure. Although we do not subscribe to the particular implementation of neuro-genesis in Mareschal et al.’s model (recruitment of new hidden units via an error-driven criterion similar to Fahlman’s (1990) cascade-correlation learning algorithm), we situate our work within the more general framework of neuro-constructivism advocated by Mareschal et al. (2007) and Quartz and Sejnowski (1997).

results. After every word-object presentation, weights are normalised so as to model limited synaptic resources:  $\sum_{ij} w_{ij}^2 = 1$ .

*Assessment of taxonomic responding.* In order to assess the performance of the model, it is necessary to introduce two definitions. The first definition introduces a measure of the magnitude of taxonomic responding. It may be interpreted as the learnability of a word. The second defines a criterion for deciding when a word is learnt. All results refer to these measures.

**Definition 1** *We will refer to the taxonomic factor as being the percentage of correct word-object associations, following one single labelling event per category.*

The taxonomic factor gives us an insight into the word-object generalisation properties of the network based on a single joint attentional event per category. As mentioned above, generalisation in classical, single-mechanism, multilayer, backpropagation, neural networks is controlled through the presentation of multiple word-object mappings. These word-object pairings must sample the input space widely to define the boundaries of individual categories. After multiple supervisory learning events, backpropagation networks can learn to identify the features relevant to category formation and discard irrelevant aspects of the input data. Traditional, single-mechanism, multi-layer, backpropagation, neural networks, including auto-encoders, would have a very low taxonomic factor.

In contrast, the model we propose first learns to categorise objects or labels in an unsupervised fashion. Through unimodal presentations, the maps self-organise and form categories out of a complex input space. All category boundaries are defined during the unsupervised learning phase. Object categorisation and word-form categorisation are thereby determined during unsupervised learning activities in the model. In the supervised phase of learning, single cross-modal labelling events need only to bind the categories on each map. The joint attentional events themselves provide the supervision needed in order to generalise accurately word-object associations. Word learning therefore corresponds to a supervised, cross-modal learning activity in the model.<sup>7</sup>

The taxonomic factor is a global measure of the networks' generalisation properties based on a single joint attentional event per category. It is also possible to provide a more detailed, microscopic picture of individual word acquisition in the model, following multiple joint attentional events. We assume that word learning involves more than simple associations of a visual token and an acoustic token. It is known that neonates (Morrongiello, Fenwick, & Chance, 1998) and dogs (Adachi, Kuwahata, & Fujita, 2007) can learn simple cross-modal associations, but few would consider these instances of genuine word learning.

---

<sup>7</sup>Note that it would also be possible to train two auto-encoders using backpropagation to learn about object and word-form categories in a uni-modal fashion and then "fast map" the hidden unit representations of each auto-encoder using a Hebbian style learning algorithm. This approach is very similar to the strategy adopted in the current modelling endeavour, insofar as the auto-encoders, like the SOMs, would use the statistical regularities in the input stream to discover the categorical structure of the learning environment, so that Hebbian associations can rapidly bind the acoustic and visual categories together. However, the approach differs in that auto-encoders require an error-correction learning algorithm. Furthermore, we shall see that the distributed representations produced by auto-encoders across hidden unit layers are not suitable for the application of efficient pruning mechanisms, that can improve the representations of label-object associations.

We follow Nazzi & Bertoncini’s (2003) proposal that real words (at least count nouns) involve a mapping between an object category and a well-defined speech pattern (the label). We therefore set a threshold for generalisation (analogous to Brown’s (1973) obligatory contexts) as our definition of successful word learning.

**Definition 2** *We will consider a word to be part of the lexicon if more than 80% of the objects of the appropriate category are associated with the correct label.*<sup>8</sup>

After training on cross-modal pairings, we assess the capacity of the network to extend the association of a presented word-object pair to non-paired items that belong to the same category. Following the presentation of the word-object pairs, weights are fixed. All objects in the data-set are then classified according to whether the induced activity on the auditory map corresponds to the activation of the appropriate label (corresponding to the minimal Euclidian distance). One of the outcomes of this procedure is that it implements a forced choice among all labels known to the network. A unique label is activated with the presentation of a given object token. Similarly, acoustic tokens are classified according to the activity induced on the visual map. The average of both classification measures is the taxonomic factor. When training the model on more than one labelling event per category, we also report the total number of words learnt. All results reported are averaged over 35 independent simulations, unless otherwise stated.

## Results

We present a series of experiments highlighting contributing factors to successful word learning in the model and the role of map structure for taxonomic responding. This is achieved by varying the point of onset of joint attentional activities in the network relative to the initialisation of unsupervised training of individual maps. The taxonomic factor is measured at different points in the training of the network by introducing a single labelling event per category and measuring generalisation of the event to other category members. We show that only when uni-modal maps are well-structured does the system respond in a taxonomic fashion. The discrete nature of the categories thus formed is illustrated by the presence of categorical perception effects on the maps.

Another set of experiments investigates the role played by synapto-genesis and synaptic pruning for word learning in the model. The findings indicate that early synapto-genesis supports an increase in taxonomic responding and that late pruning maintains taxonomic responding while lowering the metabolic costs associated with the synapses. Further analyses show that joint attentional events enhance identification of the candidate synapses for pruning. These simulations demonstrate how the network can account for experimentally observed developmental patterns such as the gradual reduction of over-extension errors, both through map structure improvement and joint attentional activity. It also provides an account for the observed shift from a prototype-based to an exemplar-based categorisation behaviour, despite relying on a single mechanism. Typicality effects are also considered.

In a final set of experiments, we assess the impact of the frequency of joint attentional events on the profile of vocabulary development in the network. The network is trained

---

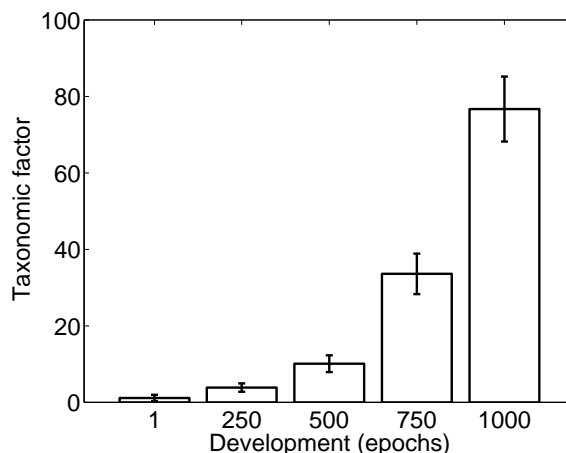
<sup>8</sup>We have run extensive simulations with different threshold criteria and found the qualitative performance of the network to be independent of the exact value for the threshold.

under two conditions: a high versus a low level of joint attentional activities. The rate of lexical acquisition undergoes a rapid increase in both conditions. However, the networks experiencing low levels of joint attentional activity have a delayed vocabulary development and lower overall sizes of vocabulary.

*The role of pre-lexical categorisation capacity*

After  $n$  epochs, each map has been trained  $n$  times on all images and words in the data-set. Simultaneously, the number of synapses between the maps increases linearly from 1% random connectivity at the start of learning to full connectivity at epoch 500, mimicking the “blind” synapto-genesis occurring from birth to early childhood.

We assess the impact of introducing joint attentional activities at different points in development by measuring the taxonomic factor after a single labelling event per category, when the maps have been trained for  $n$  epochs. Fig. 2 depicts the taxonomic factor, when joint attentional events are initialised at various points in training. Note that the taxonomic factor also approximates the likelihood that a word is learnt following a single labelling event.



*Figure 2.* Role of pre-lexical experience. As the maps gain experience with the visual and acoustic environment, taxonomic responding increases.

Fig. 2 shows that when joint attentional activities are introduced early in training, the network fails to generalise object-label associations to other members of the category. A high level of taxonomic responding is achieved only towards the end of training, when the auditory and visual SOMs have had substantial ( $>750$  epochs) experience of individual tokens of objects and words. The increase in taxonomic responding displays a distinct non-linear trend around 500–750 epochs.

This result suggests that the ability of the maps to form categories of objects and word types is crucial to successful generalisation. We demonstrate this association between the quality of the map structure and taxonomic responding by comparing two measures of the quality of map structure to taxonomic responding; the quantisation error  $Q_E$  and the topological error  $T_E$ . For  $Q_E$ , the quality of map organisation is indexed by a measure of weight alignment – the degree of correlation between the input vectors and the weight

vectors for the corresponding BMUs. A low value of  $Q_E$  corresponds to a high map quality, as the weight vectors of the BMUs match the input patterns. In contrast, a high value of  $Q_E$  indicates that the map has not aligned its internal structure to the input patterns. Quantisation error is a direct reflection of the network’s experience with input patterns.

The second index of map quality, the topological error  $T_E$ , measures the map coherence by computing the proportion of all data vectors for which the first and second BMUs are not adjacent neurones.  $T_E$  is an index of topology preservation such that low values of  $T_E$  indicate tight clustering of neurones that are BMUs from the same category whereas high  $T_E$  indicates that BMUs from the same category are dispersed on the map. Fig. 3 shows the relationship between the quality of map structure and taxonomic responding. The left panel

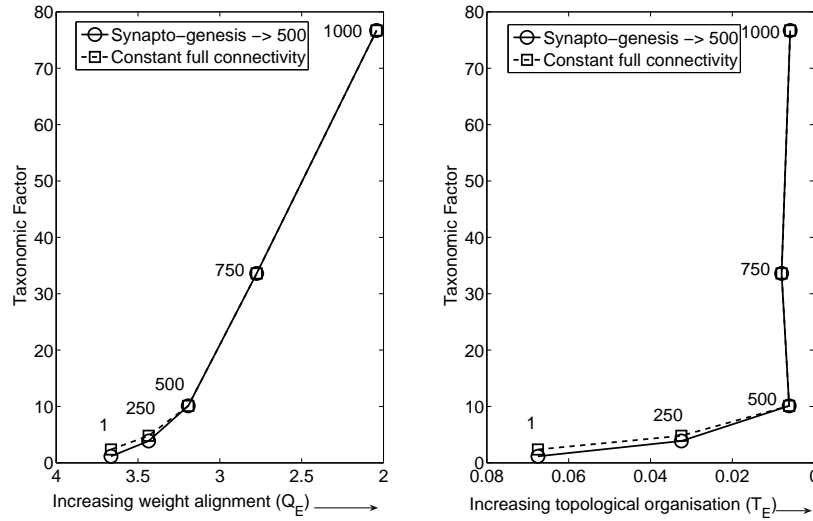


Figure 3. Impact of pre-established categorisation capacity on taxonomic responding. The taxonomic factor is a monotonic function of the quality of map structure.

of Fig. 3 depicts the average  $Q_E$  for the visual and auditory maps. The right panel of Fig. 3 depicts the average of the  $T_E$ . The left panel indicates that taxonomic responding improves monotonically with decreasing quantisation error, demonstrating that network generalisation improves with increasing weight alignment of input patterns with their BMUs. The right panel of Fig. 3 demonstrates the improvement in taxonomic responding with increasing topological organisation on the maps. Note that topological organisation is almost optimal by 500 epochs of training, indicating that input patterns from the same category activate neighbouring neurones. Further improvements in generalisation are achieved by further tuning of the weight alignment between input patterns and their BMUs (epochs 500–1000 in the left panel of Fig. 3).

Together, these two measures of map quality provide an insight into the role of map structure for taxonomic responding. Overall, generalisation of word-object associations in the network is a monotonic function of the quality of map structure. Taxonomic responding builds on the pre-established categories formed on both the auditory and visual maps. Note that the time course in the growth of the taxonomic factor is not qualitatively different

when the maps are fully-connected throughout development (Fig. 3 dashed line) than when there is a linear increase in connectivity (Fig. 3 solid line). This suggests that the fact that improvements in taxonomic responding coincide with the point when synaptic density reaches its maximum (at 500 epochs) is accidental, since a constant full connectivity still exhibits the same pattern of responding across training points. It is, therefore, ultimately the quality of map structure – corresponding to the pre-lexical categorisation skills in infants – that determines the emergence of taxonomic responding. At the same time, fast mapping abilities between word types and object categories emerge with taxonomic responding: A single labelling event leads to robust learning and accurate generalisation if and only if map structure is well-formed.

### *Categorical perception*

We now explore the nature of the category representations on the maps. In order to determine whether the objects are grouped in a categorical fashion, we identified all neighbouring visual prototype pairs in the object space. Along each of the axes connecting the neighbouring prototypes, six new objects were created, regularly spaced between the two prototypes. The new objects were presented to the network and, for each of them, the preferred category, based on the induced activation on the auditory map, was identified. This experiment involved presenting 700 novel visual stimuli to the network.

If the space of object representations has adapted to form categories, the system should exhibit a categorical perception effect; objects closer to one prototype should be unambiguously assigned the label corresponding to that prototype whereas objects closer to the other prototype should receive the alternative label. Furthermore, a pair of objects within a category should be more difficult to discriminate than a pair of objects that are equally far apart but which span a category boundary. Fig. 4A depicts the choice of labels (the identification function) associated with the new objects distributed along the axes connecting neighbouring prototypes. Distances between the prototype pairs are normalised, so that the network’s behaviour for the different pairs can be evaluated on the same scale, independently from the individual distances between prototype pairs. The black circles identify the proportion of the new objects that are associated with the label of the first prototype in the pair. The white squares measure the proportion of the new objects that are associated with the label of the second prototype in the pair. The pattern of categorisation of the new objects exhibits a typical sigmoidal identification function characteristic of categorical perception (Calder, 1996). Objects are unambiguously identified as associated with one of the category labels or the other. Note that proportions are averaged over all nearest-neighbouring prototype pairs. Therefore, the transition is slightly smoothed since some categories take more space on the map than others. The transitions for individual pairs display an even stronger nonlinearity. Figure 4b depicts the discrimination curve for pairs of objects that are equally distant. When both objects are within a category, they are not easily discriminated (the Euclidean distance between their representations on the map is small). However, when the two objects span a category boundary, the discrimination capacity improves, as their map representations differ considerably. Importantly, the network forms clear categories from which associations can be built and generalised with accuracy, even close to the transitions between category boundaries.

Note that all the objects presented in this experiment are new to the network, includ-

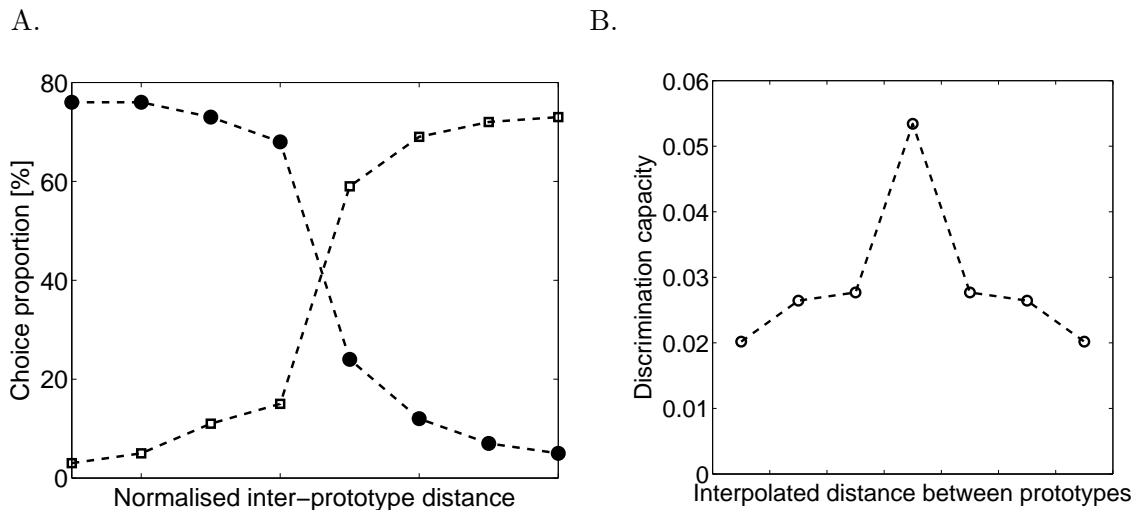


Figure 4. (A) Choice of labels associated with objects that are the product of interpolating nearest neighbouring prototypes, as a function of their normalised distance along the interpolation line. The black circles measure the proportion of objects that associate with the label of the first prototype and the white squares measure the proportion of the objects that are associated to the label of the second prototype. (B) Discrimination of equally distant pairs of objects. Objects that span a category boundary can be more easily discriminated than objects within the same category.

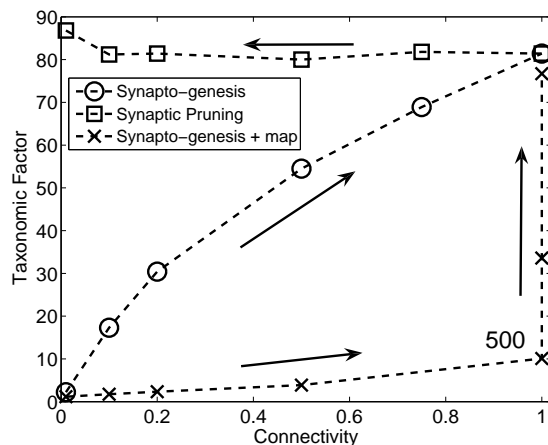
ing the prototypes and the interpolated objects themselves. The categorical responses of the network to the novel visual stimuli underscore the robust generalisation characteristics of object-word associations for any items that fit the category space that has emerged during pre-lexical development.

#### *The role of synapto-genesis and pruning*

In early childhood, the number of synapses increases during the first 2 years of life (P. Huttenlocher, 1979). In cortical regions important for language processing such as Wernicke’s and Broca’s areas, synaptic density reaches a maximum at around 3 years of age (P. Huttenlocher, 2002). The process of synapto-genesis is determined genetically and is essentially independent of environmental factors. The number of synapses remains constant until about 10 years of age, after which the number of synapses decreases, as a result of synaptic pruning. Weak synapses are eliminated so as to keep the most relevant synapses. We model both synapto-genesis and synaptic pruning in order to identify their possible impact on word learning capacity. Synapto-genesis is modelled by increasing the number of *random* connections between the maps. Once maximal connectivity is reached (full connectivity in our model) the weakest synapses are eliminated. We assess the capacity to generalise word-object associations at different points in the processes of synapto-genesis and synaptic pruning. Fully-developed maps and maps that undergo structuration are compared in these experiments. Generalisation is assessed on the basis of a single presentation of a word-object token pair per category, as in Fig. 2.

Fig. 5 shows the taxonomic factor as a function of the synaptic connectivity between the maps, both during synapto-genesis and synaptic pruning. During synapto-genesis, there





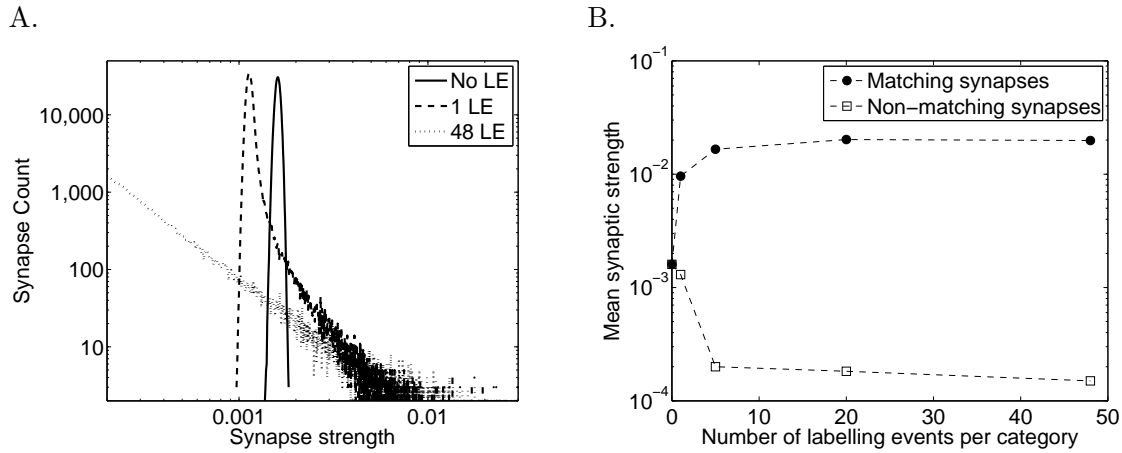
*Figure 5.* Impact of synapto-genesis and pruning on word learning capacity. In early infancy the number of synapses increases independently of the statistical regularities found in the environment. Generalisation of word-object associations improves with the number of randomly generated synapses, both for fully structured maps (o) and with maps that evolve over time (x). Later, when synapses are pruned according to their strength, the ability to learn a word remains constant and can even improve at higher levels of pruning. The full process of synapto-genesis and synaptic pruning minimises metabolic costs and optimises word learning capacity.

is a monotonic increase in taxonomic responding as the number of synapses grow. Fig. 5 displays the impact of synapto-genesis for both the fully-structured maps (o) and for maps that evolve over time (x). Note that for evolving maps full connectivity is achieved after 500 epochs. In both conditions, with very few connections between the maps at the start of word learning, generalisation is difficult because word learning is restricted to just those word-object pairings for which connections exist. The likelihood of a connection existing between a given word-object pairing at 1% connectivity is just 1%. Even if the connection exists, the network fails to generalise well because of the sparsity of connections between other members of the word-object mapping category: the ability to generalise depends crucially upon spreading activity within a map propagating to the relevant category in the other modality. If the connections do not exist, this cannot occur. In the fully-structured map condition, generalisation of word-object associations increases with higher density of connectivity. The increased likelihood that a connection exists between an untrained word-object pairing leads to improved generalisation. Fig. 5 also plots the developmental trajectory of the taxonomic factor when maps becomes structured at the same time as they increase in synaptic connectivity (x) up to 500 epochs, when synaptic density reaches a maximum. The poor quality of map structure early in learning further undermines taxonomic responding, resulting in a catastrophic failure of generalisation.

Once the system has established a full cohort of cross-modal synapses, the pruning process can begin. Fig. 5 shows the impact of pruning the weakest synapses, with fully-developed maps. Synaptic pruning is assessed by removing 25%, 50%, 80%, 90% and 99% of the synapses between the maps in order of ascending strengths, respectively. The taxonomic factor remains constant throughout the pruning process and even improves for

a connectivity of about 1%, when the normalisation procedure is not penalised by scaling irrelevant weights. This level of pruning corresponds the optimal configuration of synapses where each neurone on one map is connected to, and only to, all the the neurones of the matching category in the other modality. In other words, pruning ensures that each token of an object category is connected to all and only the instances of the appropriate label type, and each token of a label is connected to all and only the instances of the appropriate object kind.

It is noteworthy that synaptic pruning in human learners does not take place before individuals have experienced a very large sample of word-objects associations, at around 10 years of age. By this time, they will have experienced many millions of joint attentional events. We, therefore, investigate the impact of the level of joint attentional activity in the network on the application of the synaptic pruning mechanism. Fig. 6A depicts the



*Figure 6.* A. Distributions of synapses connecting the two well-structured maps, before any labelling event (LE) (solid line), after one labelling event per category (dashed line) and after 48 labelling events per category (dotted line). Labelling events flatten the distribution of synaptic strengths so that weak synapses are further weakened. In contrast, a subset of synapses increases in strength. B. Mean synaptic strengths for the synapses connecting BMUs of matching categories versus other synapses. Labelling events have a direct impact on synaptic strengths by supporting a clear cut distinction between matching synapses and non-matching synapses, facilitating the application of a pruning mechanism.

impact of labelling events on the distribution of synaptic strengths between the maps, when maps are well-structured, a likely scenario when the pruning mechanism begins to operate. Before any labelling event takes place, the distribution of synaptic strengths is compact (solid line), and normally distributed, corresponding to the random initialisation of weights in the learning process. After one labelling event per category, the distribution of cross-modal synapses is skewed (dashed line): a subset of synapses is strengthened whereas the rest of the synapses are weakened, through the normalisation procedure. After 48 repetitive labelling events per category (dotted line), the distribution of synaptic strengths is flattened. Weak synapses have a near-zero strength whereas strong synapses remain. Fig. 6B plots the mean values of synaptic strengths, for synapses connecting BMUs of matching categories and other synapses. Fig. 6B demonstrates that increasing the number of labelling events

forces the network to heighten the distinction between matching and non-matching synapses, strengthening the matching synapses and weakening the non-matching ones. By so doing, labelling events prepare the system for pruning the connections in an efficient and reliable manner. Once the distribution of synaptic strengths is stable, the choice of which synapses to prune is facilitated and can be done without affecting word learning performance. Moreover, the reduction in the number of synapses minimises metabolic costs. This finding indicates that prolonged experience of word-object associations can refine underlying representations even though no new associations have been acquired.

### *Evaluating errors in the model*

Although the network is able to generalise single label-object associations in a taxonomic fashion after object and label categories are well-formed on the maps, the same cannot be said for the network during earlier stages of training. After limited exposure to the visual and object environments, the network is prone to generalise newly acquired label-object associations in an inappropriate manner. For example, labels may be used to identify objects outside of the appropriate category. In the child language literature, inappropriate use of labels in this manner is referred to as “over-extension” (Barrett, 1978; Rescorla, 1980; J. Huttenlocher & Smiley, 1987). Early over-extension errors in production are thought to be based on featural similarities (perceptual, functional, etc.) between objects. For example, ‘cat’ may be over-extended to dogs because they are furry or are four-legged. Likewise, ‘ball’ might be over-extended to rolling objects because of shared functional features.

Over-extension errors are contrasted in the literature to mismatches or random errors where the child labels an object in a spurious manner that does not appear to reflect any obvious featural similarity. It is important, therefore, to examine whether the distribution of over-extension errors in the model parallels the feature-based characteristics of early over-extension errors in children. However, since the model does not exploit discrete features, we investigate the role of object similarity in the input space in determining the distribution of over-extension errors. Furthermore, recall that the training objects that populate the input space are not organised in a hierarchical fashion. Clusters of random dot figures generated from individual prototypes populate the input space in a heterogeneous manner, i.e., some clusters are closer than others but there is no simple hierarchical structure. As a consequence, it is not possible in the model to draw a dichotomous distinction between similarity-based over-extension errors and random errors. The structure of the input environment forces us to analyse the distribution of errors in the model in comparison to the distribution that would be obtained by randomly assigning labels to objects. Similarity-based errors can be detected by determining whether the distribution of errors deviates from chance or not.

The structure of the model allows us to examine a variety of error types. It is important to clarify how we interpret the distinctions between them. We consider a *production* event to occur when a pattern of activation occurs on the *auditory* map as result of the presentation of a visual object to the visual map and subsequent propagation of activity through the Hebbian connections to the auditory map. A *comprehension* event occurs when a word is presented to the auditory map resulting in a pattern of activation on the *visual* map via the cross-modal Hebbian connections.

Production errors result from inappropriate connections from the visual to the audi-

tory map. However, these mis-mappings can have two sources: Objects may be confused on the visual map or words may be confused on the auditory map. We analyse over-extension errors in production as resulting from confusions on the visual map. The model calls a cat a ‘dog’ because of the confusion between the representations of the objects cat and dog which are instantiated on the visual map. We take this interpretation of over-extension errors in the model to be a reasonable proxy to its usage in the child language literature. We distinguish over-extension errors in production from confusions on the auditory map which are akin to “slips of the tongue” though, of course, the model has no tongue. For example, the model might call a dog a ‘doll’ because of the proximity of the representation of ‘dog’ to ‘doll’ on the auditory map.

Comprehension errors can also derive from confusions on either the visual or auditory maps. We analyse over-extension errors in comprehension as resulting from confusions on the visual map. The model misunderstands the word ‘dog’ if it activates an inappropriate object representation on the visual map, such as a cat. Again, we take this interpretation of over-extensions in comprehension in the model to be a first approximation to its usage in the child language literature. In contrast, confusions on the auditory map during comprehension events can be interpreted as insensitivity to mispronunciations of words, such as confusing the acoustic tokens ‘dog’ and ‘doll’. Thus, hearing the word ‘doll’ might lead the model to incorrectly identify the object dog. In summary, over-extension errors arise from confusions on the visual map. Confusions on the auditory map can be interpreted as insensitivity to mispronunciations of a word (in comprehension) or slips of the tongue (in production). The literature on lexical errors in early language development has most commonly described over-extension errors in production. Therefore, we shall describe the model’s performance in this respect first.

*Over-extension errors in production.* The distribution of correct and incorrect word-object associations in production is determined by presenting every object to the visual pathway of the network and identifying the corresponding label category activated on the auditory map, following a single labelling event in each category, at different points in training. We measure the Euclidian distance between the object prototype associated with the produced label and the prototype of the correct object. For example, if the image of a cat is presented and the label ‘dog’ is activated, we measure the distance between a prototypical cat object and a prototypical dog object. If the associated label is within the category boundary of the appropriate label, we define the Euclidian distance as 0. Otherwise, a small Euclidian distance indicates that the label is over-extended to a similar but different object category (e.g., cat for dog) while a large Euclidian distance corresponds to an incorrect label category whose object is dissimilar from the target category (e.g., flower for dog). This procedure permits us to identify over-extension errors in production at different points in training and thereby evaluate the model’s capacity to mimic infant over-extension errors.

An accurate description of over-extension errors in the model can only be provided at the level of the distribution of errors, not at the level of individual erroneous mappings. An examination of the developmental trajectory of over-extension errors in production is made at different points in training (1, 500, 750 and 1000 epochs). At all stages of map training, the average distance of over-extension errors from the target is well below the

average inter-prototype distance, indicating that over-extension errors tend to be used for objects that are more similar to the correct object ( $p \ll 0.00001$ ). So as to compare the model's behaviour to experimental data, we simplify the analysis by assuming that when the distance between the associated prototype and the correct prototype is less than the average inter-prototype distance, it is considered to be a (similarity-based) over-extension. Fig. 7 depicts the number of similarity-based over-extensions as a function of the number

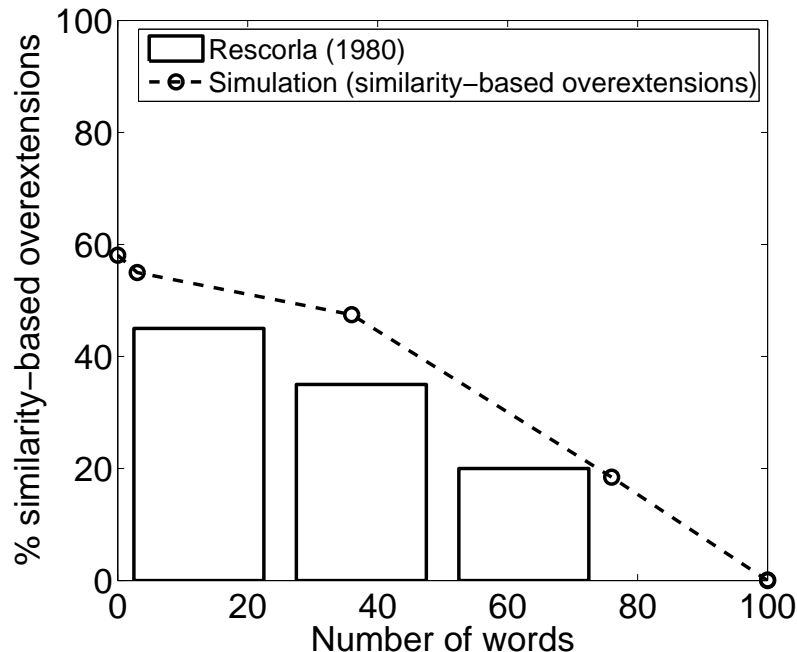


Figure 7. Percentage of similarity-based over-extensions as a function of the number of words (in production) in the lexicon, for the model ( $\circ$ ) and from data published by Rescorla (1980) (bars).

of words in the model's lexicon (circles), as defined in Definition 2. The fraction of words that are over-extended to similar objects decreases regularly from about 60% for very small lexicons (less than 10 words) and disappears when the model has learned all words in the lexicon (when additional labelling events have eliminated all errors in production).

The child language literature emphasises the overwhelming importance of object similarity in constraining over-extension errors in production. For example, J. Huttenlocher and Smiley (1987) suggest that similarity-based over-extensions outnumber low-similarity errors by a factor of 7 to 1 (also see Kay & Anglin, 1982). A similar pattern is described in Rescorla (1980) who also provides a developmental trajectory for over-extension errors in production, as shown in Fig. 7 (bars). In this study, the percentage of similarity-based over-extensions (simply referred to as over-extensions in Rescorla, 1980) decreases from about 45% for infants having a vocabulary size smaller than 25 words, to 35% when infants possess a lexicon size of 25-50 words and 20% when their lexicon reaches 75 words. It is noteworthy that the precise figure for the fraction of the infant's lexicon being over-extended is not undebated and a series of publications have reported different levels of over-extensions at

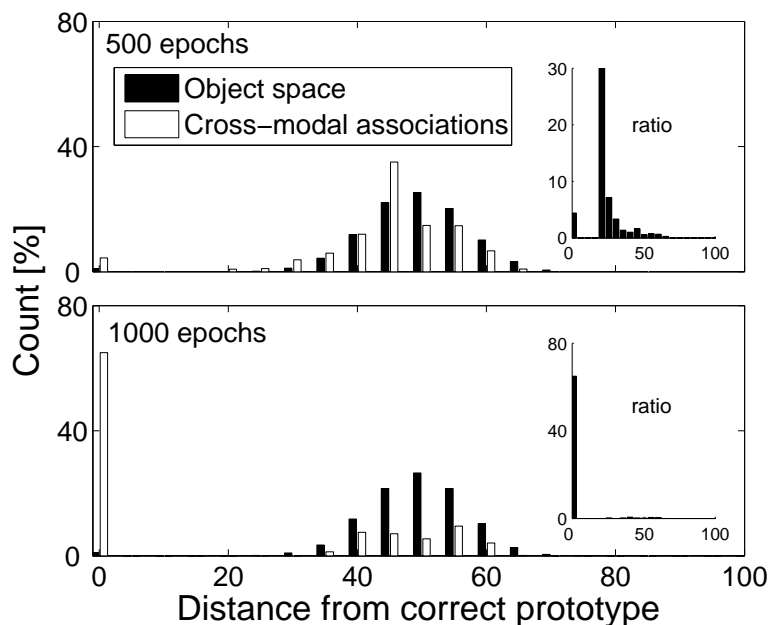
different ages. For example, Gruendel (1977) suggests that over-extensions peak at 17 to 18 months of age (5-7 out of their 50-word lexicon is over-extended) and tend to disappear at 25 months of age. Mandler and McDonough (1998) report that 14-month-olds over-generalise around 50% of their words while Mandler (2004) reports that this has decreased to 30% for 2-year-olds. What is shared among these different studies is both the reduction of similarity-based over-extensions with increasing lexicon size and age and most report that over-extensions tend to disappear from about two years of age. The number of random errors, often called mismatches in the literature are acknowledged but not quantitatively. Clark (1993, p.36) suggests that more often than not, mismatches go unnoticed by the parents. The model faithfully mimics this diminishing proportion of over-extension errors in production with increasing vocabulary size.

*Over-extension errors in comprehension.*

The pattern of over-extension errors in comprehension reported for infants and young children has been subject to considerable debate during the last 30 years. In an early review of the literature, Fremgen and Fay (1980, p. 205) argued that “children never over-extend in comprehension when tested properly”. This view was endorsed in Huttenlocher and Smiley’s (1987) research on over-extension errors. However, comprehension errors in these studies were only evaluated for words and objects which children had also over-extended in production. Later work by Gelman, Croft, Fu, Clausner, and Gottfried (1998) reported that 2-4 year old children made over-extension errors in comprehension based on taxonomic relatedness as well as commonality of shape of objects. Using a preferential looking task, Naigles and Gelman (1995) demonstrated that 21-27 month-olds, presented with pictures of a dog and a cat and asked for a ‘cow’ preferred to look at the dog. The authors of both of these studies argued that infants and young children are prone to make over-extension errors in comprehension based on perceptual or taxonomic similarity. More recently, Arias-Trejo and Plunkett (In press), also using a preferential looking task, demonstrated that infants correctly identified target objects when labelled with their basic-level name provided the distractor object was perceptually dissimilar or taxonomically unrelated: However, when infants were presented with distractor objects from the same taxonomic category which were perceptually similar to the target object (e.g., cat-dog), 18-24 month-olds failed to identify the correct object when it was named. These findings point to a strong influence of perceptual and taxonomic relatedness in over-extension errors in comprehension.

Analogously to production, the distribution of correct and incorrect word-object associations in comprehension is determined by presenting every label to the auditory pathway of the network and identifying the corresponding object category activated on the visual map, following a single labelling event in each category, at different points in training. We measure the Euclidian distance between the prototype of the associated object and the prototype of the correct object for that label. For example, if the label ‘cat’ is presented and the representation of a dog is activated, we measure the distance between a prototypical cat and a prototypical dog.

Fig. 8 depicts the distribution of over-extension errors as a function of distance from the target prototype after 500 epochs and 1000 epochs. After 500 epochs of training, when map structure is incomplete, the distribution of over-extensions errors in comprehension is skewed towards prototypes that are similar to the target prototype (top panel of Fig. 8,

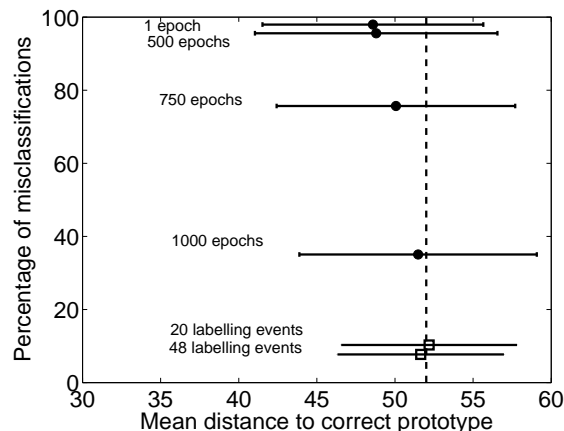


*Figure 8.* Label over-extensions in comprehension during the course of learning. Distributions of inter-category distances (black) and of word-object associations (white). Top panel: After 500 epochs of training, the distribution of over-extensions is skewed when compared to the inter-category histogram. For most misclassifications, the label is over-extended to nearby neighbouring categories. Bottom panel: After extensive learning (1000 epochs of training) most word-object associations are correct. The distribution of over-extension errors is markedly different to the inter-category distribution. In the inset, the ratio of both distributions is plotted.

white bars). In other words, when map structure has not evolved properly, over-extension errors in comprehension are driven by similarity to the target prototype. After 1000 epochs of training, when map structure is complete (see bottom panel of Fig. 8, white bars), the distribution of over-extension errors is random and there is no evidence for errors being driven by similarity.

A more detailed examination of the developmental trajectory of over-extension errors in comprehension is provided in Fig. 9 which plots the average distance of errors from the target prototype at different points in training (1, 500, 750 and 1000 epochs). After 1 epoch of training, the average distance of over-extension errors from the target is well below the average inter-prototype distance (dashed line), indicating that over-extension errors tend to be used for objects that are similar to the correct object ( $p \ll 0.00001$ ). This remains true until 1000 epochs (all  $p$ 's  $\ll 0.0001$ ), when over-extension errors are randomly distributed ( $p=0.69$ ). Once map structure is fully formed, over-extensions in comprehension are no longer determined by similarity.

The disappearance of similarity-based over-extension errors in comprehension is accompanied by a dramatic reduction in the overall number of errors (see Fig. 8). However, a substantial minority of random errors remain after a single labelling event per category when maps are well formed. Naming one object token just once in each category leads to



*Figure 9.* Percentage of the word-object associations that are erroneous, in comprehension, as a function of their mean distance (and standard deviation) to the correct object prototype. From top to bottom, the circles correspond to performance at 1, 500, 750 and 1000 epochs of training after one labelling event per category. The two bottom squares correspond to performance at 1000 epochs of training after 20 and 48 labelling events per category. Up to 750 epochs of training, the mean distance to the correct prototype is smaller than the typical distance between prototypes (dashed line). After extensive learning, at 1000 epochs, residual errors are distributed randomly. The addition of more labelling events eliminates most of the residual errors (squares).

correct word associations for 65% of all acoustic tokens that the network has ever experienced.<sup>9</sup> Although this corresponds to a high level of generalisation performance, further training of the network is needed to eliminate these residual errors. Fig. 9(□) plots the error rate after training the network on 20 and 48 labelling events per category. This additional training reduces the error rate to 10.3% and 7.7%, respectively, highlighting the importance of additional joint attentional activities for fine tuning the accuracy of the model.

This analysis indicates that there are two primary contributing factors in the elimination of over-extension errors in comprehension. First, the emergence of map structure eliminates the incidence of purely similarity-based over-extensions in the model, as a result of the formation of coherent categories, i.e., dissimilar objects from the same category are now treated as similar (see Fig. 4). Second, the residual over-extensions are removed by further direct joint attentional events that target erroneous mappings. Since the network only receives supervision on the occasion of a joint attentional event, a substantial number of such joint attentional events may be required for all over-extensions to be eliminated, particularly because they are randomly distributed across objects. However, realistic interactions between the caregiver and the infant would suggest that whenever a mistake (over-extension or random error) is made, the infant is provided with another labelling event, pairing the correct label with the object. The added information serves to reduce the number of random errors.

The transition from similarity-based errors to random errors parallels the transition

<sup>9</sup>This corresponds to the infant correctly generalising 1460 label-object associations (in our artificial training environment) after just 100 joint-attentional events. Fast mapping experiments with infants typically involve multiple exposures to object-label associations within a single object category.



from prototype-based to exemplar-based processing reported to take place during adult category learning (J. Smith & Minda, 1998) and points to a common underlying cause. J. Smith and Minda (1998) presented adults with nonsense CVCVCVC words generated around two prototypes. They were asked to classify variations around these prototypes as belonging either to the category of the first or the second prototype. Performance was monitored so as to record the time course of categorisation. The participant’s strategy was then compared for each trial to either a prototype model or an exemplar model. They showed that when participants were exposed to large, relatively differentiated categories, they displayed a shift from a prototype-based categorisation strategy to an exemplar-based strategy. Comprehension errors in the current model successfully mimic this transition from an early similarity-based generalisation pattern when exposed to few labelling events, relying on the similarity metrics developed in the maps, to an exemplar-based strategy, capitalising on the increasing number of labelling events and multiple exposure to exemplars. Note that this transition emerges naturally during training and is implemented in the model as a purely bottom-up process without a need for top-down strategic processing.

Another point is worth mentioning; whenever maps are not fully-formed in the model, the number of similarity-based over-extensions is relatively high. In this case, errors are not so much “errors” in assigning the labels to the incorrect object, but rather a correct association to an immature category, or a category that does not fit the model’s category boundaries. Effectively, immature maps determine boundaries based on partial data. Labels are then mapped to categories that overlap only partially with the mature, fully-formed categories. Likewise, Rakison (2000) argues that infant categories do not exactly match adult categories. On this view, over-extension of early labels in infants can also be regarded as the mapping of words to immature categories.

In summary, the model mimics the pattern of similarity-based over-extension errors observed in infant comprehension. Unfortunately, we do not have empirical data regarding the developmental trajectory of over-extensions in comprehension for infants and young children. However, the model predicts that such errors will diminish with increasing vocabulary in a manner similar to that observed for production.

#### *Slips of the tongue and mispronunciation effects.*

Erroneous mappings in production and comprehension can be caused by immature and/or imperfectly segregated maps, auditory and visual. Thus far, we have described errors in production and comprehension from the perspective of confusions on the visual map. We now turn to errors in comprehension and production that arise from confusions on the auditory map, characterised earlier as mispronunciation errors and slips of the tongue, respectively.

First, we examine slips of the tongue whereby the presentation of a dog results in the model producing ‘doll’. We have analysed these production errors in the model, from the perspective of similarity of word-forms between the target word and the produced word. This analysis revealed that for immature maps (less than 1000 epochs of training), erroneous productions are similar to the target word-forms (all  $p \ll 0.0001$ , rejecting the hypothesis that word-forms are randomly chosen). Despite the lack of empirical data for slips of the tongue in early childhood, it is noteworthy that careful analyses of anomalous utterances in adulthood has revealed that “the similarity of the phonological form of words appears to

play a role in word substitutions” (Fromkin, 1971, p.49). Similarly, Fay and Cutler (1977) identified word substitution errors in adults that resemble the intended word in sound but not in meaning and argued that such production errors arise through the mis-selection of a neighbour. Overall, there is a tendency for word-form similarity to impact the number of slips of the tongue: “the likelihood of error increases with increasing similarity between phonemes” (Cutler, 1995, p.118). Although the words used in our model are not presented as phonological features, the present model captures this tendency at the lexical level, and predicts that such slips of the tongue will also occur during infancy and early childhood.

The model also highlights the role of label similarity for errors in comprehension. Recall that after a single joint attentional event per category, the model is successful in generalising the label-object associations to an average of 65% of the other word tokens in each category. Fine-tuning of the remaining associations is provided by further joint attentional activity. Prior to mature levels of performance some word tokens will be incorrectly associated with an object category. For example, the model might select the representation of a cat when hearing ‘bat’. Studies of the impact of word-form similarity on learning label-object associations indicate that infants have difficulty learning two new names that differ by a single phonological feature (Stager & Werker, 1997) or learning a new name that is a phonological neighbour of an existing name (Swingley & Aslin, 2007). Other studies have demonstrated that infants can detect mispronunciations of familiar words provided they have had sufficient exposure (Swingley & Aslin, 2000, 2002; Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Mani & Plunkett, 2007). Furthermore, recent evidence suggests that 19-month-olds display a graded sensitivity to the severity of the mispronunciation of a familiar word (K. White & Morgan, 2008). In the model, comprehension is only ever assessed using pre-exposed word-forms. Thus, sensitivity to mispronunciations in the model is the result of confusions of word-forms rather than systematic manipulation of these pre-exposed forms. Nevertheless, an evaluation of similarity-based confusions in the model is a reasonable first approximation to the graded sensitivity observed in K. White and Morgan (2008). In the model, presenting word tokens that are either close to the prototypical word-form or distant from the typical word-form constitutes a proxy for White and Morgan’s (2008) experiment. We have analysed comprehension errors in the model, from the perspective of similarity of word-forms. This analysis revealed that throughout training, erroneous mispronunciations are similar to the target word-forms (again, all  $p \ll 0.0001$ , rejecting the hypothesis that the distribution of distances between word-forms is random). Insensitivity to mispronunciations in the model is eliminated when map structure is well-formed and additional labelling events fine-tune the residual errors. These findings suggest that the model offers a viable computational framework for explaining the emergent sensitivity to mispronunciations of familiar words and phonological neighbourhood effects in learning new words reported for infants.

In summary, erroneous associations between word-forms and objects can be analysed from a wide range of perspectives. The current implementation of the model highlights the potential common basis of a broad range of phenomena that have rarely been considered as related: Over-extensions in production, slips-of-the-tongue, over-extensions in comprehension and sensitivities to mispronunciations.

*The role of typicality in naming events*

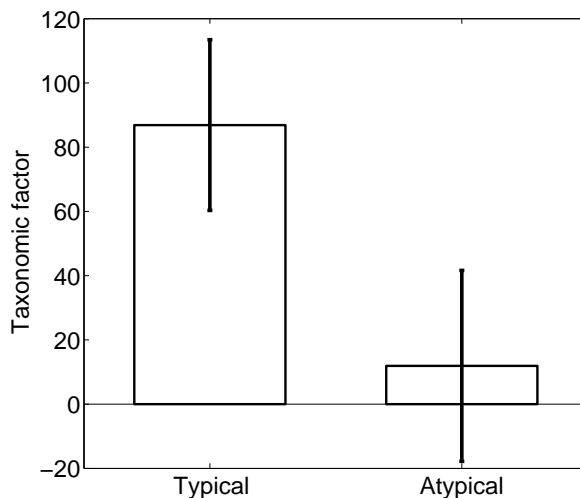
The model also allows for a comparison of generalisation properties when the labelled object in each category is a typical member of that category as opposed to an atypical member of the category. Theoretical accounts have suggested that for categories possessing a family resemblance structure, such as basic level categories (Rosch et al., 1976), some objects are considered good exemplars because of their maximal similarity to members of their own category and minimal similarity to other categories, whereas other objects are considered poor exemplars when they possess a relatively low similarity to members of their own category and, potentially, higher similarity to other categories. A prediction of such an account is that typical members of a category have a privileged status compared to atypical members by offering a superior basis for generalisation. Mervis and Pani (1980) tested this prediction experimentally by exposing children and adults to novel categories and by labelling either good exemplars or poor exemplars. The pattern of generalisation in both children and adults confirmed the prediction of the best-exemplar theory; categories were learned more accurately after exposure to a typical object (or good exemplar) than to an atypical object (or poor exemplar).

We simulated Mervis and Pani's (1980) experiment in the model. We compared the generalisation patterns of the model when the named exemplar in each category was closest to the prototype (in Euclidean distance) to the generalisation behaviour when the model was exposed to labelling events including exemplars from the training set furthest from the prototype. All exemplars were named after maps had been trained for 1000 epochs. Figure 10 displays the taxonomic factor in each condition. In the condition where typical objects are chosen for labelling, generalisation is accurate and 86.9%(SD=26.6) of the objects in the training set are named appropriately. In contrast, when atypical exemplars are named, generalisation is limited to 11.9% of the objects (SD=29.7). A two-tailed t-test confirmed the high significance of the effect ( $p \ll 0.00001$ ).

The model provides a mechanistic implementation of the claim that label-object associations are best learned after a single labelling event when the labelled object is a typical member of that category. It supports the suggestion that generalisation relies heavily on similarity to known object categories and that generalisation is more accurate when the labelled object has a maximal similarity to other members of the category and minimal similarity to other categories.

*The time course of lexical acquisition: vocabulary spurt and impact of labelling events*

The first experiment demonstrated that a pre-established categorisation capacity is a requirement for taxonomic responding, i.e., generalising a label to all members of a category after a single joint attentional labelling event per category. However, during development, the infant is highly likely to be offered more than one joint attentional event per category. We have shown that joint attention has a significant impact on the distribution of synapses between the maps, promoting growth of matching synapses and deterioration of non-matching synapses. Joint attention thereby permits the pruning mechanism to operate in a robust and efficient manner. Joint attentional events also play an important role in the reduction of residual label over-extensions. The frequency of joint attentional events therefore determines the tuning of network sensitivity to object-label associations at a mi-



*Figure 10.* Comparison of generalisation capacity when named objects are typical exemplars or atypical exemplars. Generalisation is more accurate when good exemplars are labelled, mimicking Mervis & Pani’s (1980) results.

croscopic level. Finally, we examine the impact of the frequency of joint attentional events on the macroscopic, quantitative aspects of word learning.

The network is presented with two environments; one with a small number of joint-attentional activities (0.2 labelling events per epoch) and one with a high number of joint-attentional activities (0.5 labelling events per epoch). The lower level of joint attentional activity corresponds to 8.5% of the objects in each category being labelled by the end of training compared to 21% for the higher level of joint attentional activity. These two environments model variation in the amount of joint attentional events observed in parent-infant interactions (Hart & Risley, 1992). The model is trained for 1000 epochs. Synaptic connectivity between the maps increases linearly from 1% at the beginning of training to full connectivity after 500 epochs. The network experiences joint attentional events throughout training at the specified rate. All other parameters are as previously specified, and results are the average of 35 simulations.

It will be recalled from Definition 2 that a word is declared to have been acquired when the network has achieved correct usage in at least 80% of obligatory contexts. The network performs very poorly in both training environments during the earliest stages of learning because map structure is incoherent and generalisation of individual label-object associations is poor (see Fig. 11). However, around 250 epochs, the model begins to show evidence of word learning in the high joint attentional condition. An equivalent level of performance in the low joint attentional condition is not achieved until around 400 epochs. Thereafter, vocabulary size increases dramatically under both training conditions, demonstrating accelerated growth from the 500 epoch mark onwards. However, vocabulary size in the low joint attentional condition is always delayed compared to the high joint attentional condition, even after 1000 epochs of training.

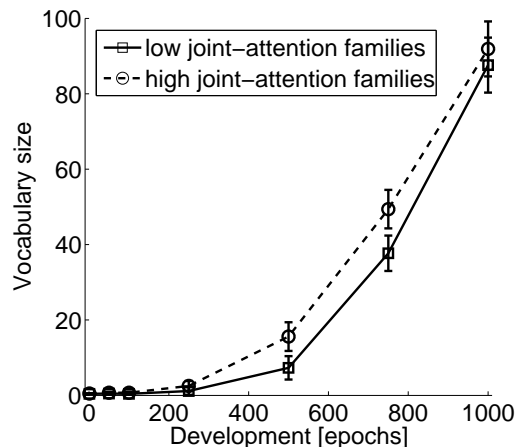


Figure 11. Time course of vocabulary growth when immersed in an environment with a high number of joint attentional activities (dashed line) and with a low number of joint attentional activities (solid line). The model replicates experimental findings that joint attentional activities are a predictor of language development. In both conditions, the rate of lexical acquisition undergoes a rapid increase, similar to a vocabulary spurt.

It is noteworthy that the non-linear rate of increase in vocabulary size at 500 epochs coincides with the non-linear increase in the growth of the taxonomic factor (see Fig. 2). This result is unsurprising since generalisation is central to the definition of word learning (see also Nazzi & Bertoncini, 2003). The overall pattern of vocabulary development is immune to the variability in joint attentional activities: both training regimes exhibit very poor levels of performance during early training and exhibit a spurt-like growth when map formation has begun to emerge. Nevertheless, low joint attentional activity in the model leads to systematically lower levels of vocabulary size, replicating empirical findings that the amount of joint attention correlates with later language skills (Tomasello & Farrar, 1986; Hart & Risley, 1995).

## Discussion

We have proposed a model of early word learning based on core aspects of infant development. As infants explore their environment, their perceptual systems become structured to reflect the regularities they experience. Structure emerges in their visual system so that they are able to group together objects that look alike. Similarly, in the auditory system, the infant learns to identify word-forms. The perceptual systems unfold in service of efficient word learning. Through joint interactions (labelling events) with the caregiver, the infant learns to associate a label with its corresponding object. Based on a single labelling event (or very few of them), the infant excels at generalising this pairing to associate the word category to the object category.

The model learns to categorise words and objects in an unsupervised manner, using self-organising maps. Once topological structure emerges in the maps as a result of separate exposure to individual word and object tokens, supervised joint presentations of words and objects are generalised efficiently following a single labelling event: Members of a

word category generalise to members of the appropriate object category, thereby offering a neuro-computational account of the emergence of taxonomic responding and fast mapping. The model offers a mechanism for generalising associations between auditory and visual categories, based on the association of one exemplar in each category. This process is independent of the specific metric constraining the formation of these categories. Categories do not need to be purely perceptually based. The only assumption required is that categories exhibit a *family resemblance* structure. Family resemblance can derive from perceptual and/or functional relations.

The representation of object categories in the model highlights the privileged status of prototypes. The identification of labels associated with the prototypes outperforms associations with other objects from the same category (see Fig. 4). The effect is not a frequency artefact, since neither the prototypes nor interpolated objects used in testing were used in training. The model developed object categories (and by symmetry of the system, word-form categories) with graded membership. Nevertheless, the generalisation of word-object associations remained accurate, even to atypical members of the category. Sudden shifts in label identification and discrimination occurred at category boundaries demonstrating the categorical nature of the representations created by the model. The model's capacity to respond in a categorical manner at the same time as it exhibits graded membership effects mimics established findings from the adult literature (McMurray, Tanenhaus, & Aslin, 2002) and infant research where categorical effects in the auditory and visual domains abound (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Quinn, Eimas, & Rosenkrantz, 1993) but the privileged status of prototypes over atypical members of a category are also prevalent (Barrett, 1978; Meints, Plunkett, & Harris, 1999). Likewise, generalisation in the model is more accurate when the named object is a typical member of the category as compared to an atypical member of the category in the supervised joint attentional phase of learning, providing an implementation of the best-exemplar theory (Rosch et al., 1976) and of its experimental validation (Mervis & Pani, 1980). These findings confirm the prominent role of similarity when generalising word-object associations to other member of a category possessing a family resemblance structure.

The mapping of labels onto pre-existing categories has often been considered to be one of the primary mechanisms in early word learning (Clark, 1983; Johnston, 1985; Mervis, 1987; K. Nelson, 1985; W. Merriman et al., 1991). The model highlights the virtue of the "mapping hypothesis" (Mervis, 1987) from a neuro-computational perspective. In the absence of pre-existing, non-linguistic categories, word learning is slow and object-specific. However, once object and word-form categories are well formed, fast mapping and taxonomic generalisation result from a single labelling event. Taxonomic responding in the model, defined equivalently as the facility of learning a word from a single labelling event, correlates with the quality of map structure. The model thereby provides a mechanistic account of experimental findings that show a positive correlation between the refinement of infants' perceptual systems and word learning capacity (F. Tsao et al., 2004).

The pattern of over-extension errors in the model derives from the developmental trajectory in the formation of object categories. In its present implementation, the model accounts for a reduction of similarity-based over-extensions in production during infancy (Rescorla, 1980; J. Huttenlocher & Smiley, 1987) as the categories formed on the maps match progressively mature categories (Rakison, 2000). The model also provides a compu-

tational framework to account for the prevalence of similarity-based over-extension errors in comprehension reported in early child language (Naigles & Gelman, 1995; Gelman et al., 1998) and the impact of taxonomic and perceptual similarity on referent identification during infancy (Arias-Trejo & Plunkett, In press). Categories are first constrained by similarity judgements—inaccuracies in the object categories induce similarity-based over-extensions during comprehension—and then residual random errors are eliminated by joint attentional events, by picking difficult members of categories and mapping them to their corresponding labels. The model thereby also provides a mechanism for the transition between prototype-based processing to exemplar-based processing reported in the adult categorisation literature (J. Smith & Minda, 1998) without the need to invoke top-down strategies.

Other types of error made by the model—owing to confusions on the auditory map—can be related to the literature on the impact of word-form similarity in early word learning (Stager & Werker, 1997; Swingley & Aslin, 2007) and sensitivity to mispronunciations in comprehension (Swingley & Aslin, 2000, 2002; Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Mani & Plunkett, 2007; K. White & Morgan, 2008). Again, the model provides a plausible computational framework for interpreting these findings as similarity-based, word-form confusions that are gradually eliminated as the model’s representation of the diverse and variable word-forms are fine-tuned and assigned to their appropriate categories. Finally, the model predicts that infants should make similarity-based slips of the tongue as reported for adults (Fromkin, 1971; Fay & Cutler, 1977; Cutler, 1995). Importantly, all these error types—over-extensions in comprehension and production, sensitivity to mispronunciations and slips of the tongue—are accommodated within a single computational account, suggesting that may share a common underlying basis in infants too.

### *Joint Attention and Cross-Modal Associations*

An increase in the number of labelling events, or an increase in the time spent in joint attentional activities impacts favourably on the size and rate of development of the model’s lexicon. Again, this finding provides a mechanistic interpretation of the positive correlation between joint attention and language acquisition (Tomasello & Farrar, 1986) and of the dependence of vocabulary size on the social environment (high-talking versus low-talking families, Hart & Risley, 1995). W. Merriman et al. (1991) point out that “when youngsters are shown each of several members of an object category the same number of times, but are not told the category name on every trial, they do not show an equal probability of learning the name for each exemplar; rather, this probability is directly related to the frequency and reliability with which each exemplar was named in training... Also, when each of several members of a category is presented equally often, the scope of children’s generalization of the category name broadens as the number of different exemplars labeled in training increases” (pg. 290). This behaviour is precisely mimicked by the model; every acoustic and visual object is presented to the maps the same number of times during the unsupervised learning (or non-linguistic) phase of training. However, the model showed a positive correlation between the number of naming events and the rate of lexical acquisition. In turn, more labelling events improved the quality of word learning.

Cross-modal associations between tokens of objects and labels can be acquired by the model whenever a supervised joint attentional event occurs. At the earliest stages of

training, these associations will necessarily be weak because the activity of the best matching neurones in each map will be low, due to poor weight alignment. Repeated presentation of the same tokens in joint attentional events lead to stronger Hebbian associations between these visual and acoustic stimuli, thereby enabling the model to predict a specific acoustic event given a visual event, or *vice versa*. The model thus possesses the properties of an associative learning device that may underlie the learning characteristics exhibited by young infants. For example, we know that very young infants can learn to associate syllables with simple visual stimuli (Slater, Quinn, Brown, & Hayes, 1999) and, by 6 months, they can orient towards an appropriate face when they hear their mother and father named (Tincoff & Jusczyk, 1999). In the model, such cross-modal associations do not generalise well during early training because the maps have not yet formed coherent categories. Likewise, it is claimed that infants' early auditory-visual associations are purely associative in nature, do not generalise appropriately (Nazzi & Bertoncini, 2003), and lack referential content (Markman, 1992). In the model, the qualitative shift from an associative to a referential use of words is heralded by the emergence of map structure and quantitatively appropriate levels of synapto-genesis. Taxonomic responding ensures that multiple tokens of a word can be used to refer to multiple tokens of an object category. In other words, the model makes the apparent transition from the associative use of proto-words to the referential use of genuine words. However, the model suggests that this transition is at the level of representation and use of words, and not at the level of the mechanisms of acquisition as previously suggested (Nazzi & Bertoncini, 2003; Markman, 1992). Hand-in-hand with the emergence of generalisation patterns in word-object associations comes the ability to fast map (Carey & Bartlett, 1978; Woodward, Markman, & Fitzsimmons, 1994). The model and infant's understanding of a word becomes richer with added joint attentional events, but the initial fast mapping capacity coupled with successful generalisation to other members of the category provide the basis for further refinement of word-meaning associations.

### *Synapto-genesis and Synaptic Pruning*

The model highlights the potential role of synapto-genesis and synaptic pruning in word learning. The blind synapto-genesis observed in early development does not guarantee the existence of appropriate connections between the visual and auditory maps. Random proliferation of connections followed by systematic pruning permits the identification of an optimal set of connections. A reduction of the number of synapses also has the advantage of lowering the metabolic demands of the brain. Maintaining only the strongest synapses can be achieved without affecting successful word-object associations.

The selective pruning process cannot take place before the learner has experienced a wide range of environmental contingencies. If synaptic pruning takes place too early in development, the relative strengths of synapses are not optimally assigned and potentially useful synapses would be wrongly suppressed, inducing sub-optimal word learning capacity. Experimental data over a range of studies reviewed by Huttenlocher et al. (2002) suggest that synaptic density peaks earlier in low level brain structures (e.g., at 3 months for the auditory cortex) than in higher level cortical areas (e.g., at 3.5 years for the medial frontal gyrus, P. Huttenlocher & Dabholkar, 1997). These findings are commensurate with the view that the low level regularities (e.g., orientations and simple shapes in visual scenes, sound frequencies in acoustic scenes) are highly accessible from early in development and



permit a rapid adaptation of dedicated brain structures (e.g., primary visual and auditory cortex). The neural organisation in early sensory cortices can therefore be *frozen* early in life.

In the model, the generalisation of word-object associations was successful only when the visual and auditory maps were well-structured. In the brain, deeper (and higher-level) structures can only start collecting stable input statistics once lower-level structures transmit reliable information following their structuration. Therefore, the pruning mechanism in high-level brain structures has to be initiated later in development in order to have accumulated enough experience of higher-order and cross-modal statistics. It is also noteworthy that the pruning mechanism is initiated slightly earlier in cortical areas dedicated to speech comprehension (Wernicke’s area) than those dedicated to speech production (Broca’s area), “possibly reflecting the earlier onset of receptive vs. expressive language” (P. Huttenlocher, 2002, p.139). The full developmental trajectory of “blind” synapto-genesis and later environmentally-driven synaptic pruning provides a biologically plausible approach to the identification of an appropriate configuration of synapses. We suppose that this combination of synapto-genesis followed by synaptic pruning would be beneficial in perfecting the execution of a wide range of cognitive processes beyond language.

The window of plasticity delimited in time by synapto-genesis and synaptic pruning can be interpreted in terms of critical periods. Before there is a sufficient number of synapses, performance is sub-optimal, since the set of possible connections between and within brain structures is limited. Similarly, after substantial synaptic elimination, cognitive performance would be affected by any modification in the basic statistics of the input to that structure. Whereas critical periods for lower-level brain structures occur early in life, new lexical entries can be added throughout development. As long as the new lexical entry can be encoded in terms of the low-level statistics that were present early in life (e.g., phonemes that are present in the native language), it can be added to the lexicon. However, once the auditory system has optimised its analysis of speech contrasts, learning a word containing speech sounds that are absent from early experience becomes difficult. Thus, although mastery of a second language can be achieved in adulthood, the speaker’s native accent will remain, due to the inaccuracy in the treatment of the fine-tuned representation of non-native sounds. In contrast, learning mappings between an object category and a word type that were initially presented only during unimodal exposures (i.e., in absence of a joint attentional event) can easily be achieved later in learning. Even associations between entirely new objects and new labels can take place before synaptic pruning, provided the new objects and the new labels can be encoded in terms of known low-level parts (e.g., object features and orientations and native speech contrasts) within the general framework of the model (see further discussion in Limitations of the Model). Careful evaluation of the categorisation capacity of SOMs in a non-stationary learning environment has shown that parameters can be varied so as to achieve good topographical organisation in the map and category discrimination while limiting the effect of catastrophic interference (Richardson & Thomas, 2008).

Note that spurious associations between words and objects in the model are naturally removed through the normalisation process implementing limited synaptic resources. Once an object has been named, hierarchies of synaptic strengths connecting the winning nodes on the map are established. If no new association is made to that object category, the con-

nection between that object category and the paired word type remains dominant throughout development. However, if new associations from that object category are formed, the hierarchy of synaptic strengths adapts through cycles of Hebbian reinforcement and normalisation. An erroneous mapping would therefore be supplanted by subsequent correct associations. Note that, similar to infants, arbitrary mappings in the model can be learned and will either fade slowly (the infant will not remember forever that an artificial object used in a laboratory setting is called a ‘dax’) or will be replaced by subsequent associations (after accumulated evidence, the infant will learn that ‘duck’ is used more often to describe the animal than ‘dax’).

The synaptic pruning process in the model is not driven by a minimisation of the error rate of word-object associations. Rather, the driving force behind the synaptic elimination mechanism is the minimisation of metabolic costs induced by synapses. In the model, the taxonomic factor remains constant during the initial phase of synaptic pruning. However, the system finally reaches an optimal number of synapses (1% connectivity) where word learning capacity improves. Knowledge about the word learning environment is thereby re-organised late in learning in the absence of any external error signal. Karmiloff-Smith (1992, p.19) has also claimed that a re-organisation of knowledge in childhood may occur in the absence of an external trigger: “Phase 1 is followed by an internally driven phase during which the child no longer focuses on the external data. Rather, system-internal dynamics take over such that internal representations become the focus of change.” Metabolic minimisation may provide one of the sources of “system-internal dynamics” driving re-organisation in the absence of error.

Note that synaptic pruning between the maps can only improve the mappings when the objects and labels are represented relatively locally, as they are in the topological organisation characteristic of SOMs. In contrast, the elimination of synapses between highly distributed and overlapping representations would be detrimental to the word-object mappings because most connections are required to maintain the relationship between the representations. For example, two auto-encoders—one for auditory processing and one for visual processing—whose representations in the hidden layer might be connected by Hebbian synapses would exhibit a degradation in the mappings between the word and object representations under pruning, a degradation that is not observed in infants. For two SOMs that possess an isomorphic connectivity between localised representations, synaptic pruning can operate in a focused and efficient manner.

### *Labels and Categorisation*

A basic assumption of the model is that category formation takes place in the absence of supervision: “In real life, people notice distinctive classes of objects by themselves without someone telling them that these are different sorts of things, or instructing them in the name.” (Murphy, 2004, p.161). In the model, we contrast this with word learning which we have characterised as a supervised process driven by joint attentional events. However, the processes underlying word learning in the model do not penetrate the categorisation processes themselves. In other words, there is no opportunity in the model for words to impact the process of object categorisation, nor for objects to impact the process of word-form categorisation. This constraint on the model was deliberately imposed to provide a conservative test of taxonomic responding and fast mapping: Our goal in this modelling endeavour was

to demonstrate that a single labelling event could be adequate for successful word learning. Any pre-exposure to word-object pairings during the non-linguistic categorisation phase of training would undermine claims for fast mapping and spontaneous generalisation in the model.

Nevertheless, there is ample evidence that labels can impact the process of categorisation during infancy (Balaban & Waxman, 1997; Plunkett et al., 2008; Waxman & Markow, 1995; Sloutsky, Lo, & Fisher, 2001), a phenomenon that has also been observed with adults (Lupyan et al., 2007; Yamauchi & Markman, 2000). Moreover, W. Merriman et al. (1991) have suggested that “acquisition of the new word meaning is more the joint product of nonverbal concept development and linguistic experience than a simple matter of mapping word onto concept” and “that even when acquisition of some word meaning is contingent on possession of some nonverbal concept, it is still possible that the initial acquisition of the nonverbal concept is influenced by linguistic experiences” (pg. 289). Other connectionist models (Plunkett et al., 1992; Colunga & Smith, 2005; Regier, 2005) have permitted the process of word learning to influence the categorisation process itself. Indeed, object categorisation in Regier’s (2005) model is entirely dependent on word learning. However, the contrast between unsupervised, non-linguistic object categorisation and label-based categorisation processes in infancy (or adulthood) need not be mutually exclusive. Formation of some categories may benefit from the provision of labels (Lupyan et al., 2007; Plunkett et al., 2008) whereas others may emerge in the absence of any linguistic input (K. Nelson, 1973; Younger, 1985).

Likewise, the type of mechanistic account needed to fully explain the processes of category formation and word learning may involve both supervisory and non-supervisory learning systems. For example, Gliozzi et al. (2009) used a slightly modified version of the present architecture, where labels were treated as another object feature in order to drive category formation in the self-organising map. In Gliozzi et al.’s model, labels impact categorisation not as names (word learning) but as features acting in an unsupervised fashion. However, there is no principled reason why the current architecture could not also be adapted for the word learning process itself to influence category formation by permitting joint attentional events to contribute to the delineation of category boundaries. Such a process would be useful both in fine-tuning the extension of a name but could also be useful for creating completely new categories on the basis of shared names. These new name-based categories could themselves become the object of analysis perhaps through the application of SOMs. Clearly, the current framework has wider applicability for the implementation of name-based and feature-based approaches to categorisation (also see Plunkett, in press).

Note that, in its present implementation, the model can learn two labels for the same object. After subsequent pairings of two different labels (e.g., ‘cat’ and ‘animal’) to an object (cat) two synaptic pathways are reinforced. Presentation of either the label ‘cat’ or ‘animal’ would induce a cat response on the object map. Conversely, presentation of the object cat would induce a co-activation of the two labels.<sup>10</sup> It is, however, unlikely that the patterns

---

<sup>10</sup>The capacity of the model to activate two labels for a single object does not contradict the earlier claim that the model can eliminate spurious associations in favour of the correct ones. Spurious associations will occur as a result of very low frequency events. The Hebbian connections thereby formed will be relatively weak and over-written by other learning events. In contrast, multiple correct labels for an object will be consistently reinforced through joint attentional events and so other mechanisms are required to control

of activation associated to the two labels would have the same magnitude, as synaptic reinforcement depends on many factors, such as the neural activation on each map, the order of presentation of the stimuli, frequency, etc. The forced choice implemented in the model would therefore lead to a unique response. However, implementation of contextual biases to the model would enable the system to opt for either of the labels, as both activation levels outperform other spurious associations. A strict mutual exclusivity constraint would, however, predict that infants cannot learn two labels for the same object. A potential implementation of mutual exclusivity in the model would be through the limitation of synaptic resources at the level of each neurone on the maps, instead of having an overall limited resource for the network. This additional constraint on synaptic resources would effectively implement a competition between associations being formed to the same object. It is, however, beyond the scope of the present paper.

#### *Individual variation and atypical development*

When constructing this model, we were careful to ensure that parameters that could affect learning, such as learning rate, frequency of presentation of stimuli, inter- and intra-category variability of auditory and visual stimuli, were constant throughout the training process. This constraint enabled us to ensure that any changes observed in the behaviour of the model were not an artefact of arbitrary changes in the values of these parameters. In the real world, however, these parameters may vary considerably both across and within individual infants. Such variability will certainly have an impact on the performance of the model. We demonstrated one such variable outcome by systematically manipulating the number of labelling events experienced by the model for each epoch of training and showed how this had an impact on vocabulary growth. Other models using similar kinds of neural architectures have been able to demonstrate individual variability in the acquisition of grammatical categories such as nouns and verbs and age-of-acquisition effects (Li et al., 2004), and selectivity of impairment in reading tasks (Mäkkiläinen, 1997). The implications of the current model for individual variability (normal and atypical) in word learning and object-label generalisations can be explored in many ways.

Among possible lines of investigation, manipulation of frequency of individual input presentations, differentiated category sizes, and order of presentation would allow for the identification of individual developmental trajectories as well as the investigation of atypical development. Relaxation of current working hypotheses in the model such as having the same learning parameters in both the visual and auditory map, presenting the network with the same number of objects and words in each category or having individual frequency of word and object presentations would further enrich the set of findings the model can provide.

An example is provided by Williams Syndrome (WS) infants and children who possess relatively large vocabularies when compared to their other, impaired, cognitive skills. Generally, their language acquisition is delayed (N. Harris, Bellugi, Bates, Jones, & Rossen, 1997) and their vocabulary spurt is less marked than for typically-developing infants (Mervis & Bertrand, 1997). Exhaustive categorisation was found to be absent at 4 years of age, even though their vocabulary is already large (Nazzi & Karmiloff-Smith, 2002). When evaluat-

---

their use.

ing language-related generalisation capacities, WS infants “reveal a deficit in generalization of past-tense regularities to novel forms” (Thomas & Karmiloff-Smith, 2003) and do not display taxonomic responding (Stevens & Karmiloff-Smith, 1997). Some researchers have reported that children with WS do not display the production-comprehension asymmetry (Paterson, 2000) and may even produce more words that they comprehend (N. Harris et al., 1997). This pattern of findings led Thomas and Karmiloff-Smith (2003) to conjecture that WS infants behave “as if they were memorizing and using phonological forms without a firm grip of their semantic underpinnings”. Similarly, Nazzi and Bertoncini (2003) concluded that WS infants “capitalise on good speech perception and memorisation capacity ... to acquire a large “proto-lexicon” (with several exemplars of a category being gradually attached to their appropriate sound pattern) by continuing using in childhood the associationist sound pattern-specific object mechanism.”

Our model would readily accommodate these findings and conclusions; preventing the visual map from forming categories so as to mimic WS lack of categorisation skills would lead to delayed vocabulary acquisition, absence of taxonomic responding and absence of a vocabulary spurt. However, through repetitive labelling events, the system would still be able to map the codes of object exemplars to labels in a fashion described in Nazzi and Bertoncini (2003), that is to map several exemplars of a category to their appropriate sound patterns, thereby leading to the formation of a surprisingly large vocabulary.

### Predictions of the Model

The research described in this paper is primarily a labour of synthesis. We have taken a reasonably well-understood neuro-computational architecture (self-organising maps combined with Hebbian learning mechanisms), a plausible input environment (random-dot patterns for objects and frequency-modulated acoustic stimuli for words) and shown how they can be combined to simulate the targeted phenomena, namely taxonomic responding and fast-mapping. It might be argued that a firm grasp of the computational properties of the architecture and an appreciation of the statistical regularities inherent in the input environment we employed renders the outcome of these simulations unsurprising. As modellers, we would certainly acknowledge that our choices of architectural and input assumptions were not arbitrary. These choices constitute the mechanistic explanation of the phenomema targeted by the model. However, we also believe that it is incumbent upon the modeller to identify empirical predictions that flow from the model. It is insufficient to fit the existing data. The generality of a model is often best evaluated against data we do not already possess. We have characterised the precise behavioural characteristics of the model in the Results section and have outlined potential extensions of the model to capture behaviours beyond the core phenomena. We now clarify the behavioural traits of the model that constitute its core empirical predictions.

The core prediction of the model is that the *extension* or generalisation of a label to objects depends on the way that objects are categorised. This prediction is aligned with the claims of the “mapping hypothesis” that labels map onto pre-existing categories. Evidence in support of this hypothesis also corroborates the behaviour of the model (Clark, 1983; Johnston, 1985; Mervis, 1987; K. Nelson, 1985; W. Merriman et al., 1991). Predictions of the model can be derived from analysis of the trajectory of its categorisation behaviour. The most important prediction of the model is that infants and young children will automatically

generalise a label following a single supervised event to all members of the appropriate object category once the object category is well-formed. This is, of course, the taxonomic constraint that the model was built to embody. However, an important corollary of this prediction is that in the absence of well-formed categories, infants and young children (or for that matter older children and adults) will fail to generalise newly learnt labels appropriately. Unfortunately, there is scant evidence to evaluate this claim. We know that infants form individual label-object associations as early as 6-months of age (Tincoff & Jusczyk, 1999)<sup>11</sup> and can learn new ones in the laboratory by 10 months (Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). However, these studies did not test generalisation of labels to new members of the same category. The model predicts that the impact of a single labelling event is not the same, from a generalisation perspective, if it happens early as opposed to late in category formation. A potential test of this prediction would be to teach infants a novel category and label a member of the category under each of two conditions: either naming the first presented object in each category before further experience with category members, or naming after prolonged experience with the objects. The model predicts better label generalisation with object experience first than initial label exposure.

Another core prediction of the model is that errors are based primarily on similarity relations (be they perceptual, functional, etc.) until category formation on the maps is complete. This core prediction provides a natural account of over-extension errors in production and comprehension during infancy. It should be noted that the model was not constructed with this goal in mind, though with hindsight the success of the model in this respect is unsurprising. Perhaps more surprising is the model's capacity to accommodate the findings that word learning is affected by word-form similarity (Stager & Werker, 1997; Swingley & Aslin, 2007) and that mispronunciation effects in a referent identification task are graded (K. White & Morgan, 2008). The model makes the clear and novel prediction that infants, like adults, will make similarity-based slips of the tongue.

The model readily captures typicality effects in naming events: generalisation is more accurate when the labelled object has a maximal similarity to other members of the category and minimal similarity to other categories (Mervis & Pani, 1980). However, the model also predicts that label generalisation will be more accurate with a typical token of the word. Recall that individual tokens of a given word-form also form family resemblance clusters. Operationalisation of this prediction from the model is that the mother's voice should promote better generalisation to appropriate objects than a stranger's voice since the mother's rendition of a word-form is likely to be a more typical member of the word-form category than a stranger's. We know of no studies that test this prediction.

The model predicts that the strength of cross-modal associations is modulated by the frequency and distribution of joint attentional events. In a recent review, C. Nelson (2007) identified the impact of cognitive, linguistic and psycho-social deprivation on the development of children's brains, from a neurobiological perspective. For example, Eluvathingal et al. (2006) examined the white matter connectivity in one group of children raised in an orphanage. White matter connectivity was found to be reduced in pathways of communication between brain areas involved in higher cognitive and emotional functions, with respect to a control group matched for overall development. The authors concluded that

---

<sup>11</sup>The model can readily learn individual tokens of object-label associations in the absence of well-established maps provided they are sufficiently frequent in the input.

connectivity between brain regions is negatively affected by early institutionalisation. Our results suggest that brain connectivity is not only affected negatively when children suffer from cognitive, linguistic or psycho-social deprivation: We predict that children raised in an environment rich in terms of social interactions and linguistic diversity would possess an enhanced density of connections between higher-level brain areas.

### Limitations of the Model

Any modelling endeavour that attempts to replicate aspects of human behaviour is inherently committed to a set of simplifying assumptions. This commitment is a double-edged sword. On the one hand, simplification offers parsimony: the fewer assumptions required to account for the data, the better the explanation offered by the model. On the other hand, simplification typically involves highlighting some aspects of behaviour at the expense of others. Furthermore, simplification can sometimes trivialise explanation. Preferably, simplification offers a plausible first step to a more complex but nevertheless principled account. Our model is not exceptional in making simplifying assumptions such that important aspects of behaviour and processing are not addressed. Here, we describe some of the shortcomings of the model and attempt to outline how the model might be extended to overcome them in a plausible and principled fashion.

Let us first consider the input assumptions. Visual inputs were random dot figures generated from prototypes similar to those used by Posner and colleagues (1962, 1964; 1967; 1968, 1970) to study concept formation in adults. These random dot figures sparsely populate the input space in a manner that fosters the natural clustering capabilities and map formation of SOMs. Although these random dot figures are intended to capture the sparse, natural clustering of many object categories observed in the real world (Murphy, 2004), it is not the case that all categories honour these sparse clustering criteria. Obvious exceptions include tools, furniture, games and other artifactual inventions. As overlap between categories increases and the criteria that partition them become blurred, the coherence and stability of the topological organisation on the SOMs will deteriorate. As we have demonstrated, an important prerequisite for taxonomic responding and fast mapping in the model is robust category formation. Blurring of the distinctions available to the model to discriminate object categories will undermine the appropriate generalisation of labels. Insofar as such categories are readily “fast mapped” by infants and young children, additional machinery will be required by the model to behave in a similar fashion.

We have argued that the abstract coding scheme adopted for visual objects allows us to remain agnostic as to the types of attributes involved in category formation, so that input categories can be interpreted as instantiating family resemblance characteristics defined across, say, either functional or perceptual attributes. Hence, taxonomic responding in the model need not be interpreted as constrained by a single type of categorical or similarity relation. However, the generation of visual inputs to the model did not involve any inter-category similarity constraints other than sparse distribution in the input space and low probability that high level distortions of a given category overlapped with the space occupied by other categories. This results in a category distribution that is best described as a heterarchy, not a hierarchy. Insofar as human category differentiation honours a hierarchical structure, neither the input space in the model nor the resultant organization on the visual map reflects this constraint. This may not be a fatal limitation of the model since it

is questionable that human categories, in infants or adults, are hierarchical (again, see Murphy, 2004, Chapter 10, for an extensive review). However, the nature of the similarity relations between categories does warrant caution in the interpretation of over-extension errors in the model. Over-extension errors in the model are described as similarity-based or random. The overwhelming majority of errors in the model are similarity-based, just as soon as the visual map has achieved a semblance of organization. However, since the model does not implement any kind of between-category hierarchical organization, it is not possible to discriminate between hierarchical, taxonomically-based over-extension errors (e.g., calling a rabbit a ‘cat’) and over-extensions outside the taxonomy (such as ‘plate’ for moon). J. Huttenlocher and Smiley (1987) point out that children’s over-extension errors in production are generally restricted to within the taxonomy rather than out-of-taxonomy errors. Our model does not honour this distinction, though it is more likely to over-extend from visual categories close in the input space. A further implementation of the taxonomic structure of categories in the input space would be needed, however, to evaluate the model’s capacity to mimic Huttenlocher & Smiley’s (1987) findings.

Regarding the acoustic inputs to the model, we have adopted an holistic approach to word-forms by populating the acoustic input space with frequency-coded vectors of the entire word-form filtered to take account of the frequency sensitivities of the human ear. This approach allows us to side-step the difficult problem of pre-determining specific aspects of the speech input that drive word-form categorisation, and allows the model to use whatever statistical regularities that are available in the speech input to carry out the task. However, the model ignores the role of time and the incremental nature of word-form processing, nor does it compute any abstract representations of phonetic or phonemic constituents of the speech signal. Again, insofar as the categorisation of early word-form representations are influenced by such incremental processes and abstract representations, the model will fail to honour the facts of infant word recognition. For example, developmental studies of infant word recognition have shown that infants process words in an incremental fashion (Fernald, Swingley, & Pinto, 2001; Mani & Plunkett, 2007) and that this processing seems to rely on the identification of sub-segmental phonological features (K. White & Morgan, 2008). These considerations will impact the similarity, and hence likely confusions, between word-form categories in a manner that is unlikely to be captured by the model, although it is possible to adapt the model to handle incremental speech input (Kohonen, 1988). We will return to the lack of phonetic/phonemic constituents in the auditory pathway of the model shortly.

Let us now turn to what might best be considered as the computational/architectural limitations imposed by the model. Under joint attentional activity, the model learns to form Hebbian associations between active neurons on the visual map with active neurons on the auditory map. For any joint attentional event, active neurons tend to be restricted to BMUs within category boundaries. Category boundaries are best construed on the visual map as delineating “basic level” categories which emphasise the consistency in object features across exemplars from the same basic-level category and lack of overlap with features of objects from different basic-level categories, whilst category boundaries on the auditory map delineate the corresponding set of “basic level” terms. We have argued that it is possible, in principle, for any basic level category on the visual map to be assigned multiple labels. Thus, the basic level category dog might be assigned the basic level term ‘dog’ *and* the



superordinate and subordinate terms ‘animal’ and ‘chihuahua’, if they were available on the auditory map. The forced-choice procedure implemented in the current version of the model precludes this possibility, though we have suggested that an implementation of contextual biases to the model would enable the model to opt for any of these labels. It should be noted, however, that this is easier said than done! Appropriate use of the label ‘animal’ would need to extend to other basic-level categories at the same time as appropriate use of the label ‘chihuahua’ is restricted to a subset of a single basic-level category on the visual map. Given the propensity for activation to spread in the map to *all* and *only* members of the appropriate category in the current implementation of the model, the additional dual requirements of extension beyond the basic-level and restriction to a subset of the category for superordinate and subordinate labels, respectively, will involve incorporating further computational/architectural features.

It is also worth noting that thematic relations such as ‘dog’-bone are not handled by the current version of the model. This problem is deeper than it may first appear, since it is not good enough to turn off the forced-choice procedure to enable both dog and bone to be associated with ‘dog’ (or *vice versa*): A ‘dog’ is not a bone in the same way that a ‘dog’ is a dog! A suitable neuro-computational account would need to distinguish these relationships.

The inability to learn new words after the visual and auditory maps have stabilised also compromises the plausibility of the model. Novel word learning will be further compromised by synaptic pruning. We have suggested that this problem can be overcome provided new lexical entries are encoded in terms of the low-level statistics that were present early in life. However, the current implementation of the model offers no such independent stable representation for these lower level statistics (edge, frequency, orientation and other spatio-temporal feature detectors) that are encoded early in the visual and auditory cortical pathways. A potential solution to this problem is to employ hierarchies (or heterarchies) of maps in both the visual and auditory pathways of the model, mimicking aspects of the organisation of visual and auditory cortex. Such multiple-level models of visual cortical processing have already been developed for complex aspects of visual object recognition (Miikkulainen et al., 2005; Rolls, Cowey, & Bruce, 1992). These models employ computational/architectural principles entirely compatible with the SOM approach advocated here. However, we know of no models that bring together hierarchies (or heterarchies) of visual and auditory maps in the service of word learning. We suggest that such an approach has the potential to overcome a number of the limitations of the current model.

With respect to word learning, early maps in the visual and auditory pathways can provide stable representations of the low-level statistics required to construct high-level representations of novel label and object categories. It is well-established that synaptic pruning takes place earlier in low-level brain structures than high-level structures, leading to a crystallisation of the representations in low level structures (P. Huttenlocher & Dabholkar, 1997; P. Huttenlocher, 2002). Such crystallisation of lower level structures may help explain why it is so difficult to learn non-native speech contrasts. However, the solution suggested here requires the higher level maps to maintain their plasticity. Of course, we know that such plasticity is maintained in the brain (Pascual-Leone, Amedi, Fregni, & Merabet, 2005). It is another matter to identify the appropriate computational/architectural principles needed to implement this flexibility. Nevertheless, we suspect that the identification of these prin-

ciples will also point to solutions to the multiple labels problem, both taxonomically and thematically. For example, higher level maps will encode increasingly abstract information regarding an object's properties. We know that there are cross-modal associations between visual and auditory cortex from the lowest levels to the very highest levels. Hence, it is possible for labels to be associated with a variety of levels of abstraction of an object, be it encoded as 'chihuahua', 'dog' or 'animal', each instantiation enjoying its own level of taxonomic responding, subordinate, basic or superordinate respectively.

Much work remains to be done to complete this neuro-computational account of word learning and lexical processing. We do not even pretend to have provided an exhaustive listing of the potential shortcomings of the model. For example, what mechanism would constrain a label to associate with its appropriate hierarchical level in the multi-layered neuro-computational model?<sup>12</sup> However, we hope that this outline of a speculative solution represents a plausible first step to a more complex and principled account that does not trivialise the nature of the simplifications implemented to account for taxonomic responding and fast mapping in infants and young children.

### Conclusion

In attempting to construct a neuro-computationally plausible account of taxonomic responding and fast mapping we have exploited a unique blend of constraints and assumptions drawn from the developmental literature. These constraints embody processes that span a broad spectrum of factors ranging from high level social interaction (joint attention) through cognitive processes (categorisation) to low level biological constraints (synapto-genesis and synaptic pruning). The model uses two standard self-organising maps, one of audition and one of vision, and couples them together in a developmental timetable that honours our current state of knowledge concerning the emergence of these biological, cognitive and social processes during infancy. Our modelling endeavour has led us to the conclusion that the critical prerequisites for a neuro-computational system that obeys a taxonomic constraint and is capable of fast mapping include well-structured and well-connected auditory and visual maps. We have argued that these prerequisites are delivered by the biological process of synapto-genesis and the perceptual processes of unsupervised auditory and visual categorisation that occur during infancy. The developmental timetable for the emergence of joint attentional behaviour during infancy is well suited to exploit these prerequisites and thereby deliver fast mapping and taxonomic responding. Moreover, we have presented evidence that the likely neuro-anatomical substrate where binding of the appropriate high level auditory and visual maps responsible for fast mapping and taxonomic responding is to be found in infero-temporal association cortex.

Although our primary aim was to offer an account of fast mapping and taxonomic responding under neuro-computationally plausible conditions, the assumptions required to achieve this goal have led seamlessly to an account of a wide range of other features associated with early lexical development: These include an acceleration in vocabulary growth around the same time that taxonomic responding and fast mapping appear in word learning, an observation that has led many researchers to propose a qualitative shift in word learning strategy during the second year. The model also identifies a new role for joint attentional

---

<sup>12</sup>See Xu and Tenenbaum (2007) for a high level Bayesian account of the source of such constraints.

labelling events: Not only does the frequency of such events predict size and rate of vocabulary acquisition, joint attentional events also assist the synaptic pruning process, targeting the least useful connections through modulating the distribution of synaptic weights between the auditory and visual maps. The unsupervised learning processes on the auditory and visual maps yielded representations of graded category membership which result in typicality effects reported in word learning. The incidence of similarity-based over-extension errors in the model diminished with increasing vocabulary size leaving residual errors to be eliminated by further supervisory labelling events, mimicking a pattern observed in the developmental literature and suggesting a unified account of the transition from prototype based to exemplar based processing reported in the adult category formation literature. Finally, we have argued how the model might be extended to deal with other word learning constraints, such as mutual exclusivity, the impact of labels on the process of categorisation itself and individual differences in word learning both within the normal range and atypical profiles of development, such as Williams syndrome.

All of these phenomena are reproduced in a purely bottom-up processing model, where all learning components are local and self-organise without any knowledge about the behaviour of the other components. In particular, the application of a lexical constraint usually considered to be domain-specific is shown to emerge out of domain-general, associative learning principles confronted with a structured environment. It is generally assumed that the taxonomic constraint guides lexical extensions from the outset of word learning (Markman, 1990). Our account demonstrates that this correlation is no coincidence.

## References

- Abitz, M., Nielsen, R., Jones, E., Laursen, H., Graem, N., & Pakkenberg, B. (2007). Excess of Neurons in the Human Newborn Mediodorsal Thalamus Compared with That of the Adult. *Cerebral Cortex*, 17(11), 2573.
- Adachi, I., Kuwahata, H., & Fujita, K. (2007). Dogs recall their owner's face upon hearing the owner's voice. *Animal Cognition*, 10(1), 17–21.
- Ames, A. (2000). CNS energy metabolism as related to function. *Brain Research Reviews*, 34(1-2), 42–68.
- Amirikian, B., & Friston, K. (2005). A Phenomenological Theory of Spatially Structured Local Synaptic Connectivity. *PLoS Computational Biology*, 1(1), e11.
- Anglin, J. M. (Ed.). (1977). *Word, object, and conceptual development*. New York: W. W. Norton and Company, Inc.
- Arias-Trejo, N., & Plunkett, K. (In press). Perceptual and category membership effects on early word-referent identification. *Journal of Experimental Child Psychology*.
- Attwell, D., & Laughlin, S. (2001). An Energy Budget for Signaling in the Grey Matter of the Brain. *Journal of Cerebral Blood Flow & Metabolism*, 21, 1133–1145.
- Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17, 1267–1284.
- Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of experimental child psychology*, 64(1), 3–26.
- Baldwin, D. (1989). Priorities in Children's Expectations about Object Label Reference: Form over Color. *Child Development*, 60(6), 1291–1306.
- Baldwin, D. (1991). Infants' Contribution to the Achievement of Joint Reference. *Child Development*, 62(5), 875–890.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32, 159–173.

- Barrett, M. (1978). Lexical development and overextension in child language. *Journal of Child Language*, 5, 205-219.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). Cognition and communication from 9-13 months: correlational findings. *The emergence of symbols*, 69-140.
- Blakemore, C., & Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, 228, 477-478.
- Bourgeois, J., & Rakic, P. (1993). Changes of synaptic density in the primary visual cortex of the macaque monkey from fetal to adult stage. *Journal of Neuroscience*, 13(7), 2801.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard.
- Brown, T., Kairiss, E., & Keenan, C. (1990). Hebbian Synapses: Biophysical Mechanisms and Algorithms. *Annual Review of Neuroscience*, 13(1), 475-511.
- Bruner, J., & Scaife, M. (1975). The capacity for joint visual attention in the infant. *Nature*, 253, 265-266.
- Butterworth, G., & Cochran, E. (1980). Towards a Mechanism of Joint Visual Attention in Human Infancy. *International Journal of Behavioral Development*, 3(3), 253.
- Butterworth, G., & Grover, L. (1990). Joint visual attention, manual pointing, and preverbal communication in human infancy. *Attention and performance XIII*, 605-624.
- Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology*, 9(1), 55-72.
- Calder, A. (1996). Categorical Perception of Morphed Facial Expressions. *Visual Cognition*, 3(2), 81-118.
- Camaioni, L., Castelli, M., Longobardi, E., & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language*, 11(33), 345.
- Carey, S., & Bartlett, E. (1978). Acquiring a new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, 63(4).
- Chechik, G. (1998). Synaptic Pruning In Development: A Computational Account. *Neural Computation*, 10(7), 1759-1777.
- Cimpian, A., & Markman, E. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology*, 41(6), 1003-1019.
- Clark, E. (1983). Meanings and concepts. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol. 3. cognitive development*. New York: Wiley.
- Clark, E. (1990). On the Pragmatics of Contrast. *Journal of Child Language*, 17(2), 417-31.
- Clark, E. (1993). *The Lexicon in Acquisition*. Cambridge University Press.
- Colunga, E., & Smith, L. B. (2005). From the Lexicon to Expectations About Kinds: A Role for Associative Learning. *Psychological Review*, 112(2), 347-382.
- Crair, M., Gillespie, D. C., & Stryker, M. P. (1998). The role of visual experience in the development of columns in cat visual cortex. *Science*, 279, 566-570.
- Cutler, E. (1995). *Spoken-word recognition*. San Diego: Academic Press.
- Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357-366.
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception. *Science*, 298, 2013-2015.
- Desimone, R., Albright, T., Gross, C., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4(8), 2051-2062.
- Dunham, P. (1993). Joint-Attentional States and Lexical Acquisition at 18 Months. *Developmental Psychology*, 29(5), 827-31.

- Dunham, P., & Dunham, F. (1990). Effects of Mother-Infant Social Interactions on Infants' Subsequent Contingency Task Performance. *Child Development*, 61(3), 785-793.
- Durbin, R., & Mitchinson, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, 343, 644-647.
- Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Elliott, L., Hammer, M., & Scholl, M. (1989). Fine-grained auditory discrimination in normal children and children with language-learning problems. *Journal of Speech and Hearing Research*, 32, 112-119.
- Eluvathingal, T., Chugani, H., Behen, M., Juhasz, C., Muzik, O., Maqbool, M., et al. (2006). Abnormal brain connectivity in children after early severe socioemotional deprivation: a diffusion tensor imaging study. *Pediatrics*, 117(6), 2093-2100.
- Fahlman, S., Lebiere, C., Finnoff, W., Hergert, F., Zimmermann, H., Prechelt, L., et al. (1990). The Cascade-Correlation Learning Architecture. *Neural Networks*, 6, 771-783.
- Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 505-520.
- Feinberg, I., Thode Jr, H., Chugani, H., & March, J. (1990). Gamma distribution model describes maturational curves for delta wave amplitude, cortical metabolic rate and synaptic density. *Journal of Theoretical Biology*, 142(2), 149-61.
- Fernald, A., Swingle, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, 72(4), 1003-1015.
- Fremgen, A., & Fay, D. (1980). Overextensions in Production and Comprehension: A Methodological Clarification. *Journal of Child Language*, 7(1), 205-11.
- Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language*, 47, 27-52.
- Gelman, S., Croft, W., Fu, P., Clausner, T., & Gottfried, G. (1998). Why is a pomegranate an apple? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of apple and dog. *Journal of Child Language*, 25(02), 267-291.
- Giozzi, V., Mayor, J., Hu, J.-F., & Plunkett, K. (2009). Labels as features (not names) for infant categorisation: A neuro-computational approach. *Cognitive Science*, 33(4), 709-738.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, 21, 125-155.
- Gopnik, A., & Meltzoff, A. (1987). The development of categorization in the second year and its relation to the other cognitive and linguistic developments. *Child Development*, 58, 1523-1531.
- Graham, S., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language*, 26, 295-320.
- Grill-Spector, K., & Malach, R. (2004). The Human Visual Cortex. *Annual Review of Neuroscience*, 27(1), 649-677.
- Gross, C., Rocha-Miranda, C., & Bender, D. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, 35(1), 96-111.
- Gruendel, J. M. (1977). Referential extension in early language development. *Child Development*, 48, 1567-1576.
- Harris, M., Jones, D., Brookes, S., & Grant, J. (1986). Relations between the non-verbal context of maternal speech and rate of language development. *British journal of developmental psychology*, 4(3), 261-268.
- Harris, N., Bellugi, U., Bates, E., Jones, W., & Rossen, M. (1997). Contrasting Profiles of Language Development in Children With Williams and Down Syndromes. *Origins of Language Disorders: A Special Issue of Developmental Neuropsychology*, 13(3), 345-370.
- Hart, B., & Risley, T. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, 28(6), 1096-1105.

- Hart, B., & Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Publishing Company, Inc., Baltimore.
- Hebb, D. (1949). *The organization of behavior: A neuropsychological theory*. John Wiley & Sons.
- Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological Cybernetics*, 82(2), 111–121.
- Hohne, E. A., & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception and Psychophysics*, 56(6), 613–623.
- Holmes, G. (1945). Ferrier Lecture: The Organization of the Visual Cortex in Man. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 132(869), 348–361.
- Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 51, 709–719.
- Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, 95(1), 27–55.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1), 106–154.
- Huttenlocher, J. (1991). Early Vocabulary Growth: Relation to Language Input and Gender. *Developmental Psychology*, 27(2), 236–48.
- Huttenlocher, J., & Smiley, P. (1987). Early Word Meanings: The Case of Object Names. *Cognitive Psychology*, 19(1), 63–89.
- Huttenlocher, P. (1979). Synaptic density in human frontal cortex—developmental changes and effects of aging. *Brain Research*, 163(2), 195–205.
- Huttenlocher, P. (2000). Synaptogenesis in human cerebral cortex and the concept of critical periods. *The role of early experience in infant development*, 15–28.
- Huttenlocher, P. (2002). *Neural Plasticity: The Effects of Environment on the Development of the Cerebral Cortex*. Harvard University Press.
- Huttenlocher, P., & Dabholkar, A. (1997). Regional Differences in Synaptogenesis in Human Cerebral Cortex. *The Journal of Comparative Neurology*, 387, 167–178.
- Janssen, P., Vogels, R., & Orban, G. (2000). Selectivity for 3D Shape That Reveals Distinct Areas Within Macaque Inferior Temporal Cortex. *Science*, 288(5473), 2054–2056.
- Johnson, M. (2001). Functional brain development in humans. *Nature Reviews Neuroscience*, 2(7), 475–483.
- Johnston, J. (1985). Cognitive prerequisites: The evidence from children learning English. *The crosslinguistic study of language acquisition*, 2, 961–1004.
- Johnston, J., & Weismer, S. (1983). Mental rotation abilities in language-disordered children. *Journal of Speech and Hearing Research*, 26, 397–403.
- Jusczyk, P. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3–28.
- Jusczyk, P., & Aslin, R. N. (1995). Infant's detection of sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.
- Kalisman, N., Silberberg, G., & Markram, H. (2003). Deriving physical connectivity from neuronal morphology. *Biological Cybernetics*, 88(3), 210–218.
- Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. MIT Press.
- Kaufmann-Hayoz, R., Kaufmann, F., & Stucki, M. (1986, Apr). Kinetic contours in infants' visual perception. *Child Development*, 57(2), 292–299.
- Kay, D., & Anglin, J. (1982). Overextension and Underextension in the Child. *Journal of Child Language*, 9(1), 83–98.
- Kellman, P., Spelke, E., & Short, K. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, 57, 72–86.
- Kelso, S., Ganong, A., & Brown, T. (1986). Hebbian Synapses in Hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 83(14), 5326–5330.

- Kisilevsky, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effect of experience on fetal voice recognition. *Psychological Science*, 14, 220–224.
- Kohonen, T. (1984). *Self-organization and associative memory*. Berlin: Springer.
- Kohonen, T. (1988). The "neural" phonetic typewriter. *Computer*, 21(3), 11–22.
- Kohonen, T. (1993). Physiological interpretation of the self-organizing map algorithm. *Neural Networks*, 6(7), 895–905.
- Kruschke, J. K. (1992). Alcové: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kuhl, P., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: implications for the 'critical period'. *Language Learning and Development*, 1.
- Landau, B., Smith, L. B., & Jones, S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3, 299–321.
- Landauer, T., & Dumais, S. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104, 211–240.
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345–1362.
- Li, P., Zhao, X., & MacWhinney, B. (2007). Dynamic Self-Organization and Early Lexical Development in Children. *Cognitive Science*, 31(4), 581–61232.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083.
- Macnamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review*, 79, 1–13.
- Malinow, R., & Miller, J. (1986). Postsynaptic hyperpolarization during conditioning reversibly blocks induction of long-term potentiation. *Nature*, 320(6062), 529–530.
- Mandler, J. (2004). Thought before language. *Trends in Cognitive Sciences*, 8(11), 508–513.
- Mandler, J., & McDonough, L. (1998). Studies in Inductive Inference in Infancy. *Cognitive Psychology*, 37(1), 60–96.
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57(2), 252–272.
- Mareschal, D., Johnson, M., Sirios, S., Spratling, M., Thomas, M., & Westermann, G. (2007). *Neuroconstructivism: How the brain constructs cognition*. Oxford University Press New York, NY.
- Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge: MIT Press.
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14, 57–77.
- Markman, E. M. (1992). Constraints on word learning: speculations about their nature, origins and domain specificity. In M. R. Gunnar & M. P. Maratsos (Eds.), *Modularity and constraints in language and cognition: The Minnesota symposium on child psychology* (pp. 59–101). Hillsdale, NJ: Erlbaum.
- Markman, E. M., & Hutchinson, J. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, 16, 1–27.
- Markman, E. M., & Jaswal, V. K. (2004). Acquiring and using a grammatical form class: Lessons from the proper-count distinction. In D. Hall & S. Waxman (Eds.), *Weaving a lexicon* (pp. 371–409). MIT Press.
- Maunsell, J., & Newsome, W. (1987). Visual Processing in Monkey Extrastriate Cortex. *Annual Reviews in Neuroscience*, 10(1), 363–401.
- Maurer, D., & Mondloch, C. J. (2006). The infant as synesthete? In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance* (Vol. XXI, p. 449–472). Oxford University Press.

- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995, Jul). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3), 419–457.
- McClelland, J. L., & Rogers, T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4(4), 310–322.
- McMurray, B., Tanenhaus, M., & Aslin, R. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), 33–42.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (2002). A precursor of language acquisition in young infants. *Psycholinguistics: Critical Concepts in Psychology*.
- Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird: The role of prototypes in early word comprehension. *Developmental Psychology*, 35, 1072–1078.
- Merriman, W., Schuster, J., & Hager, L. (1991). Are names ever mapped onto preexisting categories? *Journal of Experimental Psychology: General*, 120(3), 288–300.
- Merriman, W. E., & Bowman, L. L. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development*, 54(3–4), Serial Number No. 220.
- Mervis, C. (1987). Child-basic object categories and early lexical development. *Concepts and conceptual development: Ecological and intellectual factors in categorization*, 201–233.
- Mervis, C., & Bertrand, J. (1997). Developmental relations between cognition and language: Evidence from Williams syndrome. *Research on communication and language disorders: Contributions to theories of language development*, 75–106.
- Mervis, C., & Pani, J. (1980). Acquisition of basic object categories. *Cognitive Psychology*, 12, 496–522.
- Mervis, C., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, 32(1), 89–115.
- Miikkulainen, R. (1997). Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language*, 59(2), 334–366.
- Miikkulainen, R., Bednar, J., Choe, Y., & Sirosh, J. (2005). *Computational Maps In The Visual Cortex*. Springer.
- Miller, K., Keller, J., & Stryker, M. P. (1989). Ocular dominance and column development: analysis and simulation. *Science*, 245, 605–615.
- Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1997). Language comprehension and cerebral specialization from 13 months to 20 months. *Developmental Cognitive Neuropsychology*, 13, 397–445.
- Morrongiello, B., Fenwick, K., & Chance, G. (1998). Crossmodal learning in newborn infants: Inferences about properties of auditory-visual events. *Infant Behavior and Development*, 21(4), 543–553.
- Müller, R., & Basho, S. (2004). Are nonlinguistic functions in Brocas area prerequisites for language acquisition? fMRI findings from an ontogenetic viewpoint. *Brain and Language*, 89(2), 329–336.
- Murphy, G. (2004). *The Big Book of Concepts*. Bradford Book.
- Murphy, G., & Brownell, H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of experimental psychology. Learning, memory, and cognition*, 11(1), 70–84.
- Naigles, L., & Gelman, S. (1995). Overextensions in comprehension and production revisited: Preferential-looking in a study of dog, cat, and cow. *Journal of child language(Print)*, 22(1), 19–46.
- Nazzi, T., & Bertoncini, J. (2003). Before and after the vocabulary spurt: two modes of word acquisition? *Developmental Science*, 6(2), 136–142.



- Nazzi, T., & Karmiloff-Smith, A. (2002). Early categorization abilities in young children with Williams syndrome. *NeuroReport*, 13(10), 1259.
- Nelson, C. (2007). A Neurobiological Perspective on Early Human Deprivation. *Child Development Perspectives*, 1(1), 13–18.
- Nelson, K. (1973). Some evidence for the cognitive primacy of categorization and its functional basis. *Merrill-Palmer Quarterly*, 19, 21–37.
- Nelson, K. (1985). *Making Sense: The Acquisition of Shared Meaning*. Academic Press.
- Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. (2005). The plastic human brain cortex. *Annual review of neuroscience*, 28, 377.
- Paterson, S. (2000). The development of language and number understanding in Williams syndrome and Down syndrome: Evidence from the infant and mature phenotypes. *Unpublished doctoral thesis, University College London*.
- Pinel, J. (2008). *Biopsychology*. Pearson Education, Limited.
- Plunkett, K. (2006). Learning how to be flexible with words. In Y. Munakata & M. H. Johnson (Eds.), *Processes of change in brain and cognitive development: Attention and performance* (Vol. XXI, pp. 233–248). Oxford University Press.
- Plunkett, K. (in press). How labels impact infant categorisation. In L. Oakes, C. H. Cashon, M. Casasola, & D. H. Rakison (Eds.), *Infant perception and cognition: Recent advances, emerging theories, and future directions*. Oxford University Press.
- Plunkett, K., Hu, J., & Cohen, L. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665–681.
- Plunkett, K., & Nakisa, R. (1997). A Connectionist Model of the Arabic Plural System. *Language and Cognitive Processes*, 12, 807–836.
- Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, 4, 293–312.
- Posner, M. (1962). *An informational approach to thinking*. DTIC Research Report AD0276136.
- Posner, M. (1964). Uncertainty as a predictor of similarity in the study of generalization. *Journal of Experimental Psychology*, 68, 113–8.
- Posner, M., Goldsmith, R., & Welton Jr, K. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, 73(1), 28–38.
- Posner, M., & Keele, S. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.
- Posner, M., & Keele, S. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, 83, 304–308.
- Poulin-Dubois, D., Frank, I., Graham, S., & Elkin, A. (1999). The role of shape similarity in toddlers' lexical extensions. *British Journal of Developmental Psychology*, 17.
- Pruden, S., Hirsh-Pasek, K., Golinkoff, R., & Hennon, E. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, 77(2), 266–280.
- Quartz, S., & Sejnowski, T. J. (1997). The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences*, 20, 537–596.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, Mass.: MIT Press.
- Quinn, P., Eimas, P., & Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22(4), 463–475.
- Rakison, D. (2000). When a Rose Is Just a Rose: The Illusion of Taxonomies in Infant Categorization. *Infancy*, 1(1), 77–90.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29(2), 115–135.
- Rescorla, L. A. (1980). Overextension in early language development. *Journal of Child Language*, 7, 321–335.

- Richardson, F., & Thomas, M. (2008). Critical periods and catastrophic interference effects in the development of self-organizing feature maps. *Developmental Science*, 11(3), 371–389.
- Roe, A., Pallas, S., Hahm, J., & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, 250(4982), 818–820.
- Rogers, T., Lambon Ralph, M., Garrard, P., Bozeat, S., McClelland, J., Hodges, J., et al. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychological Review*, 111(1), 205–35.
- Rogers, T., & McClelland, J. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Bradford Books.
- Roland, P. (1993). *Brain activation*. Wiley-Liss New York.
- Rolls, E. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human neurobiology*, 3(4), 209–222.
- Rolls, E., Cowey, A., & Bruce, V. (1992). Neurophysiological Mechanisms Underlying Face Processing within and beyond the Temporal Cortical Visual Areas [and Discussion]. *Philosophical Transactions: Biological Sciences*, 335(1273), 11–21.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.
- Rosch, E., & Mervis, C. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Rosch, E., Mervis, C., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Psychology*, 7, 573–605.
- Rouw, R., & Scholte, H. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, 10(6), 792–797.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Parallel distributed processing: Explorations in the microstructure of cognition. In (Vol. 2: Psychological and Biological Models, pp. 318–362). The MIT Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Samuelson, L., & Smith, L. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33.
- Schwartz, E., Desimone, R., Albright, T., & Gross, C. (1983). Shape Recognition and Inferior Temporal Neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 80(18), 5776–5778.
- Schwarzlose, R., Swisher, J., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 105(11), 4447.
- Schyns, P. (1991). A modular neural network model of concept acquisition. *Cognitive Science*, 15, 461–508.
- Siesjö, B. (1978). *Brain Energy Metabolism*. John Wiley & Sons.
- Slater, A., Quinn, P., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants learning of arbitrary auditoryvisual pairings. *Developmental Science*, 2(3), 333–338.
- Sloutsky, V., Lo, Y., & Fisher, A. (2001). How Much Does a Shared Name Make Things Similar? Linguistic Labels, Similarity, and the Development of Inductive Inference. *Child Development*, 72(6), 1695–1709.
- Smith, J., & Minda, J. (1998). Prototypes in the Mist: The Early Epochs of Category Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411–1436.
- Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 51–80). Oxford: Oxford University Press.

- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than word learning tasks. *Nature*, 388, 381–382.
- Stevens, T., & Karmiloff-Smith, A. (1997). Word learning in a special population: do individuals with Williams syndrome obey lexical constraints? *Journal of Child Language*, 24(03), 737–765.
- Stryker, M., & Harris, W. (1986). Binocular dominance blockade prevents the formation of ocular dominance columns in cat visual cortex. *Journal of Neuroscience*, 6, 2117–33.
- Suga, N., & O'Neill, W. (1979). Neural axis representing target range in the auditory cortex of the mustache bat. *Science*, 206(4416), 351–353.
- Suter, P., Suter, S., Roessler, J., Parker, K., Armstrong, C., & Powers, J. (1994). Spatial-frequency-tuned channels in early infancy: VEP evidence. *Vision Research*, 34(6), 737–45.
- Swingle, D., & Aslin, R. (2007). Lexical competition in young children's word learning. *Cognitive psychology*, 54(2), 99–132.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
- Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, 110, 647–682.
- Tincoff, R., & Jusczyk, P. (1999). Some Beginnings of Word Comprehension in 6-Month-Olds. *Psychological science*, 10, 172.
- Tomasello, M., & Farrar, M. (1986). Joint Attention and Early Language. *Child Development*, 57(6), 1454–1463.
- Tsao, D., Freiwald, W., Tootell, R., & Livingstone, M. (2006). A Cortical Region Consisting Entirely of Face-Selective Cells. *Science*, 311(5761), 670–674.
- Tsao, F., Liu, H., & Kuhl, P. (2004). Speech perception in infancy predicts language development in the second year of life: a longitudinal study. *Child Development*, 75.
- Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, 1, 139–156.
- Waxman, S., & Kosowski, T. (1990). Nouns Mark Category Relations: Toddlers' and Preschoolers' Word-Learning Biases. *Child Development*, 61(5), 1461–1473.
- Waxman, S., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, 29, 257–302.
- Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Unpublished doctoral dissertation, Harvard University.
- Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7, 49–63.
- Werker, J., & Yeung, H. H. (2005). Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*, 9(11), 519–527.
- Westermann, G., & Miranda, E. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, 89(2), 393–400.
- White, K., & Morgan, J. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*.
- White, L., Coppola, D., & Fitzpatrick, D. (2001). The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature*, 411, 1049–1052.
- Wilson, H. (1988). Development of spatiotemporal mechanisms in infant vision. *Vision Research*, 28(5), 611–628.
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30, 553–566.
- Xu, F., & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272.

- Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26(3), 776–795.
- Younger, B. (1985). The segregation of items into categories by ten-month-old infants. *Child Development*, 56, 1574–1583.
- Zeki, S. (1993). The visual association cortex. *Current Opinion in Neurobiology*, 3(2), 155–9.
- Ziegler, J., Pech-Georgel, C., George, F., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 14110–14115.