



A University of Sussex DPhil thesis

Available online via Sussex Research Online:

<http://sro.sussex.ac.uk/>

This thesis is protected by copyright which belongs to the author.

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Please visit Sussex Research Online for more information and further details

An Investigation of Fast and Slow Mapping

Katherine Elizabeth Twomey

Thesis (by papers) submitted for the degree of
Doctor of Philosophy

University of Sussex

September, 2012

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part, to another University for the award of any other degree.

Signed:

University of Sussex

Katherine Elizabeth Twomey

An Investigation of Fast and Slow Mapping

Summary

Children learn words astonishingly skilfully. Even infants can reliably “fast map” novel category labels to their referents without feedback or supervision (Carey & Bartlett, 1978; Houston-Price, Plunkett, & Harris, 2005). Using both empirical and neural network modelling methods this thesis presents an examination of both the fast and slow mapping phases of children's early word learning in the context of object and action categorisation. A series of empirical experiments investigates the relationship between within-category perceptual variability on two-year-old children's ability to learn labels for novel categories of objects and actions. Results demonstrate that variability profoundly affects both noun and verb learning.

A review paper situates empirical word learning research in the context of recent advances in the application of computational models to developmental research. Data from the noun experiments are then simulated using a Dynamic Neural Field (DNF) model (see Spencer & Schöner, 2009), suggesting that children's early object categories can emerge dynamically from simple label-referent associations strengthened over time. Novel predictions generated by the model are replicated empirically, providing proof-of-concept for the use of DNF models in simulations of word learning, as well emphasising the strong featural basis of early categorisation.

The noun data are further explored using a connectionist architecture (Morse, de Greef, Belpaeme & Cangelosi, 2010) in a robotic system, providing the groundwork for future research in cognitive robotics. The implications of these different approaches to cognitive modelling are discussed, situating the current work firmly in the dynamic systems tradition whilst emphasising the value of interdisciplinary research in motivating novel research paradigms.

Dedication

This thesis is dedicated to my mother, who would have been very proud,
and my father, who taught me that taking things apart to see how they work is the best
fun anyone could have.

Acknowledgements

The work presented in this thesis was funded by ESRC +3 Quota award ES/H014942/ and ESRC Advanced Quantitative Methods Stipend. I warmly thank the caregivers and children who took part in the empirical studies, without whom my research would be impossible

I would also like to thank Dr. Lynne Murphy, Dr. Melanie Green and Dr. Barbara Allen from the Department of English Language and Linguistics at the University of Sussex, whose teaching and encouragement during my undergraduate degree sowed the seed of fascination that grew into this thesis. Finally, thank you to my PhD supervisor, Dr. Jessica Horst, a generous and inspiring mentor who taught me not only how to ask “how”, but also gave me the tools to find out – a gift for which I will always be grateful.

Table of Contents

Introduction	1
Terminology and background	5
Categorisation	5
Fast mapping and word learning	5
Theoretical accounts of categorisation	7
Empirical tests of categorisation and word learning	10
Computational simulations of categorisation and word learning	13
Developmental robotics	16
Overview of current research	17
Paper 1	17
Paper 2	18
Paper 3	19
Paper 4	19
Paper 5	20
Paper 6	21
Discussion	22
References	25
Paper 1: That's More Like It: More Exemplars Facilitate Word Learning	41
The effect of repetition: single versus multiple exemplars	44
The scope of the effect of variability: narrow versus broad categories	46
The effect of presentation order: blocked versus mixed presentation	46

Experiment 1	49
Method	49
Participants	49
Stimuli	50
Procedure and design	51
Results and discussion	53
Experiment 2A	56
Method	56
Participants	56
Stimuli	56
Procedure and design	60
Results and discussion	60
Experiment 2B	63
Method	63
Participants	63
Stimuli, procedure and design	64
Results and discussion	64
General Discussion	67
Comparison facilitates word learning	68
Flexible categorisation	70
References	74
Paper 2: Twinkle Twinkle Little Star: Exemplar Variability Facilitates	
Early Verb Learning	83
The complexity of the verb learning task	86

The unclear effects of variability on verb learning	87
Experiment 1	89
Method	91
Participants	91
Procedure	91
Results	92
Discussion	93
Experiment 2	94
Method	94
Participants	94
Stimuli	95
Procedure and design	98
Results	99
Familiarisation phase	100
Test phase	100
Discussion	100
Experiment 3	101
Method	101
Participants	102
Stimuli	102
Procedure and design	103
Results	104
Habituation phase	104
Looking times during test trials	105

Verb discrimination	107
Effect of vocabulary	108
Discussion	108
Labels drive categorisation following single exemplars	108
Multiple exemplars facilitate verb categorisation	109
General discussion	110
References	115
Paper 3: Two Households, Both Alike in Dignity: Bayesian Versus Emergentist Models of Word Learning and Categorisation	124
The modeller's challenge	127
Bayesian models	128
Theoretical background	128
Mathematical implementation and structure	131
Assumptions	133
Emergentist models	135
Theoretical background	135
Mathematical implementation and structure	138
Assumptions	139
Important differences in modelling word learning	141
Modelling the shape bias	144
A Bayesian shape bias model: Kemp, Perfors & Tenenbaum (2007)	145
An emergentist shape bias model: Colunga & Smith (2005)	149
Evaluating Bayesian and emergentist shape bias models	152
Modelling cognitive representation	152

Modelling the environment	154
Converging support for the emergentist approach	157
General discussion	158
Bridging the gap between Bayesians and emergentists	158
References	161
Paper 4 All Things Considered: Dynamic Field Theory Captures the Effect of Category Variability on Young Children's Word Learning	179
New insights into the word learning puzzle	181
Supporting empirical data	182
Method	182
Participants	182
Stimuli	183
Procedure and design	183
Results	183
Referent selection	183
Test trials	184
Discussion	185
Dynamic Neural Field simulation	185
Architecture	187
Stimuli and procedure	189
Referent selection	190
Test phase	191
Results	191
Discussion	192

General Discussion	195
References	198
Paper 5: Testing a Dynamic Neural Field Model of Children's Category	
Labelling	206
Simulation	209
Dynamic Neural Fields	209
Categorisation by correlated features	210
Method	211
Architecture	211
Stimuli	211
Procedure and design	212
Results and discussion	215
Empirical task	215
Method	216
Participants	216
Stimuli	216
Procedure and design	220
Results and discussion	222
General discussion	225
References	229
Paper 6: An embodied model of young children's categorisation and	
word learning	238
A bidirectional relationship between categories and labels	242
A step forward in understanding cognitive development	245

Empirical basis of the current project	246
Empirical task	248
Procedure	248
Results	249
Flexible categorisation	250
iCub implementation	251
The iCub	253
Neural architecture and inputs	254
Procedure, robot task	259
Results	263
Discussion	267
General discussion	268
References	273
Afterword	288
References	291

Running head: FAST AND SLOW MAPPING: AN INTRODUCTION

An Investigation of Fast and Slow Mapping: An Introduction

Katherine E. Twomey

University of Sussex

An Investigation of Fast and Slow Mapping: An Introduction

“It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection.” *E. Sapir, “The Status of Linguistics as a Science”, 1929.*

“... perception, action and cognition are rooted in the ... dynamic processes of pattern formation: patterns that are reflections of the experiences of acting in and perceiving the world.” *E. Thelen & L. Smith, “A Dynamic Systems Approach to the Development of Cognition and Action”, 1996.*

“My contention is that machines can be constructed which will simulate the behaviour of the human mind very closely. They will make mistakes at times, and at times they may make new and very interesting statements, and on the whole the output of them will be worth attention to the same sort of extent as the output of a human mind [... Then] instead of trying to produce a program to simulate the adult brain, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain.” *A. Turing, “Intelligent Machinery, A Heretical Theory”, 1951.*

Imagine you are an astronaut. You finally arrive at an unexplored planet, after nearly a year of travelling. But when you step out of your spaceship, you're in for a shock: the planet is populated by little green men. The light here is dazzlingly bright, and you find it difficult to pick out shapes and textures. Even worse, because gravity here is different than it is on your spaceship, you can't balance, walk, or interact with your physical environment. And because the little green men communicate using sounds only barely comparable to any sound you have heard before, you can't understand the signals they produce. However, in a matter of twelve months, you have learned to move around, reach for and grasp objects and even understand and reproduce

some of the verbal and gestural elements the little green men use to communicate in their deep-space home. Two years later, you are communicating almost as fluently as your extraterrestrial friends. When you landed here, you knew almost nothing of this strange population and their alien environment; so how did you learn all this so incredibly quickly?

This daunting task is undertaken daily by human infants all over our planet. These little adventurers navigate their way from limited *in utero* experience to adult-like interaction and communication even before formal education gives them a glimpse at the map. The paths children take through their complex perceptual, physical and linguistic environments have provoked philosophical and scientific enquiry for centuries (e.g., Aristotle 335BC/1992; Quine, 1960). This thesis provides a signpost along the road to understanding cognitive development, and cognition in general, by presenting empirical and computational investigations of how children learn to categorise and label the world.

Research in categorisation provides rich insight into the ability to treat different entities as equivalent, from evidence for categories at birth (e.g., of emotion intonation, Mastropieri & Turkewitz, 1999), to category development in human infants (Cohen, 2008) and nonhuman animals (e.g., Cangelosi, 2002), and the fundamental role of categorisation in adult cognition (Medin & Schaffer, 1978). Similarly, an equally large language acquisition literature examines how children infer the referent of an unfamiliar label (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Horst & Samuelson, 2008; Houston-Price, Caloghris, & Raviglione, 2010), how children might learn these labels (Akhtar, Jipson, & Callanan, 2001; Mather & Plunkett, 2009; Munro et al., 2012) and how children use labels refer to categories of events and objects (Brandone, Pence, Golinkoff,

& Hirsh-Pasek, 2007; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Perry, Samuelson, Malloy, & Schiffer, 2010).

Converging evidence suggests that in some contexts labels have a guiding effect on children's – and indeed, adults' – categorisation. For example, Plunkett, Hu & Cohen (2008) familiarised children with two categories of novel objects and demonstrated that when all exemplars were unlabelled, children formed two distinct categories, whereas when all exemplars were accompanied by the same label, children formed a single category. Similarly, Lupyan, Rakison & McClelland (2007) showed that adults' categorisation of novel "aliens" in a classification task was significantly facilitated by labelled stimuli relative to unlabelled stimuli. However, less is known about the effect of categorisation on labelling; that is, how does a category's representational structure affect how children learn to label its members? Further, the means by which children learn categories and words remain controversial despite decades of research. The current work therefore addresses the following overarching questions:

- a. To what extent do categorisation and word learning influence each other?
- b. Are fast mapping, word learning and categorisation governed by domain-specific or domain-general processes?
- c. Can fast mapping, word learning and categorisation be accounted for by simple, associative computational models, and if so, what does this tell us about these behaviours in the real world?

This thesis offers empirical insight into children's word learning and categorisation (Papers 1, 2 and 5), as well as contributing to the growing body of computational investigations of cognitive development (Papers 4, 5, and 6), in the context of the most recent methodological advances in cognitive psychology (Paper 3).

The following sections provide theoretical and methodological background to the papers presented in this thesis, before providing an overview of the current research.

Terminology and Background.

Categorisation

Categorisation is fundamental to adult and child cognition alike (J. D. Smith & Minda, 1998). *Categories* are groups of discriminably different entities (or *exemplars*), which are treated equivalently for the purposes of a given task (Quinn, 1987), such that categorised entities can be quickly and efficiently processed. Indeed, categorisation takes place in multiple domains, from abstract spatial orientation (Quinn, 2004), through phonemes (Rost & McMurray, 2009) and emotion expressions (Quinn, et al., 2011) to the complex domains of objects and events (Jones & Smith, 1998; Oakes, Madole, & Cohen, 1991; Quinn, Eimas, & Rosenkrantz, 1993; Paper 1; Paper 2). Although some argue that categories are structured by *a priori* conceptual knowledge, (Booth, Waxman, & Huang, 2005; Markson, Diesendruck, & Bloom, 2008; Spelke & Kinzler, 2007), categories are nonetheless demonstrably flexible and can be formed online after just minutes of experience (Horst, Oakes, & Madole, 2005; Kovack-Lesh & Oakes, 2007; Oakes, Plumert, Lansink, & Merryman, 1996; Plunkett, Hu & Cohen, 2008).

Fast mapping and word learning

During their second year of life infants begin to learn labels for the categories they regularly encounter (Booth, et al., 2005; Halberda, 2003; Houston-Price, Plunkett, & Harris, 2005). The initial stage in the word learning process is known as *fast mapping* (Carey & Bartlett, 1978). For example, on encountering a novel label alongside a novel object, children associate that label with the object as a whole, rather than part of the object, the texture of the object, the background of the visual scene, and

so on (Quine, 1960). Importantly, fast mapping is not restricted to nouns, but also forms the foundations of (lexical) verb learning (Golinkoff, Jacquet, Hirsh-Pasek & Nandakumar, 1996). Note, however, that verbs are markedly later-acquired than nouns in English-learning children, possibly by virtue of the temporal and perceptual complexity of events and actions relative to objects (Golinkoff & Hirsh-Pasek, 2008; Maguire, Hirsh-Pasek, & Golinkoff, 2006).

Given the infinite number of potential referents for each new label, children's skill in correctly fast mapping novel labels to their referents is impressive. Consequently, various *biases* or *constraints* by which children may limit the number of potential referents have been proposed: for example, pragmatic accounts based on shared knowledge between speaker and listener (Akhtar, et al., 2001; Diesendruck & Markson, 2001); taxonomic constraints, by which children assume that members of the same taxonomic category share a label (Markman & Hutchinson, 1984); the whole object assumption, by which children assume that labels refer to whole objects (Markman, 1990); and the shape bias, by which English-speaking children generalize labels for solid objects to object that share the same shape (Landau, Smith, & Jones, 1988).

Papers 1, 4, 5 and 6 examine fast mapping in the context of referential ambiguity. Specifically, when young children know the label for all-but-one item in an array, they reliably map a novel label to the unlabelled, unknown object (Halberda, 2003; Merriman & Stevenson, 1997; Woodward & Markman, 1991). Various explanations for this ability have been offered. The *novel-name-nameless-category principle* (N3C) states that children know *a priori* that novel names map to novel categories (Golinkoff, et al., 1994). The *mutual exclusivity principle* assumes that children use a process of elimination to reject any object that already has a label (Jaswal, 2010; Markman &

Wachtel, 1988), and is supported by studies demonstrating children's (and adults') attention to known competitors on hearing a novel label (Halberda, 2006; Horst, Scott, & Pollard, 2010). The current research, in particular, suggests that low-level associative mechanisms, rather than explicit reasoning, underlie this expertise (Papers 4, 5 and 6).

Importantly, fast mapping is not *word learning*. Tests of *retention* (recall of label-category mappings after a delay) and *extension* (generalisation of a category label to a new exemplar) suggest that encoding constraints limit children's ability to learn words from a single exposure (Horst & Samuelson, 2008; Munro, et al., 2012). Word learning – defined here as the ability to use new labels after a delay in a new context or in reference to a new exemplar – occurs after the *slow mapping* phase, during which repeated exposures to a label and multiple category exemplars strengthen label-category associations (Capone & McGregor, 2005; Munro, Baker, McGregor, Docking, & Arculi, 2012).

Theoretical accounts of categorisation

On encountering a new object (or animal, or event, or sound, and so on), a child must compare that object to stored memory representations of previously-encountered objects (or animals, etc.), in order to know how – or indeed, whether – to interact with it. For example, if something is small, round, made of plastic and bounces, it is probably safe to play with it, as with other small, round, plastic, bouncing things. In contrast, small, black-and-yellow-striped flying things are unlikely to make good toys. Thus, categorisation involves comparison. Consequently, facilitating comparison also promotes categorisation (Gentner & Namy, 1999; Oakes, Kovack-Lesh, & Horst, 2009; Oakes & Ribar, 2005; but see Quinn & Bhatt, 2010).

However, the structure of category representations is contentious (see Murphy, 2004 for a review). Accounts of categorisation fall broadly into two camps: prototype

and exemplar theories. According to prototype theories, across encounters with different category exemplars the learner abstracts commonalities between these exemplars and forms a single, schematic “prototype” representation. Categories are therefore structured by “family resemblances”. That is, the majority of members of the BIRD “family” have wings, feathers and a beak. However, not all birds are “good” typical exemplars of the category; for example, in category judgement studies a robin is rated as a more typical member of BIRD than a flamingo (Mervis & Rosch, 1981; Rosch, 1975; Rosch & Mervis, 1975). Prototype accounts are supported by empirical studies with adults; for example, Posner & Keele’s (1968) seminal work demonstrates that after training with abstract dot patterns, adults more readily categorise novel prototypic stimuli than novel atypical stimuli (see also Posner, Goldsmith, & Welton Jr, 1967; Strange, Keeney, Kessel, & Jenkins, 1970). Similar prototype effects have been demonstrated in infants from as young as 3 months (Bomba & Siqueland, 1983; Quinn, 1987; Strauss, 1979). Further, recent computational evidence suggests that prototype-based algorithmic models reliably simulate adult categorisation of stimuli drawn from large, well-structured categories (Minda & J. D. Smith, 2001; J. D. Smith & Minda, 1998).

In contrast, according to exemplar-based accounts, a category consists of snapshot-like stored representations of each and every encounter with that category’s exemplars. Rather than comparing new items to a prototype, then, items are compared to each stored representation in turn -- with overall similarity determining categorisation of novel objects (Medin & Schaffer, 1978; Nosofsky, 1984). Exemplar theories are supported by empirical and computational research (Kruschke, 1992; Lamberts, 1994; Nosofsky & Johanson, 2000) and exemplar-based algorithmic models account for

category learning when training exemplars are drawn from small, poorly-structured categories, unlike their prototype-based counterparts (J. D. Smith & Minda, 1998).

At present, a tenuous consensus states that different categorisation strategies become relevant at different stages of learning, dependent on task and processing demands (Horst, Oakes & Madole, 2005; Iverson & Kuhl, 2000; Juslin, Olsson & Olsson, 2003; Minda & J. D. Smith, 2001; Quinn, 2005; J. D. Smith & Minda, 1998). Evidence from Paper 4 (and the pilot data presented in Paper 6) challenges the Exemplars-versus-Prototypes debate (see also Spencer, Blumberg, McMurray, Robinson, Samuelson & Tomblin, 2009): indeed, the overriding historical influence – and success – of the leading exemplar and prototype models may have given rise to a false dichotomy in theoretical accounts of categorization. That is, representational structure in these models is explicitly (mathematically) predefined by the modeller. In prototype models new exemplars are compared to an overall similarity measure (Minda & J. D. Smith, 2001; Murphy, 2004), whereas in exemplar models new exemplars are compared to individual representations (e.g., Generalised Context Model; Nosofsky et al., 2000). Thus, because these models simulate rigid mathematical computations rather than flexible neuronal interactions they imply that categorisation must be either *entirely* exemplar-based or *entirely* prototype-based. In stark contrast, neural network models such as the simulation presented in Papers 4, 5 and 6 demonstrate that a broad, graded category representation (that is, a prototype) can emerge from individual memory traces laid down over several encounters with category exemplars. Thus, Paper 4 discusses an integrative view of categorisation in which exemplar and prototype effects emerge flexibly from the interaction of learning history and task demands (Ellis & Oakes, 2006; Mareschal & Quinn, 2001; Mareschal & Tan, 2007).

Empirical tests of categorisation and word learning

Categorisation has been tested in infants and children across development by measuring behavioural responses to a range of stimuli, based on the assumption that because categorisation results in equivalent responding to different stimuli, reliably *different* responding to a new stimulus indicates discrimination (that is, lack of categorisation). However, not all tasks are relevant to all ages – pointing tasks, for example, are unsuitable for measuring newborns’ categorisation until sufficient motor control is acquired. Likewise, recording increases in sucking rates in response to novel stimuli is not an appropriate measure of preschool children’s categorisation. Thus, experimental paradigms have been developed in which infants’ and children’s categorisation responses can be demonstrated age-appropriately (for a review, see Mareschal & Quinn, 2001). Here I will consider measures of object and verb categorisation, as they are most relevant to this thesis. However, it is important to bear in mind that various innovative paradigms have been developed to test infant and adult categorical perception across modalities (e.g., change detection in colour categorisation, e.g., Franklin, Pilling, & Davies, 2005; conditioned leg-kick in newborn categorisation, Rovee-Collier & Dufault, 1991; event-related potentials in tone perception, e.g., Zheng, Minett, Peng, & Wang, 2012).

Categorisation from birth is readily examined by recording infants’ looking time to a stimulus. In preference procedures, two images are presented manually or via video, and looking time (or head turn) to each is recorded (Fantz, 1958; Fantz & Fagan, 1975; Fantz, 1964; Quinn, 1987). Based on the assumption that very young infants display a familiarity preference (e.g., Maurer & Salapatek, 1976), an increase in fixation to one stimulus over another indicates that the preferred stimulus is an exemplar of a familiar category. However, given sufficient familiarisation, infants will habituate to the

familiar stimulus and begin to prefer novelty (Houston-Price & Nakai, 2004). Thus, increase in fixation to a given stimulus may not indicate a familiarity preference, but in fact a habituation artefact, with implications for the accurate interpretation of results (see also Hunter & Ames, 1988; Yurovsky, Hidaka, Yu, & Smith, 2010).

However, the habituation phenomenon has formed in its own right the basis of habituation and familiarisation studies. Here, children are presented repeatedly with a single stimulus for a fixed number of trials (familiarisation) or until looking time decreases below a predetermined threshold (habituation). A novel stimulus is then presented. Based on older infants' documented novelty bias (Fagan, 1984; Horst, Samuelson, Kucker, & McMurray, 2011; Shinskey & Munakata, 2005) an increase in looking time indicates detection of a novel stimulus; lack of increase indicates categorisation of the novel stimulus with the habituated category. Importantly, as with preference procedures, familiarisation/habituation studies must be carefully designed to avoid confounding in-task habituation effects with categorisation (Oakes, 2010). Despite the care required in interpretation, well-controlled experiments using looking-time procedures are plentiful and enormously informative (e.g., Arias-Trejo, 2010; Bhatt & Quinn, 2011; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Horst, Oakes, & Madole, 2005; Houston-Price, Plunkett, & Harris, 2005; Rakison, Cicchino, & Aslin, 2011).

As children gain experience with objects in their environment and their motor control improves, tests of categorisation involving object manipulation can be employed (from around six months; Oakes, Madole, & Cohen, 1991). In sequential touching studies (e.g., Oakes & Plumert, 2002; Rakison & Cohen, 1999) children are simply presented with an array of toys from two or more categories, and the order and manner in which the child touches the toys is recorded. In the related object examining

paradigm (e.g., Oakes, Coppage, & Dingel, 1997; Oakes et al., 1991; Oakes & Spalding, 1997), children are presented with category exemplars over several familiarization and test trials and the time spent manipulating or looking at the object is recorded. Again, differences in time spent engaging with objects are taken to indicate categorisation (or lack thereof).

Tests of word learning employ related techniques, not least because, as noted, words label categories (and categories and their labels are intimately linked). Looking-time procedures have been adapted to include auditory stimuli in the Intermodal Preferential Looking paradigm (IPL, Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Because looking time (rather than label production or pointing) is the dependent variable in IPL this technique is particularly useful in noun learning studies with infants, and has provided evidence of verbal processing long before the onset of speech (Tincoff & Jusczyk, 1999; Xu, 2002). However, studies using visual paradigms often have high attrition rates - up to 61.9% according to an analysis of studies carried out between 1985 and 2005 (Slaughter & Suddendorf, 2007). Thus, for older children, paradigms in which children can engage with stimuli and the experimenter rather than passively observe may be more suitable (e.g., Papers 1 & 5, this thesis). Nonetheless, IPL and similar intermodal habituation/familiarisation designs are well-suited for the investigation of verb learning, where presentation of moving stimuli via video controls variability of stimuli between subjects (e.g.; Maguire et al., 2010; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008; Paper 2).

Noun learning is also commonly studied using forced-choice trials (e.g., Jaswal & Markman, 2001; Markson & Bloom, 1997; Akhtar et al., 2001). In fast mapping studies, children are familiarised with novel object labels via “referent selection” trials (e.g., Horst & Samuelson, 2008; Mervis & Bertrand, 1994, Wilkinson, Ross & Diamond,

2003). Children are presented with an array of objects (frequently a single novel object and several known objects; e.g., Papers 1 and 5, this thesis; Horst, Scott, & Pollard, 2010) and asked to choose known or novel exemplars as the referents of known or novel nouns (for example “which one’s the *blicket*?”). Children’s retention of these novel nouns is then tested by presenting the just-encountered novel objects in a new context (e.g., alongside other novel objects; Horst, Scott & Pollard, 2010). Importantly, this paradigm can be used not only to examine learning of single label-object mappings (e.g., Axelsson, Churchley & Horst, 2012) but also of label-category mappings, by presenting children with several category exemplars alongside the same novel label (see Papers 1 and 5). In this latter case, category generalisation can also be tested by presenting children with completely novel category exemplars and examining whether children extend newly-learned labels to them, giving insight into the inclusiveness of the category representation learned during the referent selection phase (see also Quinn, Eimas & Rosenkrantz, 1993).

Computational simulations of categorisation and word learning

Recently, computational models have augmented the rich categorisation and word learning literatures. These simulations are mathematical formalisations of the environment pertaining to a particular task, the task itself, and the processes or computations by which the task is completed (see also Simmering, Triesch, Deák & Spencer, 2010). Behaviour in empirical tasks and longer-term developmental transitions have been simulated using networks of mathematically idealised neurons in neural network models (McMurray, Horst, & Samuelson, in press; Papers 4, 5 and 6; Rogers & McClelland, 2004), structured formalisation of statistical inference in probabilistic models (Shafto, Kemp, Mansinghka, & Tenenbaum, 2011; Xu & Tenenbaum, 2007), and mathematical descriptions of representational space in

algorithmic models (Minda & J. D. Smith, 2001; Nosofsky, 1986). Importantly, computational models serve not only to replicate existing data, but also to generate novel, testable predictions, which provide new insights into the cognitive processes that support complex categorisation and word learning phenomena. Neural network and probabilistic models, being highly prevalent in developmental research, are discussed in depth in Paper 3.

Papers 4 and 5 present a neural network model of the effect of category variability on children's noun learning, specifically, a Dynamic Neural Field (DNF) simulation. DNFs are a recently-developed family of neural networks which specifically implement Dynamic Field Theory (DFT; Spencer & Schöner, 2003), in turn a mathematical formalisation of Dynamic Systems Theory (DST; Thelen & Smith, 1994). At its outset, DST represented a radical new way of understanding cognitive and behavioural development, eschewing discrete developmental stages, and instead emphasising the emergence of stable behavioural and cognitive structure from the interaction of the brain, body and the environment across developmental timescales. DNFs have been used to simulate various aspects of development, offering strikingly simple accounts of formerly little-understood phenomena. For example, children's A-not-B error has been shown to emerge from motor-visual associations formed over time, based not only on perceptual salience of the reached-for object, but also on where the child's body is situated in space and time. That is, the incorrect perseverative reaching observed in the A-not-B error is easily explained based on the dynamic formation and decay of memory traces formed during previous reaches (Schutte & Spencer, 2002; Smith, Thelen, Titzer & McLin, 1999). As members of the neural network family, DNFs are based on biologically-plausible, bottom-up processing of low-level input (Spencer, Thomas & McClelland, 2009).

Papers 4 and 5 describe the first DNF model of unsupervised (that is, without feedback) fast mapping under referential ambiguity without supervision representing an extension of the existing DFT literature to a new domain (but see McMurray, Horst, & Samuelson, in press, for a connectionist model of this phenomenon). This simulation also replicates the data presented in Paper 1. Thus, Paper 3 predicts that fast mapping and categorisation may arise from activation dynamics rather than explicit or higher-level reasoning. The model was then used to generate novel predictions about children's categorisation and noun extension based on correlated perceptual features. The empirical replication of the model's predictions described in Paper 4 bear this prediction out, confirming that DNFs can serve as informative models not only of children's word learning, but also of their categorisation in fast mapping tasks (see Perone, Spencer, & Schöner, 2007 for a DNF simulation of categorisation in looking tasks).

The findings from the DNF simulation are further supported by the embodied model presented in Paper 6. Here, a connectionist neural network architecture in an embodied system also exhibits unsupervised fast mapping under referential ambiguity. Inputs to this model are taken directly from auditory and video capture of the same stimuli and experimental context presented to the children in Paper 3, without preprogrammed instructions as to how to reason using, for example, a mutual exclusivity principle¹ In line with Papers 3 and 4, then, this simulation exhibits complex behaviour with a simple associative substrate. The paper presents pilot data for a second replication of the empirical data in Paper 1, and suggests fruitful extension

¹ Although note that *initial* perceptual processing is dealt with by purpose-built camera and speech recognition “modules” and is therefore not biologically plausible).

³ Preliminary analyses revealed a violation of the univariate assumption of sphericity (Mauchly's $W = .685$, $p = .006$) and significant correlations between dependent variables (known

of the system to related fast mapping studies (e.g., Axelsson et al., 2012; Horst et al., 2010) as well as addressing complementary connectionist investigations of word learning (e.g., Gliozzi, Mayor, Hu, & Plunkett, 2009).

Developmental Robotics.

Current interdisciplinary work has seen the integration of developmental psychology and computational modelling with robotics in the emerging field of *developmental* or *epigenetic* robotics (e.g., Berthouze & Metta, 2005). Historically, roboticists aimed to engineer machines to optimally perform a specific task (Asada, et al., 2009). In contrast, developmental robotics focuses on the robot as a model of the developmental transitions observed in infants and children and emphasises the self-organisation of new behaviours as the task environment changes. The term *epigenetic* highlights the field's emphasis on the emergence of behaviour over developmental time (Morse, Belpaeme, Cangelosi, Smith, Ohlsson & Catrambone, 2010). Unlike purely computational models, robotic implementations of computational models allow researchers to investigate *embodiment*; that is, the assumption that the body fundamentally shapes cognition.

Humanoid robotic systems such as the iCub (Metta, et al., 2010) employ “cognitive” architectures which take proprioceptive (spatial location and movement) feedback from their actuators and sensors (e.g., limb position, gaze direction) as inputs to their cognitive architecture alongside the more commonly-included visual or auditory modalities (Morse, de Greeff, Belpaeme, & Cangelosi, 2010). Such robotic systems have successfully simulated embodied phenomena such as children's use of spatial location to bind labels to objects (Morse et al., 2010), use of object function in adjective learning (Yürüten, Uyanik, Çalışkan, Bozcuoğlu, Şahnin & Kalkan, 2012), or the facilitatory effect of congruent body motion on verb recognition times (Farkaš, Malík,

& Rebrová, 2012). Further, inputs to robotic systems are commonly less abstract than inputs to computational models – for example, video input from cameras rather than abstract feature vectors – addressing criticisms of lack of ecological validity sometimes directed at computational simulations (e.g., Diesendruck & Graham, 2010). Paper 5 describes an embodied model of the early stages of the slow mapping process and lays the groundwork for new investigations of embodied language acquisition.

Overview of current research

This thesis takes a strongly interdisciplinary approach to investigating word learning, categorisation and, more broadly, cognitive development. Consisting of six papers, it presents three empirical studies (Papers 1, 2 and 5), a review of computational models of word learning (Paper 3), two computational simulations of categorisation via fast mapping (Papers 4 and 5) and a robotic implementation of the fast mapping task (Paper 6).

Paper 1. That's more like it: Multiple exemplars facilitate word learning

Although research indicates that noun learning facilitates object categorisation (e.g., Ellis & Oakes, 2006; Samuelson & Smith, 1999), evidence for the effect of categorisation on noun learning is contradictory. Similarly, whether the well-known spaced learning effect (e.g., Childers & Tomasello, 2002) applies to children's categorisation over the course of a single experiment is also unclear. This study investigated whether within-category variability affects noun learning by familiarising 30-month-old children with three novel categories via a referent selection task, in either blocked or mixed presentation, and testing their ability to retain (E1 and E2) and extend (E2 only) newly-encountered novel labels after a five-minute delay. In Experiment 1, children fast mapped novel labels either to a single category exemplar repeatedly or to multiple category exemplars in blocked presentation. In Experiment 2a, all children

encountered multiple exemplars in blocked presentation; however half the children encountered low-variability exemplars (*narrow* condition), and half encountered high-variability exemplars (*broad* condition). Experiment 2b explored the effect of presentation order on noun learning by presenting either *narrow* or *broad* categories in mixed presentation. Overall, within-category variability affected children's ability to retain or extend newly-fast-mapped labels for those categories. Specifically, children who encountered low within-category variability retained novel labels (but did not extend them), and children who encountered high within-category variability extended novel labels (but did not retain them).

These results confirm that categorisation does indeed affect noun learning. However, it is not the case that findings concerning noun learning can simply be generalised to other types of word learning. This provided the motivation for Paper 2, which investigated the effect of within-category variability on verb learning.

Paper 2. Twinkle twinkle little star: Exemplar variability facilitates verb learning

Infants' verb learning and action categorisation are little understood relative to infants' early noun learning and object categorisation. Paper 1 demonstrates that within-category variability affects object category labelling, but the effect of such variability on infants' action category labelling is unclear. This paper addresses this issue by habituating 24-month-old children to a single action category paired with a novel label. Importantly, during habituation infants either encountered the same exemplar repeatedly, or multiple exemplars from the same category. Infants learned the action/verb mapping in both conditions, however only infants who encountered multiple exemplars formed a robust, perceptually-based category. Overall, this paper demonstrates that category variability affects verb learning just as it does noun learning.

More broadly, this suggests that noun and verb learning emerge from domain-general perceptually-based processing.

Paper 3. Two households, both alike in dignity: Bayesian versus emergentist models of word learning and categorisation.

Computational models offer unprecedented flexibility in testing and refining existing accounts of cognitive processes. Although such models have offered exciting new insights into cognitive development, at present the field is divided in terms of theoretical accounts of cognition. For example, probabilistic Bayesian models are based on cognitive computation over *a priori* structure, assuming specific organisation for specific cognitive domains. In contrast, emergentist neural network models are based on the simulation of simple neuronal interactions, assuming that a variety of cognitive structures can emerge from domain-general associative processes.

This review provides a historical, theoretical and implementational overview of these two established approaches to cognitive modelling, then compares two models of the *shape bias* (e.g., Landau, Smith & Jones, 1988): the phenomenon by which English-learning children generalise labels for solid objects according to shape. Although on the surface both models replicate the substrate empirical data, on closer examination only the emergentist model succeeds, and further, only the emergentist model generates novel, testable predictions. The paper concludes that both emergentist and Bayesian models offer valuable insights into the processes and computations underlying cognition, but that these insights apply to very different domains. As such, emergentist models currently contribute most to our understanding of word learning, categorisation, and cognitive development more generally.

Paper 4. All things considered: Dynamic Field Theory captures the effect of category variability on young children's word learning

Papers 1 and 2 demonstrate that within-category variability profoundly influences young children's word learning and categorization. This paper describes a Dynamic Neural Field simulation of data presented in Paper 1. The model was presented with a simulated referent selection task and familiarised either with low-variability categories (*narrow* condition) or with high-variability categories (*broad* condition). Like the children in Paper 1, the model was better able to retain category labels in the *narrow* condition, and better able to extend category labels in the broad condition. Simulation data suggest that simple associative mechanisms can underlie complex behaviours, specifically children's ability to fast map novel labels to novel objects in cases of referential ambiguity, as well as their flexible and online categorisation.

However, as with any computational model of cognition, it is important to expose the model to a new task environment in order to generate novel, testable predictions about the behaviour in question. Only when these predictions have been empirically replicated can it be inferred that the computational processes driving the model's behaviour reflect the cognitive processes underlying the human behaviour. Thus, Paper 5 describes such a prediction, and presents its empirical replication.

Paper 5. Testing a Dynamic Neural Field model of children's category labelling

This paper describes an empirical test of predictions generated by the DNF model of categorisation and labelling presented in Paper 4. Given the same experimental design but new stimuli, the model predicted that children will extend novel labels to new exemplars that share many perceptual features with previously-encountered category exemplars (*many features* condition), but not to exemplars that share few features (*few features* condition). To test this prediction, 30-month-old children were familiarised with three novel categories in a referent selection task. After

a five-minute delay, all children retained newly-encountered novel labels. On extension trials, only children in the *many* condition extended novel names. The empirical data therefore confirm the model's prediction.

The empirical replication suggests that DNFs provide an informative model of infant categorisation and word learning. Further, taken together with Papers 1 and 3, this work sheds new light on a behaviour often ascribed to higher-level, conscious reasoning - for example, apparent use of mutual exclusivity in the fast mapping tasks described here (e.g., Markman & Wachtel, 1988). Here, this behaviour is driven by the inhibitory processes intrinsic to the neural network model. Similarly, the model demonstrates the emergence of categories in real-time, from interactions between learning history, recent experience, and in-the-moment perceptual input. Thus, this thesis provides evidence that apparently complex cognitive structure can in fact emerge bottom up from simple, perceptual-associative processes.

Paper 6. An embodied model of young children's categorisation and word learning

Recent innovation in cognitive research has seen the implementation of computational models of cognitive development in embodied robotic systems (Morse, Belpaeme, Cangelosi, & Smith, 2010). This paper presents a second exploration of children's behaviour in Paper 1, this time using a connectionist neural network, implemented in the iCub robot (Metta, et al., 2010). Given referent selection trials in the same manner as in Papers 1 and 5, the robot was able to correctly map novel labels to novel objects. Further, pilot data suggest that the robot was capable of retaining novel labels after encountering a *narrow* category, and generalising novel labels after encountering *broad* category. Thus, this study again demonstrates "mutual exclusivity" emerging from simple associations between labels and objects made in real-time.

Importantly, implementing the neural architecture in an embodied system provides a considerably more ecologically valid experimental setting: inputs to this model include real-time video and audio, as well as the environmental noise experienced by children in a laboratory setting.

These data again point to associative learning and dynamic systems accounts of children's categorisation. This demonstration of unsupervised learning of category labels in an embodied system pushes the boundaries of what has so far been achieved in developmental robotics, and provides the foundation for exciting future interdisciplinary research examining the influence of other cues (e.g., social) to word learning and categorisation, as well as the tantalising possibility exploring the emergence of grammatical structure from the same embodied associative processes. Overall, this paper highlights the enormous benefits of integrating computational and robotic approaches with developmental science for a deeper understanding of cognition.

Discussion

This thesis answers the questions posed earlier as follows:

- a. *To what extent do categorisation and word learning influence each other?*

Categorisation and word learning are mutually influential and tightly coupled. Papers 1 and 2 provide evidence that visual variability helps word learning and support investigation of the effects of variability in other modalities on word learning. Paper 5 in particular suggests future empirical research into the specific perceptual status of labels when applied to objects, whilst Paper 2 points to specific investigation of the interaction between different types of verbs and their action-category referents.

- b. *Are fast mapping, word learning and categorisation governed by domain-specific or domain-general processes?*

Fast mapping, word learning and categorisation clearly *can* emerge from domain-general processes: perceptual variability affects both noun and verb learning (Papers 1 and 2); fast mapping, word learning and categorisation are all underpinned by simple associations in both the DNF (Papers 4 and 5) and embodied (Paper 5) simulations. Whether more abstract systems such as syntax in natural language can also emerge from such low-level process will be the focus of future empirical and computational work.

c. Can word learning and categorisation be accounted for by simple, associative computational models, and if so, what does this tell us about these phenomena in the real world?

Yes. Fast mapping, word learning and categorisation can be accounted for by emergentist computational models (Papers 4, 5 and 6), suggesting that complex real-world behaviour is underpinned by the same dynamic-associative mechanisms that drive neural network models.

Taken together, the empirical and computational studies in this thesis demonstrate that categorisation and word learning emerge from domain-general associative learning mechanisms. There is evidence here that children's experience with labels influences their categorisation: for example, the relationship between vocabulary and verb categorisation in Paper 2 (see also e.g., Lupyan, 2005; Lupyan, Rakison, & McClelland, 2007; Plunkett, Hu, & Cohen, 2008). There is also evidence that children's experience with categories influences their labelling: for example, the facilitatory effect of within category variability on object category label learning in Paper 1 (see also, Perry, Samuelson, Malloy, & Schiffer, 2010; Vlach, Sandhofer, & Kornell, 2008). The present work therefore illuminates the dynamic nature of categorisation and word learning (Colunga & Smith, 2008; Samuelson & Horst, 2007;

Smith, Colunga, & Yoshida, 2010). Similarly, the computational simulations presented capture apparently high-level cognitive phenomena emerging from the temporal, physical, and environmental components of the dynamic cognitive system, extending existing paradigms to new domains and forming a solid basis for future research.

This integration of computational modelling and developmental robotics with empirical work in gives a fine-grained insight into the flexible, emergent processes underlying fast mapping, word learning and categorisation. This thesis therefore treads new methodological ground, pushing the boundaries of our existing conception of development, and pointing to exciting new research paths in the journey to understanding the complex dynamics of cognition.

References

- Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, 72(2), 416-430. DOI: 10.1111/1467-8624.00287
- Arias-Trejo, N. (2010). Young children's extension of novel labels to novel animate items in three testing conditions. *International Journal of Behavioral Development*, 34(3), 206–217. DOI: 10.1177/0165025409350951
- Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, 105(1-2), 63–80. DOI 10.1016/j.jecp.2009.10.002
- Aristotle (1992). *Eudemian Ethics: books I, II, and VIII*. (Woods, M. J., Trans.), Oxford University Press, USA.
- Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., et al. (2009). Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, 1(1), 12-34.
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3. PMID:22291676
- Berthouze, L., & Metta, G. (2005). Epigenetic robotics: modelling cognitive development in robotic systems. *Cognitive Systems Research*, 6, 189-192. DOI: 10.1016/j.cogsys.2004.11.002
- Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, 35(2), 294–328. DOI: 10.1016/0022-0965(83)90085-1

- Booth, A. E., Waxman, S. R., & Huang, Y. T. (2005). Conceptual information permeates word learning in infancy. *Developmental Psychology*, 41(3), 491-505. DOI: 10.1037/0012-1649.41.3.491
- Brandone, A. C., Pence, K. L., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Action speaks louder than words: Young children differentially weight perceptual, social, and linguistic cues to learn verbs. *Child Development*, 78(4), 1322-1342. DOI: 10.1111/j.1467-8624.2007.01068.x
- Cangelosi, A. (2002). Language evolution in apes and autonomous agents. *Behavioral and Brain Sciences*, 25(05), 622–623. DOI: 10.1017/S0140525X0224011X
- Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech Language and Hearing Research*, 48(6), 1468–1480. DOI: 10.1044/1092-4388(2005/102)
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967–978. DOI: 10.1037/0012-1649.38.6.967
- Cohen, L. B. (2010). A bottom-up approach to infant perception and cognition: A summary of evidence and discussion of issues. In Johnson, S. P (Ed.). *Neoconstructivism: The new science of cognitive development* (Vol. 1, pp. 335–367). Oxford, UK: Oxford University Press.
- Colunga, E., & Smith, L. B. (2008). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology*, 26(2), 174–192. DOI:10.1016/j.newideapsych.2007.07.012

- Diesendruck, G., & Graham, S. A. (2010). Kind matters: a reply to Samuelson & Perone. *Cognitive Development*, 25, 149-153. DOI: 10.1016/j.cogdev.2010.02.003
- Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, 37(5), 630-641. DOI: 10.1037/0012-1649.37.5.630
- Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology*, 42(6), 1000–1011. DOI: 10.1037/0012-1649.42.6.1000
- Fagan, J. F., (1984). The relationship of novelty preferences during infancy to later intelligence and later recognition memory. *Intelligence*, 8(4), 339–346. DOI: 10.1016/0160-2896(84)90016-3
- Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record*, 8, 43-47.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science*, 146(668-670).
- Fantz, R. L., & Fagan, J. F. (1975). Visual attention to size and number of pattern details by term and preterm infants during first 6 months. *Child Development*, 46(1), 3–18. DOI: 10.2307/1128828
- Farkaš, I., Malík, T., & Rebrová, K. (2012). Grounding the Meanings in Sensorimotor Behavior using Reinforcement Learning. *Frontiers in Neurorobotics*, 6. DOI:10.3389/fnbot.2012.00001
- Franklin, A., Pilling, M., & Davies, I. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, 91(3), 227–248. DOI:10.1016/j.jecp.2005.03.003

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories.

Cognitive Development, 14(4), 487–513. DOI: 10.1016/S0885-2014(99)00016-7

Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names)

for infant categorization: A neurocomputational approach. *Cognitive Science*,

33(4), 709–738. DOI: 10.1111/j.1551-6709.2009.01026.x

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends*

in Cognitive Sciences, 12(10), 397-403. DOI: 10.1016/j.tics.2008.07.003

Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young-

Children and Adults Use Lexical Principles to Learn New Nouns.

Developmental Psychology, 28(1), 99–108. DOI: 10.1037/0012-1649.28.1.99

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have

it: lexical and syntactic comprehension in a new paradigm. *Journal of Child*

Language, 14(01), 23–45. DOI:10.1017/S030500090001271X

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical

principles may underlie the learning of verbs. *Child Development*, 67(6), 3101–

3119. doi:10.1111/j.1467-8624.1996.tb01905.x

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(1), 125-155. DOI: 10.1017/S0305000900008692

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004).

A theory of causal learning in children: causal maps and Bayes nets.

Psychological Review, 111(1), 3. DOI: 10.1037/0033-295X.111.1.3

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1),

B23-B34. DOI: 10.1016/S0010-0277(02)00186-5

- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), 310-344. DOI: S0010-0285(06)00025-9
10.1016/j.cogpsych.2006.04.003
- Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infants categorize. *Child Development*, 76(3), 614-631. 10.1111/j.1467-8624.2005.00867.x
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI: 10.1080/15250000701795598
- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, 118(2), 234-244.
10.1016/j.cognition.2010.10.015
- Horst, J. S., Scott, E. J., & Pollard, J. P. (2010). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706-713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125-150. DOI: 10.1111/j.1532-7078.2009.00009.x
- Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development*, 13(4), 341-348. DOI: 10.1002/icd.364
- Houston-Price, C., Plunkett, K., & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language*, 32(1), 175-189. DOI: 10.1017/S0305000904006610
- Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research*, 5, 69-95.

- Jaswal, V. K. (2010). Explaining the disambiguation effect: Don't exclude mutual exclusivity. *Journal of Child Language*, 37(1), 95–113. DOI: 10.1017/S0305000909009519
- Jaswal, V. K., & Markman, E. M. (2001). Learning proper and common names in inferential versus ostensive contexts. *Child Development Vol 72(3) May-Jun 2001*, 768-786. DOI: 1467-8624.00314
- Jones, S. S., & Smith, L. B. (1998). How children name objects with shoes. *Cognitive Development*, 13(3), 323-334. DOI: 10.1016/S0885-2014(98)90014-4
- Juslin, P., Olsson, H., & Olsson, A. C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, 132(1), 133–156. DOI:10.1037/0096-3445.132.1.133
- Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, 98(2), 69-93. DOI 10.1016/j.jecp.2007.05.001
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological review*, 99(1), 22. DOI: 10.1037/0033-295X.99.1.22
- Kucker, S. C., & Samuelson, L. K. (2011). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*. DOI: 10.1111/j.1532-7078.2011.00081.x
- Lamberts, K. (1994). Flexible tuning of similarity in exemplar-based categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1003. DOI: 10.1037/0278-7393.20.5.1003
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. DOI: 10.1016/0885-2014(88)90014-7

- Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. *Progress in Neural Processing*, 16, 87. DOI: 10.1142/9789812701886_0008
- 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. DOI: 10.1111/j.1467-9280.2007.02028.x
- Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A Unified Theory of Word Learning: Putting Verb Acquisition in Context *Action meets word: How children learn verbs*. (pp. 364-391).
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, 11(4), 628-634. DOI: 10.1111/j.1467-7687.2008.00707.x
- Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, 5(10), 443-450. DOI: 10.1016/S1364-6613(00)01752-6
- Mareschal, D., & Tan, S. H. (2007). Flexible and context-dependent categorization by eighteen-month-olds. *Child Development*, 78(1), 19-37. DOI: 10.1111/j.1467-8624.2007.00983.x
- Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1), 57-77. DOI: 10.1207/s15516709cog1401_4
- Markman, E. M., & Hutchinson, J. E. (1984). Childrens Sensitivity to Constraints on Word Meaning - Taxonomic Versus Thematic Relations. *Cognitive Psychology*, 16(1), 1-27. DOI: 10.1016/0010-0285(84)90002-1

- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, 20(2), 121-157. DOI: 10.1016/0010-0285(88)90017-5
- Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Developmental Science*, 11(2), 204-208. DOI: 10.1111/j.1467-7687.2007.00666.x
- Mastropieri, D., & Turkewitz, G. (1999). Prenatal experience and neonatal responsiveness to vocal expressions of emotion. *Developmental Psychobiology*, 35(3), 204-214. DOI: 10.1002/(SICI)1098-2302(199911)35:3<204::AID-DEV5>3.0.CO;2-V
- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, 14(1), 60-76. DOI: 10.1080/15250000802569702
- McMurray, B., Horst, J. S., & Samuelson, L. K. (in press). Word learning as the interaction of online referent selection and slow associative learning. *Psychological Review*.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review; Psychological Review*, 85(3), 207. DOI: 10.1037/0033-295X.85.3.207
- Merriman, W. E., & Stevenson, C. M. (1997). Restricting a familiar name in response to learning a new one: Evidence for the mutual exclusivity bias in young two-year-olds. *Child Development*, 68(2), 211-228. DOI: 10.2307/1131846
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual review of psychology*, 32(1), 89-115. DOI: 10.1146/annurev.ps.32.020181.000513

- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, 23(8-9), 1125-1134. DOI: 10.1016/j.neunet.2010.08.010
- Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: the effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(3), 775. DOI: 10.1037/0278-7393.27.3.775
- Morse, A. F., Belpaeme, T., Cangelosi, A., & Smith, L. B., Ohlsson, S. & Catrambone, R. (2010). *Thinking With Your Body: Modelling Spatial Biases in Categorization Using a Real Humanoid Robot*. Paper presented at the 32nd Annual Conference of the Cognitive Science Society, Austin, TX.
- Morse, A. F., de Greeff, J., Belpeame, T., & Cangelosi, A. (2010). Epigenetic Robotics Architecture (ERA). *Autonomous Mental Development, IEEE Transactions on Autonomous Mental Development*, 2(4), 325-339. DOI: 10.1109/TAMD.2010.2087020
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3. PMid:22291676
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104. DOI: 10.1037/0278-7393.10.1.104
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*; *Journal of Experimental Psychology: General*, 115(1), 39.

- Nosofsky, R. M & Johansen, M. K. (2000). Exemplar-based accounts of “multiple-system” phenomena in perceptual categorization. *Psychonomic Bulletin and Review*, 7(3), 375–402. DOI: 10.1037/0096-3445.115.1.39
- Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, 9(4), 247–255. DOI: 10.1111/1467-9280.00051
- Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3), 255–268. DOI:10.1080/15248371003699977.
- Oakes, L. M., Coppage, D. J., & Dingel, A. (1997). By land or by sea: The role of perceptual similarity in infants’ categorization of animals. *Developmental Psychology*, 33(3), 396–407. DOI: 10.1037/0012-1649.33.3.396
- Oakes, L. M., Kovack-Lesh, K. A., & Horst, J. S. (2009). Two are better than one: Comparison influences infants’ visual recognition memory. *Journal of Experimental Child Psychology*, 104(1), 124–131. DOI: 10.1016/j.jecp.2008.09.001
- Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants Object Examining - Habituation and Categorization. *Cognitive Development*, 6(4), 377-392. DOI: 10.1016/0885-2014(91)90045-F
- Oakes, L. M., & Plumert, J. M. (2002). Variability in thirteen-month-old infants’ touching patterns in the sequential-touching task. *Infant Behavior & Development*, 25(4), 529–549. DOI: 10.1016/S0163-6383(02)00149-2

- Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior & Development, 19*(4), 425-440. DOI: 10.1016/S0163-6383(96)90004-1
- Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7*(1), 85-98. DOI: 10.1207/s15327078in0701_7
- Oakes, L. M., & Spalding, T. L. (1997). The role of exemplar distribution in infants' differentiation of categories. *Infant Behavior & Development, 20*(4), 457-475. DOI: 10.1016/S0163-6383(97)90036-9
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning. *Psychological Science, 21*(12), 1894-1902. DOI: 10.1177/0956797610389189
- Pfeifer, R., & Bongard, J. (2007). *How the Body Shapes the Way We Think*. Cambridge, Mass.: MIT Press.
- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition, 106*(2), 665-681. DOI: 10.1016/j.cognition.2007.04.003
- Posner, M. I., Goldsmith, R., & Welton Jr, K. E. (1967). Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology, 73*(1), 28.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology, 77*(3p1), 353.
- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge: MIT Press.

- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, 27(2), 145-179. DOI: 10.1016/0010-0277(87)90017-5
- Quinn, P. C. (2004). Visual perception of orientation is categorical near vertical and continuous near horizontal. *Perception*, 33(8), 897-906. DOI: 10.1068/p3453
- Quinn, P. C. (2005). Young Infants' Categorization of Human Versus Nonhuman Animals: Roles for Knowledge Access and Perceptual Process. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building Object Categories in Developmental Time* (pp. 107–130). Mahwah, NJ: Lawrence Erlbaum Associates.
- Quinn, P. C., Anzures, G., Izard, C. E., Lee, K., Pascalis, O., Slater, A. M., et al. (2011). Looking across domains to understand infant representation of emotion. *Emotion Review*, 3(2), 197-206.
- Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception*, 39(6), 795-806. DOI:10.1068/P6639
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for Representations of Perceptually Similar Natural Categories by 3-Month-Old and 4-Month-Old Infants. *Perception*, 22(4), 463-475. DOI: 10.1068/p220463
- Rakison, D. H., Cicchino, J. B., & Aslin, R. N. (2011). Correspondences between what infants see and know about causal and self-propelled motion. *Cognition*, 118(2), 171–192. DOI:10.1016/j.cognition.2010.11.005
- Rakison, D. H., & Cohen, L. B. (1999). Infants' use of functional parts in basic-like categorization. *Developmental Science*, 2(4), 423–431. DOI: 10.1111/1467-7687.00086
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

- Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology - General*, 104(3), 192–233. DOI: 10.1037/0096-3445.104.3.192
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4), 573–605. DOI: 10.1016/0010-0285(75)90024-9
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339–349. DOI: 10.1111/j.1467-7687.2008.00786.x
- Rovee-Collier, C., & Dufault, D. (1991). Multiple contexts and memory retrieval at three months. *Developmental Psychobiology*, 24(1), 39–49. DOI: 10.1002/dev.420240104
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. DOI: 10.1080/15250000701795598
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1–33. DOI: 10.1016/S0010-0277(99)00034-7
- Schutte, A. R., & Spencer, J. P. (2002). Generalizing the dynamic field theory of the A-not-B error beyond infancy: Three-year-olds' delay- and experience-dependent location memory biases. *Child Development*, 73(2), 377–404. DOI: 10.1111/1467-8624.00413
- Shafto, P., Kemp, C., Mansinghka, V., & Tenenbaum, J. B. (2011). A probabilistic model of cross-categorization. *Cognition*. DOI: 10.1016/j.cognition.2011.02.010

- Shinskey, J. L., & Munakata, Y. (2005). Familiarity breeds searching: infants reverse their novelty preferences when reaching for hidden objects. *Psychological science*, 16(8), 596–600. DOI:10.1111/j.1467-9280.2005.01581.x
- Simmering, V. R., Triesch, J., Deak, G. O., & Spencer, J. P. (2010). A Dialogue on the role of computational modeling in developmental science. *Child Development Perspectives*, 4(2), 152–158. DOI: 10.1111/j.1750-8606.2010.00134.x
- Slaughter, V., & Suddendorf, T. (2007). Participant loss due to “fussiness” in infant visual paradigms: A review of the last 20 years. *Infant Behavior & Development*, 30(3), 505–514. DOI 10.1016/j.infbeh.2006.12.006
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411. DOI: 10.1037/0278-7393.24.6.1411
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, 34(7), 1287–1314. DOI: 10.1111/j.1551-6709.2010.01130.x
- Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review*, 106(2), 235–260. DOI: 10.1037/0033-295X.106.2.235
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558-1568. DOI: 10.1016/j.cognition.2007.06.010
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89-96. DOI: 10.1111/j.1467-7687.2007.00569.x
- Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer

abide the nativist-empiricist debate. *Child Development Perspectives*, 3(2), 79–

87. DOI: 10.1111/j.1750-8606.2009.00081.x

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, 6(4), 392–412. DOI: 10.1111/1467-7687.00295

Spencer, J. P., Thomas, M., & McClelland, J. (2009). *Toward A Unified Theory of Development: Connectionism and Dynamic Systems Theory Reconsidered*. Oxford Series in Developmental Cognitive Neuroscience. New York, NY: Oxford University Press. DOI: 10.1093/acprof:oso/9780195300598.001.0001

Strange, W., Keeney, T., Kessel, F. S., & Jenkins, J. J. (1970). Abstraction over time of prototypes from distortions of random dot patterns: A replication. *Journal of Experimental Psychology*, 83(3, Pt.1), 508-510. DOI: 10.1037/h0028846

Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 618. DOI: 10.1037/0278-7393.5.6.618

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass.: MIT Press.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, 10(2), 172–175. DOI: 10.1111/1467-9280.00127

Woodward, A. L., & Markman, E. M. (1991). The mutual exclusivity bias in young children's word learning. *Developmental Review*, 11(2), 137-163. DOI: 10.1016/0273-2297(91)90005-9

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, 85(3), 223-250. DOI: 10.1016/S0010-0277(02)00109-9

- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245-272. DOI: 10.1037/0033-295X.114.2.245
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414. DOI: 10.1111/j.1467-9280.2007.01915.x
- Yurovsky, D., Hidaka, S., Yu, C., & Smith, L. B. (2010). Linking learning to looking: Habituation and association in infant statistical language learning. *Proc. of the Annual Meeting of the Cognitive Science Society* (Vol. 32, pp. 1589–1594).
- Yürüten, O., Uyanık, K., Çalışkan, Y., Bozcuoğlu, A., Şahin, E., & Kalkan, S. (2012). Learning Adjectives and Nouns from Affordances on the iCub Humanoid Robot. In T. Ziemke, C. Balkenius, & J. Hallam (Eds.), *From Animals to Animats 12*, Lecture Notes in Computer Science (Vol. 7426, pp. 330–340). DOI: 10.1007/978-3-642-33093-3_33
- Zheng, H. Y., Minett, J. W., Peng, G., & Wang, W. S-Y. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes*, 27(2), 184–209. DOI: 10.1080/01690965.2010.520493

Running Head: MULTIPLE EXEMPLARS FACILITATE WORD LEARNING

That's More Like It: Multiple Exemplars Facilitate Word Learning

Katherine E. Twomey & Jessica S. Horst

University of Sussex

Author note

A version of this paper is currently under review:

Twomey, Ranson & Horst (under review). That's more like it: Multiple exemplars facilitate word learning. *Journal of Cognition and Development*.

Abstract

Previous research indicates that learning words facilitates categorisation. In the current study, we investigated whether learning about a category facilitates word learning by presenting 2-year-old children with multiple referent selection trials from the same object category. In Experiment 1, children mapped novel names either to a single category exemplar repeatedly or to multiple category exemplars across trials. All children did very well on the initial task. However, only children who encountered multiple exemplars retained labels after a short delay. Experiments 2A and 2B extended this finding by exploring the effect of within-category variability and presentation order both on retention and on extension of novel names. Children encountered exemplars from either narrow or broad categories across trials, in either blocked (E2a) or mixed (E2b) presentation. Across Experiment 2, all children did very well on the initial task. In E2a, only children who encountered narrow exemplars retained labels after a short delay. Further, only children who encountered broad exemplars extended labels to a never-before-seen exemplar. In E2b, children neither retained nor extended labels. Overall, these data suggest that comparison and within-category variability can profoundly affect children's word learning. The flexible interplay between word learning and categorisation discussed in terms of the dynamic systems approach to development and cognition more broadly.

That's More Like It: Multiple Exemplars Facilitate Word Learning

Categorisation is a fundamental cognitive skill. The ability to group items in the world and respond to them equivalently allows children to interact efficiently with their complex environment (Quinn, Slater, Brown & Hayes, 2001). In order to communicate with those who share that environment, children must also learn labels for their categories; their early vocabularies are consequently dominated by words for the categories of objects they encounter day-to-day (Samuelson & Smith, 1999; Waxman, 2003). Indeed, existing research demonstrates a close relationship between categorisation and vocabulary acquisition (e.g., Gopnik & Meltzoff, 1992; Nazzi & Gopnik, 2001; Thom & Sandhofer, 2009). For example, infants with large vocabularies categorise more flexibly than infants with small vocabularies; that is, they are able to categorise objects on more than one dimension (Ellis & Oakes, 2006; Horst et al., 2009). Similarly, toddlers with larger productive vocabularies are better able to categorise objects at the basic level (Singer-Freeman & Bauer, 1997) and more likely to make adult-like object category judgements (Imai & Gentner, 1997; Samuelson & Smith, 1999). Further, toddlers' productive vocabulary is positively correlated with their ability to categorise based on object labels, rather than based on perceptual information alone (Jaswal & Markman, 2007). Although evidence suggests a relationship between vocabulary and categorisation, however, how experience with object categories may influence word learning remains unclear.

Learning labels for object categories is a complicated process, involving both fast and slow mapping (Capone & McGregor, 2005; Horst, McMurray, & Samuelson, 2006). The first time a novel name is encountered, the child quickly forms an initial, rough hypothesis of the word's meaning — hence the term *fast mapping* (Carey, 1978). For example, when presented with a boat, a cup and a novel kazoo and asked for the

cheem, a 2-year-old child can reliably determine that *cheem* refers to the KAZOO.

However, simply forming this initial mapping does not mean that the child has really learned the word (Horst & Samuelson, 2008; Riches, Tomasello, & Conti-Ramsden, 2005) and in some contexts processing demands might prevent young children from learning the correct referent after only a single exposure (e.g., Mather & Plunkett, 2009).

In fact, full word learning emerges gradually during a period of slow mapping (Capone & McGregor, 2005; Horst & Samuelson, 2008; Munro, Baker, McGregor, Docking & Arculi, 2012). During this phase repeated encounters allow the child to strengthen the label-object association (Mather & Plunkett, 2009; see also Smith & Yu, 2008). Specifically, the statistical co-occurrence of the labels and their referents strengthens these associations (Horst, et al., 2006). For example, the *cheem*-KAZOO mapping will be strengthened each time the child hears the word *cheem* and sees the kazoo in a new context, such as seeing the kazoo with a car and later with blocks.

The effect of repetition: single versus multiple exemplars

However, children rarely encounter a single category exemplar repeatedly, but rather encounter multiple, different exemplars over time (Smith & Yu, 2008). For example, a child might play with a green kazoo at nursery, then a red kazoo at home and later a yellow kazoo at Grandma's. Whether repeated exposure to a single exemplar leads to more robust word learning than exposure to several different exemplars is unknown. On the one hand, encountering the same object repeatedly could facilitate word learning by providing children with more experience with that individual exemplar, thus strengthening the memory trace for the individual item. On the other hand, experience with multiple exemplars from the same category could invite comparison between exemplars, highlighting category-general properties and triggering

the formation of a robust category representation (Nosofsky, 1984; Quinn, 2005; Rosch, 1975).

Although encountering multiple exemplars versus a single exemplar repeatedly has been shown to facilitate categorisation in various domains (Kovack-Lesh & Oakes, 2007; Quinn & Bhatt, 2005; Rost & McMurray, 2009; Wilcox, Smith, & Woods, 2011; Younger & Cohen, 1983), it is unclear whether this effect extends to situations involving learning category labels: relatively few studies of children's word learning have examined the effect of repetition. However, a handful of recent studies have begun to explore whether single or multiple exemplars facilitate word learning – with mixed results.

Some studies support the hypothesis that multiple exemplars facilitate word learning. For example, toddlers taught a colour-word vocabulary that included six exemplars were better able to generalise novel colour labels to new instances than toddlers trained on vocabularies that included fewer exemplars (four or two, Thom & Sandhofer, 2009). Similarly, preschoolers presented with two exemplars, rather than one exemplar, were better able to extend a novel superordinate category label to items from the same taxonomic category (Liu, Golinkoff, & Sak, 2001; see also Gentner & Namy 1999).

In contrast, other recent results suggest that multiple exemplars are not essential for word learning: Mather & Plunkett (2009) demonstrated that children could learn novel labels for novel categories via mutual exclusivity with repeated exposures to single stimuli; and Maguire, Hirsch-Pasek, Golinkoff & Brandone (2008) found that encountering multiple actors in a verb learning task hindered children's ability to recall and generalise novel verbs. Further, in a word learning study focusing on the effect of spaced versus massed presentation, 3-year-old children reliably mapped novel labels to

their referents at test, but did so better after familiarisation with single rather than multiple exemplars (Vlach, Sandhofer, & Kornell, 2008). However, children in the two conditions of this last study encountered identical objects at test, and test trials were therefore not equivalent between conditions; that is, half of the children were asked to recall a previously fast-mapped exemplar while half of the children were asked to extend a previously fast-mapped label to a new exemplar. Comparison between groups in this study is therefore difficult.

The scope of the effect of variability: narrow versus broad categories

Evidently, the question of whether or not exposure to multiple exemplars helps or hinders noun learning remains open. Even less well-understood is the extent of any effect of category variability on word learning; that is, is it easier for young children to learn a label for a high-variability category or a low-variability category? In a longitudinal category training study, toddlers who encountered perceptually variable exemplars experienced a significant acceleration in vocabulary growth and were able to make adult-like label generalisations, in contrast to children who encountered perceptually similar exemplars (Perry, Samuelson, Malloy, & Schiffer, 2010). Conversely, in a novel noun generalisation task, after being given novel labels for parts of novel objects or animals, preschool children prefer to generalise novel labels to highly similar test items than to highly variable test items (Gentner, Loewenstein, & Hung, 2007). However, the effect of within-category variability on novel category label retention and extension has yet to be systematically explored.

The effect of presentation order: blocked versus mixed presentation

Just as questions remain as to the effect of variability on children's word learning, little is known about the effect of the order in which exemplars are encountered. Research in the adult domain has reached a consensus that spacing

learning events over time leads to more robust learning than a massed learning regime. Indeed, a meta-analysis of 63 adult studies concluded that the “spacing effect” was robust, albeit subject to task effects (Donovan & Radosevich, 1999). Despite this substantial body of research, however, few studies have investigated the spacing effect in children. Of the studies that do exist, results are contradictory. Some are consistent with the adult findings: for example, two-year-old children have been shown to retain novel nouns and verbs more reliably when presented with an exemplar daily for eight days than when presented with eight exemplars in one day (Childers & Tomasello, 2002, see also Merriman, Rovee-Collier & Wilk, 1997). Similarly, Vlach, Sandhofer and Kornell (2008) demonstrated that three-year-old children retained novel nouns significantly better when presented with three (identical or multiple) exemplars spaced by 30s breaks than when presented with the same three exemplars successively (although, as noted, interpretations of these results is difficult). Finally, three-year-old children demonstrated better word learning in a fast mapping task in which they encountered two novel objects successively (that is, across trials) than concurrently (that is, two novel objects on a single trial; Wilkinson, Ross & Diamond, 2003). Thus, there is evidence that children’s word learning is facilitated not only by spacing presentation of exemplars over intervals of days, but also over intervals of seconds.

Equally, however, categorisation studies suggest that children’s learning benefits from being able to compare simultaneously-presented stimuli. For example, 4-month-old children can discriminate between the categories DOG and CAT when familiarised with stimuli side-by-side in a paired comparison task, but not when familiarised with the same stimuli for an equal amount of time in a successive presentation task (Oakes & Ribar, 2005). From this perspective, concurrent presentation allows children to rapidly compare between stimuli, reducing demands on visual short-term memory, drawing

attention to commonalities among objects and allowing quick encoding. Stimuli presented at greater intervals are likely to place higher demands on visual short-term memory, and are consequently more difficult to compare and encode (Kovack-Lesh & Oakes, 2007).

Clearly, the methods and empirical questions of existing studies vary greatly. The effect of encountering single versus multiple exemplars, or narrow versus broad categories, in mixed or blocked presentation is therefore difficult to predict. In the current study we address this by systematically manipulating these factors across two experiments. In Experiment 1, we ask whether children's ability to learn labels for novel object categories is affected by exemplar repetition. Children repeatedly mapped novel names to either the same or multiple exemplars. Specifically, children encountered a novel category label three times during a block of six referent selection trials. Half of the children were repeatedly presented with the same exemplar across trials, and half of the children were presented with multiple category exemplars across trials. If encountering the same exemplar repeatedly facilitates word learning, then these children should demonstrate better retention of novel category labels at test. In contrast, if encountering multiple exemplars facilitates word learning, then these children should demonstrate better retention.

Experiment 2 addresses two issues in the word learning literature: variability and order of presentation. In Experiment 2A, we ask how within-category variability affects children's category label learning. All children encountered multiple exemplars; however, half were presented with narrow categories and half were presented with broad categories. If limited variability among category exemplars facilitates retention of newly-formed label-object mappings, then children in the *narrow* condition should retain better than children in the *broad* condition. However, if higher variability

facilitates retention, children in the *broad* condition should retain better than children in the *narrow* category.

Children were also asked to extend newly-learned category-label mappings to never-before-seen exemplars (see also Horst & Samuelson, 2008) to examine more whether variability and order of presentation also affect noun extension. Because both recall and generalisation require children to store a representation of previously fast-mapped labels, we predicted that variability should have a comparable effect during extension trials as during retention trials. That is, if limited variability facilitates extension of newly-learned label-object mappings to never-before seen exemplars, children in the *narrow* condition should extend more accurately than children in the *broad* condition, and if high variability facilitates word learning, children in the *broad* condition should extend more accurately than children in the *narrow* condition.

In Experiment 2b, we ask whether order of presentation affects children's category label learning. Children received the same referent selection trials as in Experiment 2a, but encountered these trials in pseudorandom (mixed) presentation. If learning is facilitated by intervals between learning events from a given category, then children who encountered mixed trials should demonstrate better retention and extension than children who encountered blocked trials.

Experiment 1

Method

Participants. Twenty-four typically developing, monolingual, English-speaking 30-month-old children (13 girls, $M = 30\text{m}, 17\text{d}$, $SD = 43.19$ days; range = 28m, 1d - 32m, 26d) with a mean productive vocabulary of 563 words ($SD = 81.91$ words, range = 391 - 668 words) and no family history of colourblindness participated. Children were from predominantly middle class homes. Half of the children were

randomly assigned to the *single* condition, and the other half were randomly assigned to the *multiple* condition. Children's ages and productive vocabularies did not differ between conditions. Data from two additional children were excluded from analyses due to fussiness and experimenter error. Caregivers were reimbursed for travel expenses and children received a small gift for participating.

Stimuli. Eighteen toy objects, chosen because they are highly familiar to 2-year-old children, served as known objects: six animals (bird, chicken, elephant, fish, giraffe, lion), six vehicles (boat, bus, car, motorcycle, plane, train), and six household objects (block, chair, comb, cup, toy mobile phone, spoon).

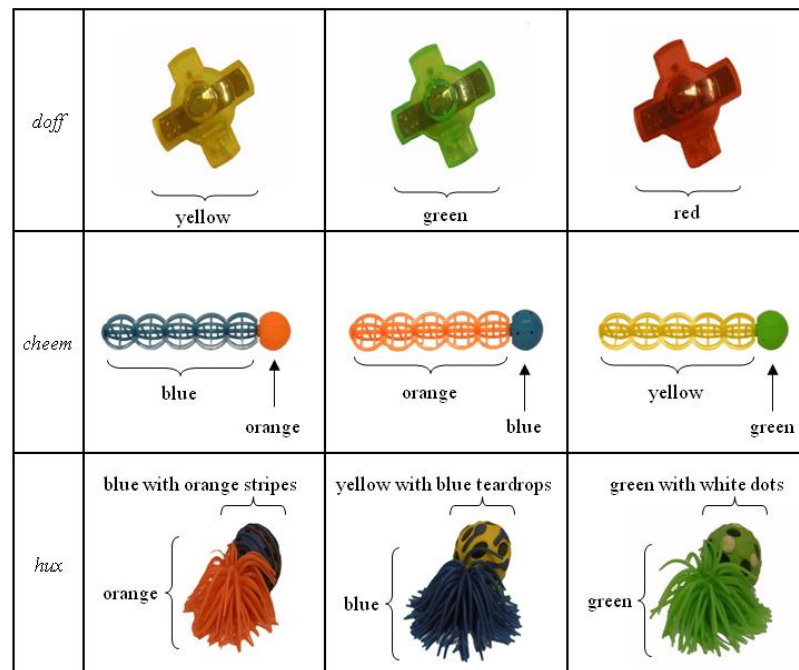


Figure 1. Novel objects used in Experiment 1.

Nine novel objects from three categories, chosen because they are not easily named by 2-year-old children, served as the target objects (see Figure 1). Consistent with other studies (e.g., Ankowski, Vlach & Sandhofer, 2021; Vlach et al., 2008), the objects in these categories varied in colour, but shared the same shape and material/texture. The DOFF category consisted of transparent plastic, plus-sign shaped tops in green, red and yellow. The CHEEM category consisted of blue, orange or yellow

plastic rods with small balls on one end in orange, blue or green, respectively. The HUX category consisted of rubber balloons with elastic strings hanging down in blue/orange, green/white and yellow/blue. The balloons contained foam balls and therefore kept their spherical shape. Stimuli were presented on a 46cm x 24cm white, wooden tray divided into three even sections. A digital kitchen timer was used to time the five-minute delay.

Procedure and Design. Before the experiment began, the experimenter showed the caregiver a booklet of colour photographs of the known and novel objects to ensure they were familiar and novel, respectively (which they were for all children). If the child knew a different label for an object (e.g., “kitty” vs. “cat”), the experimenter used that label.

During the experiment, children were seated in a booster seat across from the experimenter at a white table. Caregivers sat next to their children and completed a vocabulary checklist (Klee & Harrison, 2001) and were instructed to avoid interacting with their children, but to encourage them to respond during the warm-up trials if necessary. None of the children needed encouragement after the warm-up trials.

Warm-up trials. Each session began with three warm-up trials to introduce children to the task. On each trial, children were presented with three randomly selected known objects. First, the experimenter placed the tray of objects on the table and silently counted for three seconds to give the child an opportunity to look at the objects (see Horst, Scott, & Pollard, 2011). The experimenter then asked the child to select an object by naming it twice (e.g., “Can you find the block? Can you get the block?”) before sliding the tray forward. Children were praised heavily for correct responses and corrected if necessary. Between trials the experimenter replaced the tray on her lap and arranged the objects for the next trial out of the child’s view. The same objects were

presented on each warm-up trial, but object positions (left, middle, right) were pseudo-randomised across trials. Thus, children were asked for a different object in a different position on each trial. Warm-up stimuli were later used as known objects during the referent selection trials (see Horst & Samuelson, 2008).

Referent Selection Task. Referent selection trials immediately followed the warm-up trials and proceeded in the same manner, except that children were neither praised nor corrected. After each choice, the experimenter either said nothing or simply “OK” or “thank you.” Each child was presented with nine sets of stimuli and saw each set once on a known label trial and once on a novel label trial across a total of 18 trials. Each set included two familiar objects (e.g., boat and cup) and one novel object (e.g., top). Children in the *single* condition saw the same exemplar in each set. For example, a child might see the green top with the boat and cup, and again with the block and lion and once more with the car and spoon. Children in the *multiple* condition saw a different novel exemplar in each set. For example, a child might see the green top with the boat and cup, the red top with the block and lion and the yellow top with the car and spoon. Thus, the only difference between conditions was whether children saw one or three exemplars for each category.

Referent selection trials were presented in three blocks. For example, one child completed all trials with the top category, then all trials with the rod category and finally all trials with the balloon category. Block order was counterbalanced across participants using a Latin Square design. The order of known and novel trials was pseudorandomised in each block such that the same set was never presented on two consecutive trials and no more than two trials of either type (i.e., known or novel) were presented sequentially. Object position (left, middle, right) was randomly determined on each trial. Between the referent selection task and the retention task the child

remained at the table and coloured pictures from a colouring book during a five-minute delay period.

Retention Task. The retention task was the same across conditions. First, to re-engage children in the task, a new warm-up trial with three different known objects was presented. This was immediately followed by three retention trials, during which children saw one novel exemplar from each novel category encountered during referent selection (top, rod, balloon). The same exemplars were presented on each trial for a given child. In the *single* condition, children were presented with the exemplars encountered during referent selection. In the *multiple* condition, presentation of exemplars was counterbalanced across participants (i.e., which top, which rod, which balloon). Object positions were randomised across trials such that children were asked for a different novel object in a different position on each trial.

For all test trials reported in this paper, children's responses were included in analyses irrespective of whether they correctly mapped novel labels to novel exemplars during referent selection for that category. Theoretically, we argue that word learning via fast mapping is a slow process, and that children encode information about exemplars and labels even on incorrect referent selection trials (see Munro et al., 2012). However, analyses excluding test trials for a given category following (a) zero correct referent selection trials or (b) one or fewer correct referent selection trials generated the same overall pattern of results.

Coding. Children's responses were coded offline from DVD. A second coder blind to the experimental hypotheses coded 20% of the sessions for reliability. Inter-coder agreement was high, $M = 98.08\%$, $SD = 3.44\%$ (range = 92.31% – 100.00%).

Results and Discussion

As can be seen in the left panel of Figure 2, children in both conditions were very accurate at choosing the target object during the referent selection task. Thus, children's proportion of target choices was the same in each condition and greater than would be expected by chance (.33), $t(11) = 71.73$, $p < .0001$, $d = 20.60$ (all p s are two-tailed). On novel label referent selection trials, children's proportion of target choices was also greater than expected by chance (.33) both for children in the *single* condition, $t(11) = 4.59$, $p < .001$, $d = .84$ and for children in the multiple exemplars condition, $t(11) = 6.57$, $p < .0001$, $d = 2.38$. Again, there was no difference between conditions, $t(22) = .345$, ns . Thus, whether children encountered the same exemplar repeatedly or multiple exemplars during referent selection did not influence children's performance on either known or novel label referent selection trials.

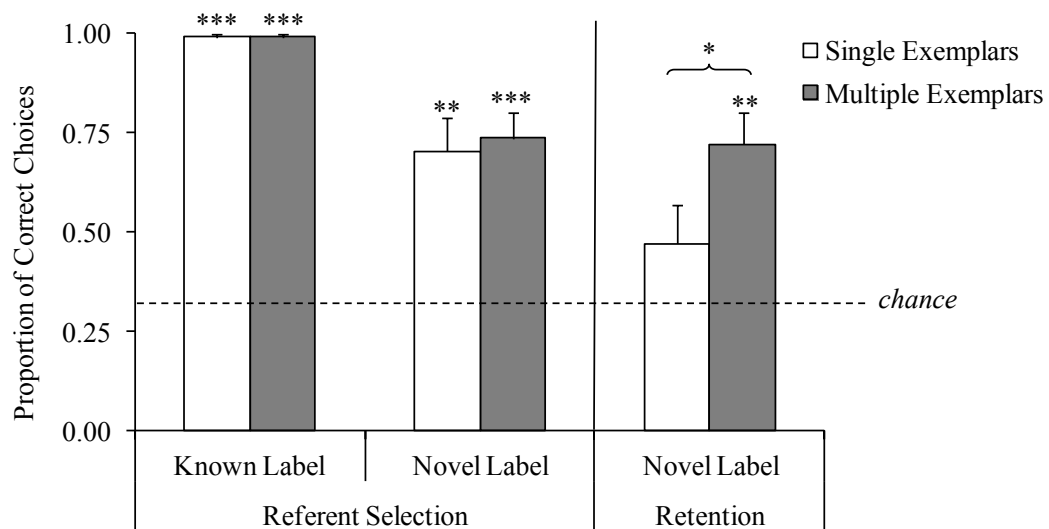


Figure 2. Children's proportion of correct choices in Experiment 1. Dotted line represents chance (.33). Error bars represent one standard error. *** $p < .0001$, ** $p < .001$, * $p = .05$. All p s are two-tailed.

In contrast, seeing multiple exemplars during referent selection did influence children's ability to retain previously fast-mapped novel labels after a five-minute delay.

As can be seen in the right panel of Figure 2, only children in the *multiple* condition retained more labels than expected by chance (.33), $t(11) = 5.00$, $p < .001$, $d = 1.46^2$. Children in the *single* condition failed to retain more words than expected by chance, $t(11) = 1.47$, ns , $d = .44$. An unpaired t test confirmed that children who encountered multiple exemplars retained significantly more words than children who encountered the same exemplars repeatedly, $t(22) = 2.06$, $p = .05$, $d = 0.89$. Note that children in the *multiple* condition were almost as accurate on the retention task ($M = 72\%$, $SD = 27\%$) as on the initial referent selection task ($M = 74\%$, $SD = 21\%$).

Importantly, in the *multiple* condition, children encountered a *category* of novel objects, while those in the *single* condition encountered the same object repeatedly. In the categorisation literature there has long been consensus that categories are collections of individual objects which share common features (Mandler, Fivush, & Reznick, 1987; Quinn & Eimas 1986; Rosch, 1975; Younger & Cohen 1985). In the DOG category, for example, the majority of members bark and have fur and four legs. Each individual exemplar, however, is discriminable from the other category members: a collie is a different shape and colour from a poodle or a dachshund. E1 demonstrates that children learn more words via fast mapping when presented with a category rather than the same object multiple times. However, the extent to which within-category variability influences word learning is unclear.

Thus, in Experiment 2a, we presented all children with multiple exemplars during the initial referent selection task. However, half the children were presented with narrow categories (i.e., low within-category variability) and the other half were presented with broad categories (i.e., high within-category variability). Further, we added extension trials to explore how encountering multiple exemplars during referent

selection may also influence how children extend novel labels to completely new category members.

Experiment 2A

Method

Participants. Twenty-four typically developing, monolingual, English speaking 30-month-old children (12 girls, $M = 30\text{m}, 5\text{d}$, $SD = 54.54$ days; range = 27m, 12d - 33m, 11d) with a mean productive vocabulary of 568 words ($SD = 106.71$ words, range = 294 - 665 words) and no family history of colourblindness participated. Children were from predominantly middle class homes. Half of the children were randomly assigned to the *narrow* condition and the other half were randomly assigned to the *broad* condition. Children's ages and productive vocabularies did not differ between conditions. Data from four additional children were excluded from analyses due to fussiness (2) and caregiver interference (2). Caregivers were reimbursed for travel expenses and children received a small gift for participating.

Stimuli. Eighteen toy objects, chosen because they are highly familiar to 2-year-old children, served as known objects: seven animals (cow, chicken, elephant, fish, giraffe, lion, turtle), six vehicles (boat, bus, car, motorcycle, plane, train), and five household objects (block, cup, fork, phone, spoon).

Eighteen novel objects from three categories, chosen because they are not easily named by 2-year-old children, served as the target objects (see Figure 3). Children in the *narrow* condition encountered three novel objects for each of the three categories during the referent selection and retention tasks. Like the objects shown to the children in the *multiple* condition in Experiment 1, the objects in these categories varied in colour but shared the same shape and material/texture (see Figure 3). For these children, the DOFF category consisted of three wooden castanets in green, red or yellow. The base

of each castanet matched the cover; each cover was patterned with flowers, hearts and hoops, respectively. The CHEEM category consisted of three small plastic kazoos in yellow, red, or blue. The HUX category consisted of three rubber pom-poms in purple, pink or blue with plastic caps on the fronds in purple, blue, yellow, pink and green.

Children in the *broad* condition also encountered three novel objects for each of the three categories during the referent selection and retention tasks (see Figure 3). For these children, the DOFF category consisted of three wooden castanets: one with a green base and a green, flowered cover, one with a red, pointed base with a yellow-and-black striped cover, and one with a pink, scalloped base with a yellow, pink-spotted cover. The CHEEM category consisted of three plastic kazoos: one small and yellow, one large and red and one large and blue with orange dots. The HUX category consisted of three rubber pom-poms in purple, blue or yellow. The purple pom-pom had caps on the ends in purple, blue, yellow, pink and green. The blue pom-pom had the same number of fronds as the purple pom-pom, but only half the number of caps (all purple). The yellow pom-pom had fewer fronds than the others but all the fronds had caps (green and pink).

Importantly, one exemplar from each category was seen by children in both conditions: the green castanet (DOFF), the yellow kazoo (CHEEM) and the purple pom-pom (HUX). The only difference between conditions was the within-category variability of the novel categories. For children in the *narrow* condition, the categories only varied in colour; however, for children in the *broad* condition the categories varied in colour, texture, size and slightly in overall shape. All children saw the remaining three novel objects on the extension trials: a green castanet with a ladybird top (DOFF), a brass kazoo (CHEEM) and a purple pom-pom with a balloon on top (HUX).






















<i>doff</i> narrow	 green	 red	 yellow
<i>doff</i> broad	 green	 red	 pink
<i>cheem</i> narrow	 yellow	 red	 blue
<i>cheem</i> broad	 yellow	 red	 blue
<i>hux</i> narrow	 multicolored	 multicolored	 multicolored
<i>hux</i> broad	 multicolored	 pink and green	 purple
<i>extension</i>	 green	 brass	 multicolored

Figure 3. Novel objects used in Experiment 2.

To confirm that the broad stimuli were more variable than the narrow stimuli, 18 adults from the university community provided similarity judgements for the novel

objects used in both experiments. Adults were tested individually in a quiet room and presented with the exemplar that was the same between conditions (e.g., the green castanet) and each of the other exemplars from that category (e.g., the red, yellow, flower-base and striped castanets) individually, one at a time. Adults were asked to rate how similar the two objects were on a scale of 1 (very similar) to 11 (not similar at all). The order in which the sets and exemplars were presented was counterbalanced across participants. Mean ratings are depicted in Table 1. An ANOVA with average similarity ratings with Stimuli (E1/*multiple*, E2/*narrow*, E2/*broad*, E2/*extension*) as a repeated measure yielded a main effect of Stimuli, $F(1.62, 27.59) = 143.98, p < .0001, \eta_p^2 = .89$ (Greenhouse-Geisser corrected). Pairwise comparisons (Bonferroni corrected, $\alpha = .008$) confirmed that the broad exemplars were less similar to each other than the narrow exemplars were to each other and also than exemplars used in Experiment 1 were to each other (both $ps < .0001$). There were no differences in the similarity ratings between the novel objects used on the extension trials and the broad exemplars, but the extension objects were more variable than the narrow exemplars ($p < .0001$). Finally, there were no differences in the similarity ratings between narrow exemplars used in Experiment 2 and the novel objects used in Experiment 1.

	Experiment 1	Experiment 2		
	Multiple Exemplars	Narrow Exemplars	Broad Exemplars	Extension Stimuli
Mean	2.87 (0.99)	2.54 (0.81)	7.30 (1.61)	7.11 (1.54)
Range	1.83-5.00	1.17-4.33	4.17-9.50	3.33-9.00

Table 1. Adult similarity ratings for the stimuli used in both Experiments. Standard deviations provided in parentheses.

Procedure and Design. All aspects of the experiment were identical to Experiment 1 with two exceptions. First, different novel stimuli were used so that within-category variability could be controlled. Second, after the retention trials children were presented with extension trials. These trials were identical to the retention trials except that the completely new exemplars from the novel categories were presented (see Fig. 3). As in previous studies, these trials always occurred after the retention trials (e.g., Horst & Samuelson, 2008).

Coding. Children's responses were coded offline from DVD. A second coder blind to the experimental hypotheses coded 20% of the sessions for reliability. Inter-coder agreement was high, $M = 97.80\%$, $SD = 3.19\%$ (range = 92.86% – 100.00%).

Results and Discussion

As can be clearly seen in the left panel of Figure 4, children in both conditions were very accurate at choosing the target object during the referent selection task. On known label referent selection trials, children's proportion of target choices were greater than would be expected by chance (.33) for both children in the *narrow* condition, $t(11) = 10.56$, $p < .0001$, $d = 3.05$, and children in the *broad* condition, $t(11) = 17.51$, $p < .0001$, $d = 5.05$. There was no difference between conditions, $t(22) = 0.28$, *ns*. On novel label referent selection trials, children's proportion of target choices was also greater than expected by chance (.33) for both children in the narrow exemplars condition, $t(11) = 5.99$, $p < .0001$, $d = 1.73$ and children in the broad exemplars condition, $t(11) = 15.67$, $p < .0001$, $d = 4.52$. Again, there was no difference between conditions, $t(22) = 0.63$, *ns*. Thus, whether children encountered narrow or broad categories during referent selection did not influence performance on either known or novel label referent selection trials.

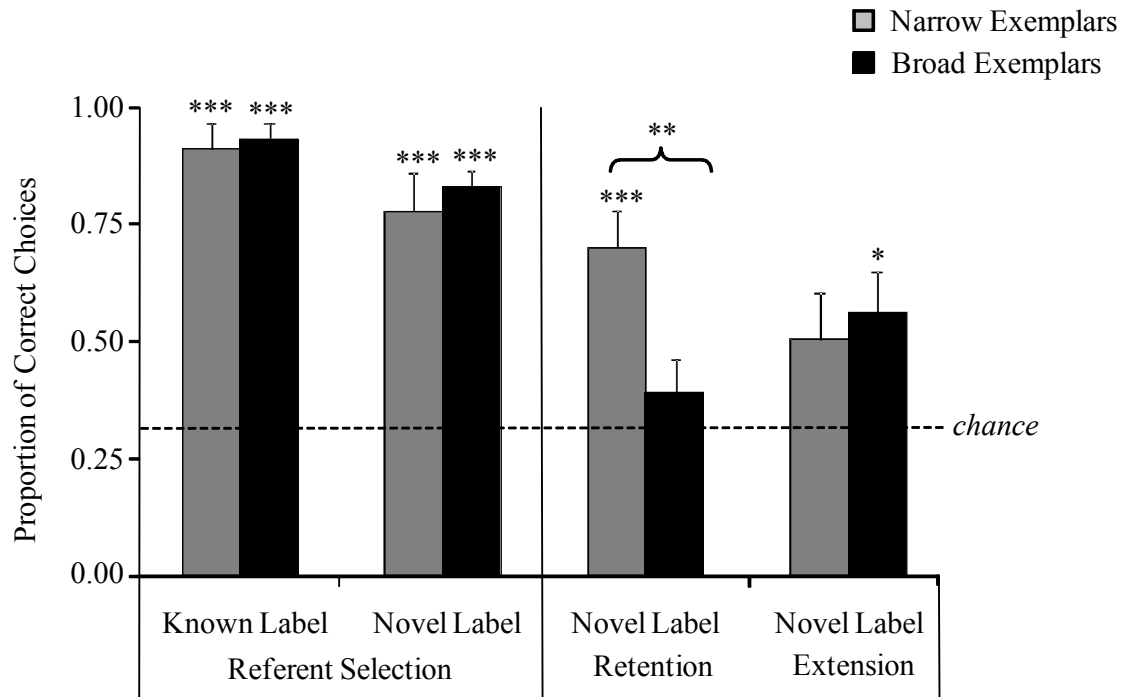


Figure 4. Children's proportion of correct choices in Experiment 2a. Dotted line represents chance (.33). Error bars represent one standard error. *** $p < .001$, ** $p < .01$, * $p < .05$.

In contrast, encountering narrow categories during referent selection did influence children's ability to retain previously fast-mapped novel labels after a five-minute delay. As can be seen in the right panel of Figure 4, only children in the *narrow* condition retained more labels than expected by chance (.33), $t(11) = 4.76$, $p < .001$, $d = 1.38$ (cf. E1, *multiple*). Children in the *broad* condition failed to retain more labels than expected by chance, $t(11) = 0.82$, ns , $d = 0.24$. An unpaired t -test confirmed that children who encountered narrow categories retained significantly more labels than children who encountered broad categories, $t(22) = 2.98$, $d = 1.27$.

However, contrary to our prediction, the opposite pattern was found for children's novel label extensions: encountering broad categories facilitated children's ability to extend previously fast-mapped novel labels. As can be seen in the right panel of Figure 4, children in the *narrow* condition failed to extend previously fast-mapped

novel labels to a never-before-seen category member, $t(11) = 1.76$, *ns*, $d = 0.51$. In contrast, children in the *broad* condition extended significantly more labels than expected by chance, $t(11) = 2.63$, $p < .05$, $d = 0.76$. However, the difference between conditions was not significant, $t(22) = 0.43$, *ns*, $d = 0.76$. Taken together, these data indicate that encountering low within-category variability facilitates retention of newly-encountered label-category mappings, and that high within-category variability facilitates extension of those mappings.

Initially, results from the *broad* condition are surprising: children reliably extend novel labels without showing any evidence of having retained them. To examine the possibility that the null result for retention in the *broad* condition constituted a Type 2 error rather than a genuine null result, the mean for retention in the *broad* condition (0.39) was used to generate a Bayes Factor (Dienes, 2011). The Bayes Factor represents the ratio between posterior probability (probability of a theory being true given observed data) and prior probability (expected probability of a theory being true before data collection, Dienes, 2008), calculated as in Equation 1:

$$(1) \quad P(h|d) = \frac{P(d|h)}{P(d|h)P(h)}$$

where $P(d|h)$ is the posterior probability and $P(h)$ is the prior probability. Bayes Factors of less than 1 suggest that the data support the null hypothesis; Bayes Factors of approximately 1 suggest that the experiment was not sensitive; Bayes Factors of greater than 1 support the experimental hypothesis over the null.

Our hypotheses were as follows:

H_0 = children do not retain novel labels after encountering broad categories.

H_1 = children retain novel labels after encountering broad categories.

For our prior probability, we specified a normal distribution of plausible scores with a mean proportion of correct choices of 0.67 (the average of the mean proportion retained

in E1 *multiple* and E2a *narrow*) and a standard deviation of 0.16, ranging from 0.33 (chance performance) to 1.00. The data of interest (from the retention trials in the *broad* condition) were corrected for small sample size (Dienes, 2008) resulting in a sample SE of .097. Sample mean difference was 0.06 (that is, $\text{mean}_{\text{observed}} - \text{mean}_{\text{expected}}$, $0.39 - 0.33$). The resulting Bayes Factor of 0.22 supports the null hypothesis. That is, children did not retain novel labels after encountering broad exemplars during referent selection.

Together, Experiments 1 and 2a demonstrate that within-category variability facilitates word learning, but only up to a point. When children encounter too much variability their retention of newly-encountered labels is inhibited. However, these findings may only hold when children encounter exemplars in blocked presentation; alternatively, encountering exemplars in mixed presentation may further support word learning via the spacing effect.

In Experiment 2b we examine the effect of temporally-distributed learning events by presenting novel label trials in a mixed order. If encountering distributed learning events facilitates learning, as predicted by the spaced learning literature, then children in Experiment 2b should retain and extend novel labels better than children in Experiment 2a. However, if encountering learning events in blocks facilitates comparison and encoding, as predicted by the categorisation literature, then children in Experiment 2a should retain and extend labels better than children in Experiment 2b.

Experiment 2B

Method

Participants. Twenty-four typically developing, monolingual, English speaking 30-month-old children (13 girls, $M = 29\text{m}, 25\text{d}$, $SD = 46.13$ days; range = 27m, 14d - 32m, 30d) with a mean productive vocabulary of 543 words ($SD = 128.76$ words, range

= 213 – 668 words) and no family history of colourblindness participated. Children were from predominantly middle class homes. Half of the children were randomly assigned to the *narrow* condition and the other half were randomly assigned to the *broad* condition. Children's ages and productive vocabularies did not differ between conditions. Data from one additional child were excluded from analyses due to fussiness (1). Caregivers were reimbursed for travel expenses and children received a small gift for participating.

Stimuli, procedure and design. Stimuli and procedure were the same as those in Experiment 2a; the design differed only in trial order. In contrast to Experiment 2a, children in Experiment 2b saw trials in a pseudorandom order such that they did not see more than two trials from the same category consecutively, or more than two familiar or novel trials consecutively.

Results and Discussion

As depicted in the left panel of Figure 5, children again performed at above-chance (0.33) levels during referent selection. On known label trials, children reliably selected the target exemplar in both conditions (*narrow*: $t(11) = 19.16, p < .001, d = 5.31$; *broad*: $t(11) = 4.54, p < .01, d = 1.36$). Similarly, on novel label trials children also reliably selected the target in both conditions (*narrow*: $t(11) = 9.17, p < .001, d = 3.15$; *broad*: $t(11) = 4.99, p < .001, d = 1.48$). However, children were unable to retain or extend newly-encountered label-category mappings after a five-minute delay (all $ps > .17$).

To examine whether mixed presentation had an effect on children's performance relative to blocked presentation, data from Experiment 2a and Experiment 2b were submitted to a MANOVA with Known Proportion Correct, Novel Proportion Correct, Retention Proportion Correct, and Extension Proportion Correct as dependent variables

and Variability (narrow, broad) and Presentation (blocked, mixed) as fixed factors³.

The MANOVA revealed significant effects of Variability ($F = 2.918, p < .05, \eta_p^2 = .22$) and of Presentation ($F = 3.60, p < .05, \eta_p^2 = .260$). Post-hoc univariate tests confirmed a significant effect of Variability on Retention Proportion Correct ($F(1,44) = 11.01, p < .01, \eta_p^2 = .20$). An unpaired t -test revealed that children who encountered *narrow* exemplars retained significantly more label-category mappings than children who encountered *broad* exemplars ($t(46) = 3.204, p < .01, d = 0.944$). Post-hoc analyses also revealed a significant effect of Presentation on Novel Proportion Correct ($F(1,44) = 9.28, p < .05, \eta_p^2 = .17$). An independent samples t -test (Bonferroni corrected, $\alpha = .0125$) revealed that children who encountered *blocked* presentation chose the target exemplar on novel referent selection trials significantly more often than children who encountered *mixed* presentation ($t(46) = 3.10, p < .01, d = 0.91$).

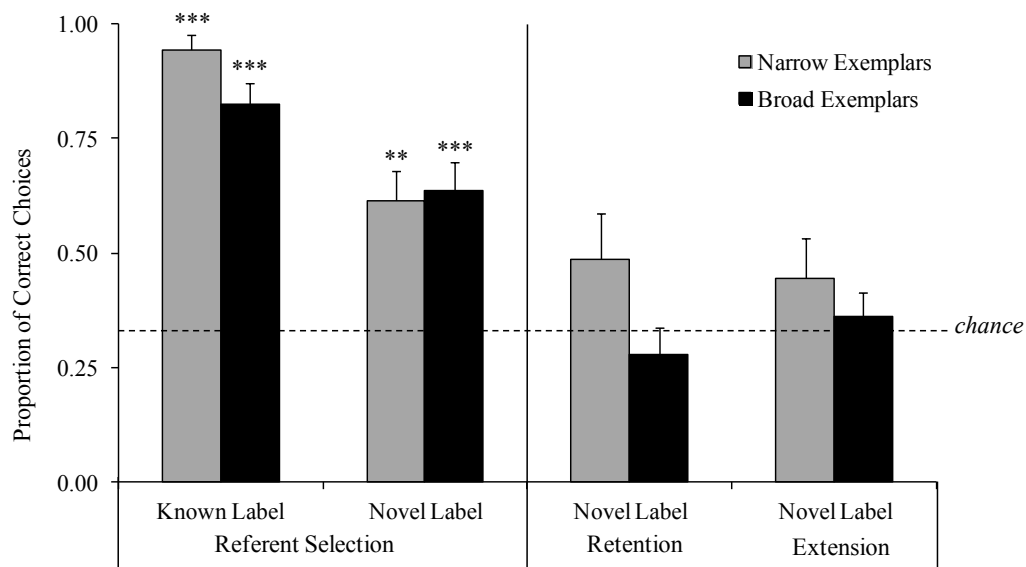


Figure 5. Children's proportion of correct choices in Experiment 2b. Dotted line

represents chance (.33). Error bars represent one standard error. *** $p < .001$, ** $p < .01$.

³ Preliminary analyses revealed a violation of the univariate assumption of sphericity (Mauchly's $W = .685, p = .006$) and significant correlations between dependent variables (known referent selection/novel referent selection, $r = .295, p < .05$; novel referent selection/extension, $r = .404, p < .01$, both 2-tailed). MANOVA therefore constitutes a more appropriate test than ANOVA, as multivariate tests do not assume sphericity (Field, 2009) and remain sensitive where dependent variables are correlated (Tabachnik & Fidell, 2001).

These data indicate that presenting trials in a mixed presentation during referent selection impairs children's ability to retain and extend novel category-label mappings, irrespective of the degree of within-category variability they encounter. That is, in E2b we found no spaced learning effect. However, in the context of the robust adult findings, this result may in fact suggest that the mixed presentation in Experiment 2b does not constitute spaced presentation. Indeed, the between-trial interval in all experiments presented here is a matter of seconds, whereas intervals between learning events in comparable experiments consisted of days (e.g., Childers & Tomasello, 2002).

Why, then, is word learning so much harder given mixed presentation? In the referent selection phases of the current experiment, blocked presentation allowed immediate comparison of representations across successive trials: that is, children needed only to retain a single exemplar-label mapping from one encounter to the next for any given category. In mixed presentation, however, children had to maintain representations of exemplars from up to three separate categories before encountering a new exemplar from the initial category, thus preventing immediate comparison (Gentner & Namy, 1999; Oakes et al., 2009). Further, children in the *mixed* conditions performed worse during novel referent selection than children in the *blocked* conditions. This is likely to be due to the reduced opportunity for learning from trial to trial: that is, children in the mixed condition may have had to solve each trial from scratch, rather than use existing representations of novel objects available in working memory from the previous trial. Put another way, mixed presentation made label-category mappings not only more difficult to remember, but also more difficult to make in the first place.

Overall, data from Experiment 2b suggest that the intervals between each encounter with a particular label-exemplar mapping were not sufficient to support the

spaced learning effect. Further, however, these intervals may also have been too long to allow children to maintain newly-formed label-exemplar mappings, preventing cross-trial comparison, and resulting in children's inability to retain or extend novel labels. Within-category variability helps word learning, then, but this effect is readily disrupted by changes in the task environment.

General Discussion

Across three experiments, we explored whether encountering multiple category exemplars facilitates word learning via fast mapping. In Experiment 1, we exposed 2-year-old children to three novel object categories via three blocks of referent selection trials. Across trials, children either encountered the same category exemplar repeatedly or multiple exemplars. Overall, all children did very well on the initial referent selection task. However, only children who encountered multiple exemplars retained previously fast-mapped novel labels after a short delay. Further, these children demonstrated significantly better retention than children who encountered the same exemplar repeatedly. In Experiment 2a, all children were presented with multiple exemplars, from either narrow (low variability) or broad (high variability) categories. Again, all children did very well on the referent selection task, however only children who saw narrow exemplars retained the novel labels. Moreover, only children who saw broad exemplars were able to extend the novel labels to never-before-seen category exemplars. In Experiment 2b, we exposed children to the same referent selection trials as in Experiment 2a, but in mixed presentation. Children were again successful in referent selection but unable to retain or extend the newly fast-mapped novel labels.

Comparison facilitates word learning

These data demonstrate that comparison of multiple exemplars facilitates word learning. Other studies that have explored the relationship between vocabulary and

categorisation have typically tested children over a timescale of several weeks (Ellis & Oakes, 2006; Perry et al., 2010). However, the current study reveals that exposing children to an object category, rather than a single category member, facilitates children's ability to learn the label for that category within minutes (see also Kemler-Nelson, O'Neil, & Asher, 2008).

We contend that children who encountered multiple (E1) or narrow exemplars (E2) retained category labels because perceptual variability across trials facilitated comparison between exemplars. Specifically, small amounts of perceptual variation (colour) highlighted the invariant features of the objects (shape, material and texture), whilst suppressing exemplar-specific features (Gogate & Hollich, 2010; Quinn & Bhatt, 2010; Rogers & McLelland, 2006). Thus in the *multiple* and *narrow* conditions, children were able to rapidly encode exemplars' invariant features, and robustly map novel labels to these representations such that children retained these labels over a five-minute delay. The results of E2b strongly support the idea that comparison is important for categorisation. Here, children did not have the opportunity to compare exemplars trial-by-trial, which clearly disrupted their ability to retain and extend novel labels.

A similar effect has been observed in 6-7-month-old infants using the preferential looking paradigm. Infants categorised stimuli organised in a bar configuration when familiarised with bar-shaped composites, the elements of which varied across trials (e.g., bars composed of crosses versus bars composed of circles, cf. multiple and narrow exemplars). When variability was encountered within trials (e.g., bars composed of crosses and circles, identical across trials, cf. single exemplars), infants did not form a category (Quinn & Bhatt, 2010). Thus, the current studies provide evidence for persistence of this phenomenon across development.

Data from extension trials in E2a are also in line with the categorisation literature (e.g., Quinn, Eimas & Rosenkrantz, 1993). Specifically, lack of variability in the *multiple* and *narrow* conditions led to formation of a narrow category, which did not include shape or texture/material variation. Representations of extension objects in E2a (which varied in shape and texture/material as well as colour) therefore fell outside the narrow category formed during referent selection, hence children's unwillingness to extend novel names in the narrow condition

However, data from the *broad* condition of E2a suggest that comparison may be a necessary but not sufficient condition for word learning: when within-category variability was high children were unable to retain category labels (but were able to extend labels to never-before-seen objects). These data support exemplar-based accounts of categorisation (e.g., Nosofsky, 1984; see J. D. Smith & Minda, 1984, for a discussion). According to these "exemplar theories", each time an exemplar is encountered, a separate, new representation is stored. Categories therefore consist of a set of individual representations, and new items are categorised via comparison with these stored representations. Importantly, stored representations degrade, such that over time, a given representation may become very different to its original (i.e., real-world) exemplar (Murphy, 2004).

From this perspective, during referent selection children formed novel categories consisting of six representations (one for each encounter with a novel exemplar), which degraded over the five-minute delay. In the E2a *broad* condition, representations degraded to the extent that they no longer matched the previously-encountered exemplars presented during retention trials, hence children's failure to apply newly-encountered novel labels to these exemplars (that is, the apparent lack of retention). In contrast, during extension trials, the degraded representations were similar enough to

the never-before-seen extension exemplars for children to generalise novel labels.

Further, due to the lack of shared, invariant features in the broad exemplars, children in the *broad* condition likely formed more diffuse category representations during referent selection, further hampering their performance in retention, but supporting generalisation to completely new exemplars.

Flexible categorisation

Clearly, children's word learning is influenced both in real time and across experience by different factors in different contexts. Here, both comparison and perceptual ability play a role in facilitating – or indeed disrupting – children's word learning abilities. Figure 6 provides a schematic explanation of how exemplar theory and comparison-based explanations apply to the data in E1 and E2A. In the top panel, depicting E1 *single*, no comparison occurs because exemplars are identical; encoding is therefore weak, and no words are retained. In the centre panel, depicting E1 *multiple*/E2a *narrow*, comparison highlights invariant features, promoting robust encoding and formation of a narrow category, and leading to successful retention but no extension. In the bottom panel, depicting E2a *broad*, lack of invariance causes weaker encoding and formation of a broad category, leading to lack of retention but successful extension.

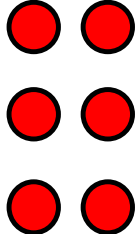
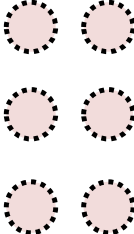

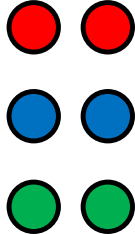
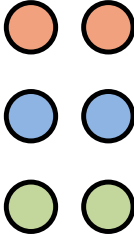


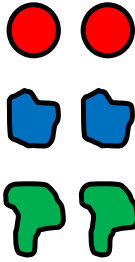
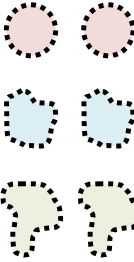


Condition	Representations formed during referent selection	Degraded representations after 5-minute break	Retention exemplar	Extension exemplar
<i>single</i> (E1)	"Cheem" x 3		"Where's the cheem?"	
				n/a
Lack of comparison → weak encoding. No retention.				
<i>multiple</i> (E1)/ <i>narrow</i> (E2a)	"Cheem" x 3		"Where's the cheem?"	"Where's the cheem?"
				 (Broad only)
Consistently invariant feature (shape) highlighted by comparison → strong encoding & narrow representation Successful retention, no extension.				
<i>Broad</i> (E2a)	"Cheem" x 3		"Where's the cheem?"	"Where's the cheem?"
				
Fewer invariant features → weaker encoding, & broad representation. No retention, successful extension.				

Figure 6. Exemplar-based account of the data in E1 and E2a.

Overall, these findings are consistent with the dynamic systems perspective on word learning (e.g., Elman, 2003; Samuelson & Horst, 2007; Samuelson, Smith, Perry, & Spencer, 2011), in which word learning emerges over a protracted period of slow

mapping from the interaction of multiple processes – in this case, previous experience with categories and in-task processes of comparison and encoding. In the current experiments, task effects clearly mattered: although every child received nine novel label referent selection trials, precisely which novel objects a child encountered during referent selection, and in which order those objects were encountered, strongly influenced that child's ability to learn and extend the object labels (see also Horst, et al., 2010; Kovack-Lesh, Horst, & Oakes, 2008; Oakes, Kovack-Lesh, & Horst, 2009; Quinn, Eimas, & Rosenkrantz, 1993; Samuelson & Horst, 2007).

The current studies add to a growing body of evidence that experience with variable categories influences young children's word learning. Importantly, these studies demonstrate that comparison and variability have a profound effect on children's word learning over a short time scale. The current studies are among the first to systematically investigate the dynamic interplay between category variability and word learning, and as such provide important groundwork for further research in the area, as well as inform our understanding of category learning and cognitive development, more generally.

Acknowledgements

This research was supported by ESRC Quota Award ES/H014942/1 to K.E.T. Portions of these data were presented at the 2009 Biennial Meeting of the Society for Research in Child Development (Denver, CO), the 2010 Biennial Meeting for the International Society for Infant Studies (Baltimore, MD) and the 2011 Society for the Study of Artificial Intelligence and Behaviour Convention (York, UK). We would like to thank Samantha Ranson, Jessica Abraham, Lauren Lush, Lucy Metcalfe and Philine Stein for help in data collection, subject recruitment and coding. We would also like to thank the caregivers and children who participated.

References

- Ankowski, A. A., Vlach, H. A., & Sandhofer, C. M. (2012). Comparison versus contrast: Task specifics affect category acquisition. *Infant and Child Development*. DOI: 10.1002/icd.1764/full
- Arias-Trejo, N., & Plunkett, K. (2010). The effects of perceptual similarity and category membership on early word-referent identification. *Journal of Experimental Child Psychology*, 105(1-2), 63-80. DOI: 10.1016/j.jecp.2009.10.002
- Capone, N. C., & McGregor, K. K. (2005). The effect of semantic representation on toddler's word retrieval. *Journal of Speech, Language, and Hearing Research*, 48(6), 1468-1480. DOI: 10.1044/1092-4388(2005/102)
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29. DOI: 10.1037/0012-1649.38.6.967
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967-978. DOI: 10.1037/0012-1649.38.6.967
- Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: The self-organization of skilled noun learning. *Developmental Science*, 11(2), 195-203. DOI: 10.1111/j.1467-7687.2007.00665.x
- Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. Basingstoke: Palgrave Macmillan.
- Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274. DOI : 10.1177/1745691611406920

- Ellis, A. E., & Oakes, L. M. (2006). Infants' flexibly use different dimensions to categorise objects. *Developmental Psychology*, 42(6), 1000-1011. DOI: 10.1037/0012-1649.42.6.1000
- Elman, J. (2003). Development: It's about time. *Developmental Science*, 6(4), 430-433. DOI: 10.1111/1467-7687.00297
- Field, A. P. (2009). *Discovering Statistics Using SPSS*. London: SAGE Publications Ltd.
- Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development*, 8(3), 285-307. DOI: 10.1080/15248370701446434
- Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, 14, 487-513. DOI: 10.1016/S0885-2014(99)00016-7
- Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117(2), 496. DOI: 10.1037/a0019049
- Gopnik, A., & Meltzoff, A. (1992). Categorisation and naming: Basic-level sorting in eighteen-month-olds and its relation to language. *Child Development*, 63(5), 1091-1103. DOI: 10.2307/1131520
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), 23-34. DOI: 10.1016/S0010-0277(02)00186-5
- Horst, J. S., Ellis, A. E., Samuelson, L. K., Trejo, E., Worzalla, S. L., Peltan, J. R., et al. (2009). Toddlers can adaptively change how they categorise: Same objects, same session, two different categorical distinctions. *Developmental Science*, 21(1), 96-105. DOI: 10.1111/j.1467-7687.2008.00737.x
- Horst, J. S., McMurray, B., & Samuelson, L. K. (2006). Online Processing is essential for learning: Understanding fast mapping and word learning in a dynamic

- connectionist architecture. *Proceedings from the 28th Meeting of the Cognitive Science Society*, 339-344. DOI: 10.1080/15250000701795598
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI: 10.1080/15250000701795598
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2011). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706-713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2), 169-200. DOI: 10.1016/S0010-0277(96)00784-6
- Jaswal, V. K., & Markman, E. M. (2007). Looks aren't everything: 24-month-olds' willingness to accept unexpected labels. *Journal of Cognition and Development*, 8(1), 93-111. DOI: 10.1207/s15327647jcd0801_5
- Kemler Nelson, D. G., O'Neil, K. A., & Asher, Y. M. (2008). A mutually facilitative relationship between learning names and learning concepts in preschool children: The case of artifacts. *Journal of Cognition and Development*, 9(2), 171-193. DOI: 10.1080/15248370802022621
- Kirkham, N. (2010). All together now: Learning through multiple sources. *Neoconstructivism: The new science of cognitive development*. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780195331059.003.0002.
- Klee, T., & Harrison, C. (2001). CDI Words and Sentences: validity and preliminary norms for British English. *Paper presented at the Child Language Seminar, University of Hertfordshire*.

- Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorisation. *Infancy*, *13*(4), 285-307. DOI: 10.1080/15250000802189428
- Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorisation. *Journal of Experimental Child Psychology*, *98*(2), 69-93. DOI: 10.1016/j.jecp.2007.05.001
- Liu, J., Golinkoff, R. M., & Sak, K. (2001). One cow does not an animal make: Young children can extend novel words at the superordinate level. *Child Development*, *72*(6), 1674-1694. DOI: 10.1111/1467-8624.00372
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628-634. DOI: 10.1111/j.1467-7687.2008.00707.x
- Mandler, J. M., Fivush, R., & Reznick, J. S. (1987). The development of contextual categories. *Cognitive Development*, *2*(4), 339-354. DOI: 10.1016/S0885-2014(87)80012-6.
- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*(1), 60-76. DOI: 10.1080/15250000802569702
- Merriman, J., Rovee-Collier, C., & Wilk, A. (1997). Exemplar spacing and infants' memory for category information. *Infant Behavior and Development*, *20*(2), 219-232. DOI: 10.1016/S0163-6383(97)90024-2
- Murphy, G. L. (2004). *The Big Book of Concepts*: Cambridge, Mass.: The MIT Press.

- Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General*, *131*(1), 5-15. DOI: 10.1037/0096-3445.131.1.5
- Nazzi, T., & Gopnik, A. (2001). Linguistic and cognitive abilities in infancy: When does language become a tool for categorisation? *Cognition*, *80*(3), B11-B20. DOI: 10.1016/S0010-0277(01)00112-3
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 104. DOI: 10.1037/0278-7393.10.1.104
- Oakes, L. M., Kovack-Lesh, K. A., & Horst, J. S. (2009). Two are better than one: Comparison influences infants' visual recognition memory. *Journal of Experimental Child Psychology*, *104*(1), 124-131. DOI: 10.1016/j.jecp.2008.09.001
- Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorisation in paired and successive presentation familiarisation tasks. *Infancy*, *7*(1), 85-98. DOI: 10.1207/s15327078in0701_7
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalisation and word learning. *Psychological Science*, *21*(12), 1894-1902. DOI: 10.1177/0956797610389189
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, *27*, 145-179. DOI: 10.1016/0010-0277(87)90017-5
- Quinn, P. C. (2005). Young infants' categorisation of human versus nonhuman animals: Roles for knowledge access and perceptual process. In L. Gershkoff-Stowe & D.

- H. Rakison (Eds.), *Building object categories in developmental time* (pp. 107-130). Mahwah, NJ: Lawrence Erlbaum Associates. DOI: 10.1111/j.0956-7976.2005.01567.x
- Quinn, P. C., & Bhatt, R. S. (2005). Learning perceptual organization in infancy. *Psychological Science, 16*(7), 511-515. DOI: 10.1111/j.0956-7976.2005.01567.x
- Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organization in infancy: The effect of simultaneous versus sequential variability experience. *Perception, 39*(6), 795-806. DOI: 10.1068/P6639.
- Quinn, P. C., & Eimas, P. D. (1986). On categorization in early infancy. *Merrill-Palmer Quarterly-Journal of Developmental Psychology, 32*(4), 331-363.
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception, 22*(4), 463-475. DOI: 10.1068/p220463
- Quinn, P. C., Slater, A. M., Brown, E., & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology, 19*, 207-218. DOI: 10.1348/026151001166038
- Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: Frequency and spacing effects. *Journal of Speech, Language, and Hearing Research, 48*(6), 1397-1411. DOI: 10.1044/1092-4388(2005/097)
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology-General, 104*(3), 192-233. DOI: 10.1037/0096-3445.104.3.192

- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339-349. DOI: 10.1111/j.1467-7687.2008.00786.x
- Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: Moment-to-moment interactions shape children's naming biases. *Infancy*, 11(1), 97-110. DOI: 10.1207/s15327078in1101_5
- Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition*, 110(3), 322-345. DOI: 10.1016/j.cognition.2008.10.017
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1-33. DOI: 10.1016/S0010-0277(99)00034-7
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, 6(12), e28095. DOI: 10.1371/journal.pone.0028095
- Singer-Freeman, K. E., & Bauer, P. J. (1997). Sorting out language and level: Examining the relation between productive vocabulary and category differentiation. *First Language*, 17(51, part 3), 241-270. DOI: 10.1177/014272379701705111
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411. DOI: 10.1037/0278-7393.24.6.1411
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: Contextually cued attention and early word learning. *Cognitive Science*, 34(7), 1287-1314. DOI: 10.1111/j.1551-6709.2010.01130.x

- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558-1568. DOI: 10.1016/j.cognition.2007.06.010
- Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, 108(3), 626-638. DOI: 10.1016/j.cognition.2008.05.002
- Tabachnik, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn & Bacon.
- Thom, E. E., & Sandhofer, C. M. (2009). More is more: The relationship between vocabulary size and word extension. *Journal of Experimental Child Psychology*, 104(4), 466-473. DOI: 10.1016/j.jecp.2009.07.004
- Tomasello, M., & Farrar, M. J. (1986). Joint Attention and Early Language. *Child Development*, 57(6), 1454–1463. DOI:10.2307/1130423
- Vlach, H., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109(1), 163-167. DOI: 10.1016/j.cognition.2008.07.013
- Waxman, S. R. (2003). Links between object categorization and naming: Origins and emergence in human infants. In D. H. Rakison & L. M. Oakes (Eds.), *Early Category and Concept Development: Making Sense of the Blooming, Buzzing Confusion*. (pp. 213-241). New York: Oxford University Press.
- Wilcox, T., Smith, T., & Woods, R. (2011). Priming infants to use pattern information in an object individuation task: The role of comparison. *Developmental Psychology*, 47(3), 886. DOI: 10.1037/a0021792
- Wilkinson, K. M., Ross, E., & Diamond, A. (2003). Fast mapping of multiple words: Insights into when “the information provided” does and does not equal “the

information perceived.” *Applied Developmental Psychology*, 24, 739–762. DOI: 10.1016/j.appdev.2003.09.006

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, 54(4), 858-867. DOI: 10.2307/1129890

Younger, B. A., & Cohen, L. B. (1985). How infants form categories. *Psychology of Learning and Motivation-Advances in Research and Theory*, 19, 211-247. DOI: 10.1016/S0079-7421(08)60528-9

Zulkipli, N., & Burt, J. (2012). *The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations*. *Memory & Cognition*, 1–12. DOI:10.3758/s13421-012-0238-9

Running head: EXEMPLAR VARIABILITY FACILITATES VERB LEARNING

Twinkle Twinkle Little Star:

Exemplar Variability Facilitates Early Verb Learning

Katherine E. Twomey & Jessica S. Horst

University of Sussex

Author note

A version of this paper is in preparation, to be submitted as follows:

Twomey, Lush, Pearce & Horst (in prep.). Twinkle Twinkle Little Star: Perceptual Variability Facilitates Early Verb Learning. *Infancy*.

Abstract

Infants' verb learning and action categorisation are little understood relative to their early noun learning and object categorisation. Evidence from the noun learning literature suggests that within-category variability facilitates noun learning, but the effect of variability on verb learning remains unclear. Across three experiments we demonstrate that noun and verb learning are subject to the same real-time, domain-general processes. In E1, adults rated the semantics of infants' early-learned verbs. Results indicate that early verb categories commonly encode path-based information. In E2, 24-month-old infants were familiarised with a single intransitive verb-action pair. At test, infants discriminated a new action from the same category, performed by the same actor. In E3, 24-month-old infants were habituated to intransitive verb-action pairs, encountering either the same pair each time (*single* condition) or three different exemplars from the same action category (*multiple* condition). At test, infants in both conditions showed evidence of having learned the mapping between verb and action. That is, infants discriminated the newly-learned mappings from novel exemplars. Taken together, these data suggest that exemplar variability affects verb categorisation, indicating that verb learning is governed by the same low-level perceptual processes as noun learning.

Twinkle Twinkle Little Star:

Perceptual Variability Facilitates Early Verb Learning

Young children acquire language with astonishing speed and apparent ease, while simultaneously learning to parse their perceptual environment into categories of objects, actions and events. Children are demonstrably adept categorisers from just a few months old (e.g., Arterberry & Bornstein, 2001; Bornstein & Mash, 2010; Quinn & Bhatt, 2009), and they begin the word learning task with substantial experience of using categories to structure their perceptual environment by generalising across exemplars (for a review, see Quinn, 2002). Moreover, children's remarkable ability to "fast map", or rapidly infer associations between new words and their correct referents, is well-documented (e.g., Carey & Bartlett, 1978; Childers & Tomasello, 2002; Markman, 1987; Seston, Golinkoff, Ma, & Hirsh-Pasek, 2009).

Word learning emerges from the interaction of fast mapping and categorisation over multiple timescales of development (Horst & Samuelson, 2008; Munro, Baker, McGregor, Docking, & Arculi, 2012). In the longer term, young children's vocabulary levels affect their categorisation (Ellis & Oakes, 2006), word learning (Gershkoff-Stowe & Hahn, 2007), and label generalisation (Jones, 2003; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Thom & Sandhofer, 2009); and children's experience with categories outside the lab affects categorisation in the experimental context (Kovack-Lesh, Horst, & Oakes, 2008). In the medium and short term, evidence from the noun learning literature suggests that within-category variability over time as well as in-task variability facilitate word learning (Perry, Samuelson, Malloy, & Schiffer, 2010; Twomey & Horst, Paper 1, this thesis).

The complexity of the verb learning task

Importantly, not all words are created equal: there are clear differences in noun and verb acquisition (Bornstein & Cote, 2004). For example, English-learning children acquire verbs later than nouns (McDonough, Song, Hirsh-Pasek, Golinkoff, & Lannon, 2011), leading some to suggest that children are initially biased to attend to objects over actions (Kersten & Smith, 2002), or that objects are “cognitively primitive” and easier to label than events or actions (e.g., the "Natural Partitions Hypothesis", Gentner, 1981). However, the precedence of nouns over verbs in order of acquisition is by no means universal. A growing body of studies reveals correspondence between verb learning and language-specific variables such as word order, vocabulary structure and caregiver speech (Choi & Gopnik, 1995; Imai, Haryu, & Okada, 2005; Tardif, Shatz, & Naigles, 1997). Such cross-linguistic differences suggest verb learning is heavily contingent on environmental context.

However, differences between verb and noun learning may also depend on semantics. In general, the referents of early-learned nouns are solid, bounded objects (Samuelson & Smith, 1999), visible in the environment, available for manual exploration and continuously perceptible. Verbs describe relations between objects (Haryu, Imai, & Okada, 2011; Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008), and as such the referents of early-learned verbs are in general actions or events, contingent on the participant(s) in those events, not independently visible, not available for physical exploration, and often only briefly perceptible. Thus, relative to learning nouns, learning verbs is a challenging task.

Based on the existing language and categorisation literatures, then, a child’s verb learning may depend on internal (e.g., vocabulary), external (e.g., perceptual concreteness), and developmental (e.g., categorisation experience) factors. However,

the mechanisms driving verb learning and action categorisation are less understood in comparison to the wealth of literature focusing on noun learning and object categorisation – and the question of whether nouns and verbs are learned via word-class-specific or domain-general processes remains open (for a discussion, see Maguire, Hirsh-Pasek, & Golinkoff, 2006).

The unclear effects of variability on verb learning

The current studies examine verb learning in the context of a recent investigation of the effect of variability on word learning, with the rationale that if word learning proceeds by domain-general processes, then verb categorisation should be affected by the same factors as noun categorisation. Specifically, Twomey & Horst (Paper 1, this thesis) exposed 30-month-old children to novel label-category pairs in which the category exemplars were the same object repeatedly (*single* condition), objects that varied in colour (*multiple* condition) or objects that varied in shape, texture and colour (*broad* condition). Children retained newly-learned word-category mappings in the *multiple* condition only, and generalised newly-learned mappings to completely novel exemplars in the *broad* condition only. Therefore, moderate variability appears to support noun learning, whilst high variability triggers formation of a broad category that includes never-before-seen category exemplars. Perry, Samuelson, Malloy & Schiffer (2010) examined exemplar variability in a longitudinal training study, with similar results: within-category variability facilitates word learning. Indeed, variability has been shown to benefit learning in a variety of domains and ages. For example, stimulus variability influences 6-7-month-old infants' categorisation of abstract visual stimuli (Quinn & Bhatt, 2010); facilitates 14-month-old infants' learning of lexical neighbours (Rost & McMurray, 2009); improves 13-month-old infants' categorisation in object

examination tasks (Oakes, Coppage, & Dingel, 1997) and speeds the emergence of novel strategies from embodied problem-solving in adults (Stephen & Dixon, 2009).

However, whether exemplar variability facilitates verb learning is not known. Perhaps surprisingly, actor variability has been shown to hinder rather than facilitate verb learning and generalisation (Maguire, et al., 2008). Specifically, in a pointing task, 2.5 and 3-year-old children generalised a novel intransitive verb to a new exemplar of a familiarised novel action only when the action was repeatedly performed by a single actor, and not when the action was performed by four different actors (see also Kersten & Smith, 2002). Similarly, in a test of the “similarity bootstrapping hypothesis”, 3- and 4-year-old children generalised transitive verbs only when both actor and novel object remained constant across exemplars (Haryu, Imai, & Okada, 2011). The authors argue that because actions (and verbs) are more complex than objects (and nouns), verb learning requires highlighting of relational similarity via consistency in lower-level perceptual features.

Unfortunately, however, comparison between existing studies is difficult, due to wide variation in terms of animacy and novelty of stimuli, and the argument structure of the verbs in question. For example, stimuli have consisted of nonhuman, animate agents performing intransitive verbs on nonhuman patients, (Kersten & Smith, 2002), a human agent performing an intransitive verb while holding a novel object inanimate participant (Haryu, et al., 2011), a human agent performing a transitive verb on a human patient (Yuan & Fisher, 2009), and so on. Clearly, it is difficult to compare the relative effects on verb learning of these different sources of variability.

Evidently, it is unclear whether or not exemplar variability helps children learn verbs, as predicted by the noun learning literature. The current studies examine infants’ early-learned verbs and the effect of exemplar variability on 24-month-old infants’ verb

learning. First, Experiment 1 examines the semantics of infants' early-learned verbs, based on the Macarthur-Bates Communicative Vocabulary Inventory (British English Adaptation) (MCDI; Klee & Harrison, 2001). Based on these findings, Experiments 2 and 3 employ rigorously-controlled intermodal stimuli to ensure results can be unambiguously interpreted (Oakes, 2010). Second, Experiment 2 demonstrates that infants can discriminate between exemplars from an intransitive verb category when the agent remains the same across exemplars. Finally, in Experiment 3 children are habituated with either invariant (*single* condition) or variable (*multiple* condition) verb categories and demonstrate that within-category variability does indeed affect on verb learning. A vocabulary analysis suggests that this effect emerges from the dynamic interaction between infants' in-task experience and their longer-term experience with verbs and action categories.

Experiment 1

Recent evidence suggests that children learning different languages may not encode aspects of a given action in the same way, due to cross-linguistic differences in verb semantics. For example, in a study with English-, Spanish- and Japanese-speaking participants, children generalised newly-learned verb-action pairs differently depending on their age and linguistic environment (Maguire, et al., 2010). Specifically, 24-month-old infants generalised a new verb from an animated character moving along a given path to an animated character moving along the same path, regardless of the manner in which this action was performed. That is, young children generalised novel verbs on the basis of path. However, older children (3-5 years old) had begun to generalise in line with their native verb semantics. Thus, at 24 months, English-learning children generalised to novel actions with the same path, but by 5 years generalised to novel actions with the same manner, in line with the underlying semantics of the majority

English verbs. These results suggest a developmental link between vocabulary structure and verb generalisation.

Similarly, in the noun learning literature, Samuelson and Smith (1999), conducted an analysis of the first 312 nouns in the MCDI (Fenson, Dale, Reznick, Thal, Bates, Hartung, et al., 1993). Based on the well-established finding that English-speaking children (and, indeed, adults) preferentially generalise novel nouns to new objects based on shape similarity (the "shape bias", Samuelson & Horst, 2008, but see also Markson, Diesendruck, & Bloom, 2008), the authors asked adults to rate each noun on three dimensions. First, raters judged whether each noun was a count noun (i.e., can be pluralised, takes "a", "the" or "those" as an article, and refers to a perceptually-bounded entity, e.g., *shoe*), a mass noun (i.e., cannot be pluralised, takes "some" as an article, and refers to a perceptually-unbounded entity, e.g., *milk*) or could be used in both senses, e.g., (*some crayon, those crayons*). Second, raters judged whether nouns labelled a solid object (e.g., *shoe*), a nonsolid object (e.g., *milk*), both (e.g., *ice cream*), or neither. Finally, raters judged whether nouns referred to objects that were categorised according to their shape (e.g., *shoe*), their material (e.g., *milk*), their colour (e.g., *lemon*), or any combination of these features. Of children's early-learned nouns, 110 were classified as count nouns, labelling solid objects which belonged to shape-based categories (for a replication with a UK sample see, Horst & Twomey, 2012). The authors argue that English-learning children's robust tendency to generalise novel nouns on the basis of shape was an emergent, probabilistic mechanism that had its roots in these statistical regularities in their early noun vocabularies (the "Attentional Learning Account"; see also Colunga & Smith, 2008).

Taking jointly the findings of Maguire et al. (2010) and Samuelson & Smith (1999) leads to two divergent predictions. Maguire et al. (2010) demonstrated that the

tendency to generalise verbs by manner emerges over development. Children's earliest verb generalisations, then, may reflect an intrinsic bias to extend novel verbs to new exemplars on the basis of shared path. Or, in line with the Attentional Learning Account, children's path-based verb generalisations may emerge from the statistical regularities (specifically, a preponderance of path-based verbs) in their early verb vocabularies.

Thus, Experiment 1 applies Samuelson and Smith's (1999) method to examine whether young children's bias towards path-based verb generalisations (Maguire et al., 2010) is reflected in their early-learned verbs. We asked adults to rate children's early-learned verbs according to whether they encoded an event's manner (for example, "the girl *ran* down the hill" vs. "the girl *jogged* down the hill") or its path (for example, "the girl *crossed* the path" vs. "the girl *followed* the path").

Method

Participants. Fifty-five native English speaking undergraduates (35 female, mean age = 18.89 y, $SD = 1.03$, range = 18 – 23y) participated for course credit. Data from one additional participant were excluded due to bilingualism. Participants were contacted via the University of Sussex online participant recruitment system.

Procedure. Participants rated a corpus of 107⁴ verbs from the "Action Words" section of the British English adaptation of the MCDI (Klee, et al., 2001) via an online questionnaire. Participants were asked to decide whether each verb encoded path, manner, neither, or any combination of these three options. The following examples were provided: "Path refers to the course followed by the person or object carrying out the verb, with respect to another object. For example, *circling* might be used to refer to how a dog moves around a fire hydrant (Maguire et al., 2010). Similarly, *crossing*

⁴ Note that due to a programming error, data from one verb, "listen," were excluded from the analysis. The MCDI - British English Adaptation lists 108 verbs.

might refer to how a woman moves from one side of a road to another. Manner refers to the way in which the person or object moves. For example, you could say ‘The woman walked/ran/scurried/jogged across the road’.” Order of presentation of words was randomised between participants, and words were presented in a single list.

Results

Percentages of ratings for each verb type are shown in Table 1.

Rating	Path	Manner	Neither	Ambiguous
Percentage (count) of verbs rated	53.28 (57)	26.17 (28)	14.95 (16)	5.61 (6)

Table 1. Percentage (count) of verb ratings

Samuleson & Smith (1999) rated nouns (e.g., “count nouns”) if and only if 85% of adults agreed on that rating. However, our agreement levels were less clear-cut, likely due to the semantic complexity of verbs relative to nouns. Thus, we classified verbs by frequency of rating. Verbs that received no one rating most often were classified as “Ambiguous”. No verbs were classified as Manner/Path, Manner/Neither, Path/Neither or Path/ Manner/Neither. As can be seen in Table 1, 57 of the 107 rated verbs were classified as path-based.

Overall proportion of rating types from the total of 5885 ratings obtained were submitted to a repeated measures ANOVA with Verb Type (path, manner, neither, path/manner, path/neither, manner/neither, path/manner/neither) as the within-subjects factor. The ANOVA revealed a significant effect of Verb Type ($F(1.982, 210.13) = 431.315, p < .0001, \eta_p^2 = .803$, Greenhouse-Geisser corrected). Planned pairwise comparisons confirmed that participants rated a greater proportion of verbs as path-based than as any other type (all $ps < .05$). Overall, these results indicate that early-

learned verbs predominantly encode motion path.

Discussion

According to adults' ratings, children's early learned verbs encode motion path significantly more frequently than they encode motion manner. Taking Maguire et al.'s (2010) finding in this context, children's path-based verb extensions may not be the result of a prelinguistic "path bias". Rather, early path-based generalisations may be a product of statistical regularities in young children's verb vocabularies: the path-dominated structure of children's early verb vocabularies may indicate to them that "new verbs are path-based" (see Smith, et al., 2002, for a corresponding account of noun generalisations). However, unlike the shape bias, which increases with development (e.g., Horst & Twomey, 2012; Landau, Smith & Jones, 1988) verb generalisations do not remain path-based, because as noted, older English-speaking children and English-speaking adults generalise novel verbs on the basis of manner (Maguire et. al, 2010). Indeed, English has been classified as an S-language, that is, one that categorises verbs according to manner (Slobin, 2003). Although outside the scope of the current studies, these data point to fruitful future research into the factors influencing this change in the basis of verb generalisation.

Experiments 2 and 3 explore the effect of within-category variability on verb learning by either familiarising (E2) or habituating (E3) children to exemplars of actions paired with a novel verb. In looking studies, children will look at a category of stimuli while encoding that category, and look away (or habituate) once the category has been learned. At test, new stimuli are presented. Low looking times indicate that the child has extended the learned category to include the new stimulus. High looking times indicate that the child has discriminated the new stimulus. Put differently, increases in looking to a given test stimulus are taken as evidence that the child perceives that

stimulus to be out-of-category (Oakes, 2010; Quinn, Eimas, & Rosenkrantz, 1993). The habituation paradigm offers an ideal method for investigating verb categorisation, as it permits both moving and multiple exemplars.

As noted, verb learning appears to be a fragile process, and little is known about the nature of children's verb categories. While even young infants can discriminate visual stimuli (e.g., Quinn, Bhatt, & Hayden, 2008) and map labels to object categories (Halberda, 2003; Horst & Samuelson, 2008; Waxman & Booth, 2001; Waxman & Braun, 2005), infants' ability to discriminate and map labels to action categories is comparatively fragile (Maguire, et al., 2006; Rakison, 2005). Thus, in Experiment 2, to ensure that 24-month-old infants are able to discriminate between exemplars from a single, intransitive verb category performed by a novel agent, we familiarised them with a single exemplar-verb mapping and tested their ability to discriminate a novel exemplar from the same category.

Experiment 2

Method

Participants. Twelve typically developing, English-learning 24-month-old infants (6 girls, $M = 24\text{m}, 22\text{d}$, $SD = 23.73$; range = 22m, 9d - 24m, 24d) took part. Participants had a mean productive vocabulary of 334.36 words ($M = 334.36$ words, $SD = 107.86$ words, range = 112 – 498 words) which included at least 1 verb ($M = 52.00$ verbs, $SD = 27.84$ verbs, range = 12 – 101 verbs). Data from five additional infants were excluded due to having no verbs in their productive vocabularies (2) and fussiness (3). Infants were from predominantly white, middle-class homes. Caregivers' details were collected via visits to nurseries in East Sussex, UK, and caregivers were contacted via email or telephone when their child's age approached 24 months. Caregivers' travel expenses were reimbursed and infants received a small gift for participating.

Stimuli. All familiarisation and test stimuli consisted of four five-second video clips. Each clip was looped to play continuously for a maximum of six times (30 seconds). Clips consisted of cartoon-style animations produced in Adobe Flash CS4, converted to Apple Quicktime format.

Figure 1 depicts the stimuli used in Experiments 2 and 3. Each clip consisted of a yellow star-shaped character with an orange outline and a smiling face, in the centre of a uniform grey screen. Data from Experiment 1 suggest that English-learning children will categorise (rather than discriminate) verb exemplars when they share the same path. Therefore, to ensure visual stimuli constituted meaningful and learnable verb referents, path was held constant, while manner provided within-category variability. Each five-second clip therefore consisted of the character travelling horizontally across the display, whilst changing shape at a constant rate. The character first travelled to the left extreme of the display, then returned to centre, then travelled to the right extreme of the display, and finally returned to centre again.

During this motion, the character shrank and changed into one of the four secondary shapes depicted in Figure 1: a circle, a triangle, a square, or a pentagon. Thus, the character was a different shape when it was halfway between the centre and the edge of the screen, and returned to its original shape by the time it reached the edge of the screen. Figure 2 depicts the time course of a single trial.



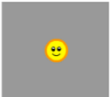
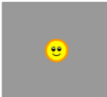







	Experiment 2	Experiment 3
Initial appearance		
Secondary shapes		
	same-category: circle	same-category: circle
		
	same-category: triangle	same-category: triangle
		
	same-category: square	same-category: square
		
	same-category: pentagon	same-category: pentagon
		
		new-category

Figure 1. Stimuli used in Experiments 2 and 3.

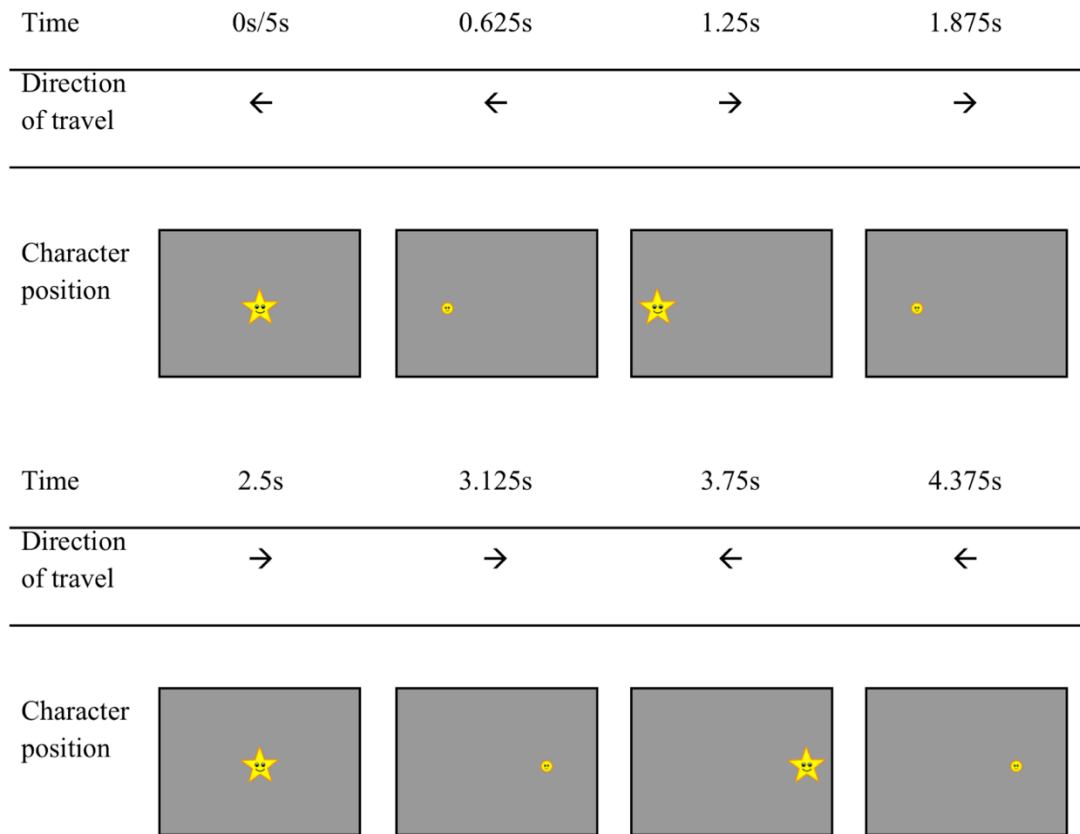


Figure 2. Time course of a single trial.

All familiarisation and test clips were accompanied by an auditory stimulus consisting of a recording of a British female voice saying “Look! He’s *dacking*! Watch him *dacking*!”. Auditory stimuli were recorded and edited for clarity using Audacity 1.2.6 sound editing software. The novel verb was chosen to be phonologically plausible in English and to follow CVC syllabic structure, which is readily discriminable by English infants from nine months (Saffran & Thiessen, 2003).

Two further video clips were used. Pre-familiarisation and post-test trials consisted of video of a novel purple toy being inverted by a human hand, adapted from Horst, Oakes & Madole (2005). The clip lasted six seconds and was looped to play continuously for a maximum of five times (30 seconds) per trial. This clip was accompanied by a whistling noise that rose and fell as the toy was inverted. Finally, an attention-getter was used between trials to reorient infants’ attention to the screen. This

consisted of a red circle on a white background that expanded and shrank repeatedly, accompanied by a staccato whistle.

Procedure and design. The experiment took place in a quiet, dark testing room and an adjoining coding room. The testing room contained a chair facing the main display and a 42-inch Samsung PDP television mounted in a ceiling-to-floor black fabric backdrop. The screen was raised 80cm from the floor. Below the screen, a camera recorded the infant's gaze. The lens of the camera was visible in the black fabric. Videos were displayed on the television screen in standard 4:3 format. Auditory stimuli were played over the television speakers with a volume of approximately 50dB. The adjoining coding room contained an Apple iMac G5 running OSX 10.4.11, running Habit 2000 (Cohen, Atkinson, & Chaput, 2000) which was used to control stimulus display and record infants' looking times. A small television displayed real-time closed-circuit video of the infant's visual responses to the stimuli.

Caregivers were seated in the testing room with the child on their lap, approximately 65cm away from the screen. Caregivers wore sunglasses with blacked-out lenses to prevent them from unconsciously of their child's responses, and were instructed to avoid engaging with their child during the experiment.

The experiment began with the attention-getter. Onset of the pre-familiarisation stimulus occurred immediately once the infant's gaze was oriented towards the screen. The pre-familiarisation stimulus played for 30 seconds or until the infant looked away from the screen for a minimum of 0.1 seconds, at which point the attention-getter played automatically. Looks away of less than 0.1 seconds were ignored and did not prompt the attention-getter to play. Familiarisation trials commenced immediately once the infant's gaze reoriented towards the screen, and proceeded in an identical manner.

Infants were familiarised with one of four action-verb pairs (see “Experiment 2” column, Fig. 1) over 8 trials.

Test trials immediately followed familiarisation trials, and again, proceeded in an identical manner. Test trials consisted of a baseline trial, in which the familiarised stimulus was presented, and a novel trial, in which a novel exemplar was presented. For example, a child might be familiarised with the “square” exemplar, and see the “circle” exemplar at test. Familiarisation and test exemplars were counterbalanced across participants such that each exemplar appeared alongside every other exemplar twice.

Following the experiment, caregivers were asked to complete the MCDI (British English Adaptation; Klee & Harrison, 2001).

Coding and reliability. Data were coded online by the experimenter. In addition, 20% of recordings were re-coded offline by a second naïve experimenter. The mean intercoder correlation was high, $r = .96$ (range = .97 – .99), and the mean absolute difference between coders was low (0.58s).

Results

Looking times for Experiment 1 are reported in Table 2.

Mean looking time (s)				
Trial 1:	Trials 2 - 9:	Trial 10	Trial 11	Trial 12
Prefamiliarisation	Familiarisation	Baseline	Novel	Post-test
			exemplar	
27.50 (4.76)	16.27 (5.87)	6.16 (3.94)	11.58 (7.96)	26.00 (8.53)

Table 2. Mean looking times in seconds for Experiment 2. Standard deviations are reported in parentheses.

Familiarisation phase. By the final familiarisation trials, infants' looking times had decreased: a paired t -test confirmed that looking during the first familiarisation trial was significantly greater than looking during the final familiarisation trial (all t -tests two-tailed, $t(11) = 16.43$, $p < .001$). However, to ensure that this decrease was the result of learning rather than fatigue, we compared looking time at the baseline to looking at the post-test stimulus, with the rationale that looking to the post-test stimulus should increase if infants were engaged in the experiment. A related-samples Wilcoxon's Signed Rank test⁵ confirmed this increase ($Z = -2.981$, $p < .01$), confirming infants' looking during familiarisation decreased due to learning.

Test phase. To explore whether infants' looking times recovered on encountering a novel stimulus, we compared looking times to the test trial with looking times to the baseline trial. A paired t -test revealed that infants looked significantly longer during the test trial than the baseline trial, $t(11) = 5.43$, $p < .05$. Thus, infants were able to discriminate between exemplars.

Discussion

Experiment 2 demonstrated that 24-month-old infants are able to discriminate between exemplars drawn from the same intransitive verb category⁶. Specifically, after familiarisation with an action category-verb pair infants' looking time to a novel exemplar from that category increased. Thus, children were able to discriminate between individual exemplars from the same category, confirming the validity of the stimuli for use in Experiment 3 (see Quinn, 1987).

⁵ As looking times to the completely novel stimulus were non-normal in both E2 and E3, appropriate nonparametric tests have been used for analyses including this variable.

⁶ Note that an identical experiment was conducted to assess the discriminability of a different set of stimuli. After testing five children, no evidence of discrimination between exemplars was found, (baseline vs. novel exemplar, $t(4) = 0.246$, $p = .81$, $d = .16$) despite indications of increased looking to the post-test stimulus relative to the baseline ($t(4) = 2.49$, $p = .07$). These stimuli were discarded, with the exception of one, which served as the "new category" stimulus in Experiment 2 (see Fig.1).

Experiment 3 examined the effect of category variability on infants' ability to learn and generalise verbs. Existing research demonstrates that category variability has a systematic effect on infants' noun learning. While too much variability interferes with infants' ability to learn novel nouns, some within-category variability facilitates noun learning (Twomey & Horst, Paper 1, this thesis). Further, Perry et al. (2010) demonstrated in a longitudinal study that training children with highly variable exemplars of object categories induced an increase in vocabulary relative to children trained with less variable exemplars. Although a recent study examined the effect of actor variability on infants' verb learning (Maguire, et al. 2008), the effect of exemplar variability on verb learning remains untested. If variability facilitates verb learning as it does noun learning, then verb and noun learning may be part of the same attentional learning system. If not, verbs and nouns may be acquired via different, specialised processes (for a discussion, see Maguire, et al., 2006).

Thus, in Experiment 3, we habituated infants to either single or multiple exemplars from a category of verbs, and tested their verb categorisation. Specifically, we employed a switch design (Werker, Cohen, Lloyd, Casasola, & Stager, 1998) to test whether infants would generalise a habituated verb-category pair to a new exemplar from the familiar category, or whether they would generalise to a new exemplar from a novel category, in the presence of either the habituated verb or a novel verb.

Experiment 3

Method

Participants. Thirty-six typically developing, English-learning 24-month-old infants (17 girls, $M = 23\text{m}, 24\text{d}$, $SD = 46.35$; range = 21m, 13d - 26m, 29d) took part. Participants had a mean productive vocabulary of 330.94 words ($SD = 192.90$, range = 61 words – 663 words), which included at least one verb ($M = 50.03$ verbs, $SD = 38.63$,

range = 1 verb - 124 verbs). Infants were from predominantly white, middle-class homes, and had not taken part in Experiment 2. Infants' ages, productive vocabularies and verb vocabularies did not differ between conditions. Caregivers' travel expenses were reimbursed and infants received a small gift for participating.

Data from 12 additional infants were excluded due to failure to complete (9), fussiness (2), chronic ear infections (1) and failure to habituate (1). Excessive fussiness was defined as crying, and/or refusal to remain still such that the child's eyes were not visible to the experimenter. The attrition rate of 25% is comparable to that reported in similar studies: a meta-analysis of 101 infant looking studies conducted between 1984 and 2005 revealed a mean attrition rate of 22% (range = 0.00% to 86.70%). Further, the meta-analysis found no correlation between exclusion rate and experimental outcome (Slaughter & Suddendorf, 2007).

Stimuli. Stimuli are depicted in the right-hand column of Figure 1. The same visual (*same-category*) stimuli and auditory stimuli (*same-verb*) stimulus used in Experiment 2 served as stimuli in Experiment 3. In addition, one further video clip (the *new-category* stimulus) was used during test trials. The *new-category* stimulus was identical to the *same-category* stimuli, except that rather than shrinking and changing global shape as it travelled across the display, the bottom two points of the star shape (its "legs") grew downwards and sideways. Speed and direction of movement as well as timing of the *new-category* stimulus were therefore the same as in the *same-category* stimuli. Finally, an additional auditory stimulus served as the *new-verb* stimulus. The novel verb *keefing* was recorded by the same female speaker and inserted into the same carrier phrase replacing the novel word *dacking* (again using Audacity 1.2.6). The only difference between the *same-verb* and the *new-verb* stimuli was the novel verb.

Procedure and design. Procedure, stimulus duration and display of pre-familiarisation stimulus, post-test stimulus and attention-getter were identical to Experiment 2, with the exception that children were habituated (rather than familiarised) to action-verb pairs via a maximum of 18 trials. Habituation was determined by a fixed three-trial window with a habituation criterion of 50% (that is, infants were considered to have habituated once looking times during a given three-trial window totalled 50% or less of looking times to the previous 3-trial window).

During habituation, infants were habituated either to a single visual stimulus repeatedly (*single* exemplars) or to blocks of three of the four action stimuli from Experiment 2 (*multiple* condition). Thus, in the *single* condition an infant might see the circle stimulus repeatedly, and in the *multiple* condition an infant might see the circle stimulus, then the square stimulus, then the triangle stimulus, then the square stimulus, and so on. In both conditions, the video clips were paired with the same-verb stimulus during habituation. All infants saw least six trials (and, in the *multiple* condition, infants saw each *same-category* exemplar at least twice). Each exemplar was seen equally often by children in either condition. In the *multiple* condition, each possible block of three exemplars (e.g., circle, square, triangle) was seen equally often across children. Trial order within blocks was pseudo-randomly determined for each infant with the constraint that no infant saw the same exemplar on two consecutive trials.

Immediately following habituation, infants were presented with a baseline measure consisting of either the habituated stimulus (*single*) or one of the three habituated stimuli (*multiple*; note the baseline consisted of a different stimulus than trial 18 for all infants in this condition). Four test trials immediately followed the baseline trial: (1) a novel *same-category* stimulus paired with the *same-verb* stimulus (“SCSV” e.g., circle + “*dacking*”); (2) novel *same-category* stimulus paired with the *new-verb*

label (“SCNV” e.g., circle + “*keefing*”); (3) the *new-category* stimulus paired with the *same-verb* stimulus (“NCSV” i.e., *new-category* + “*dacking*”); (4) the *new-category* stimulus paired with the *new-verb* label (“NCNV” i.e., *new-category* + “*keefing*”).

Order of presentation of test trials was counterbalanced across infants such that each trial type appeared in first, second, third or fourth position approximately equally often (e.g., the first test trial was SCSV five times, SCNV five times, NCSV four times, and NCNV five times).

Following the experiment, caregivers were asked to complete the MCDI (British English Adaptation; Klee & Harrison, 2001).

Coding and reliability. Data were coded online by the experimenter. In addition, 20% of recordings were re-coded offline by a second naïve experimenter. The mean intercoder correlation was high, $r = .94$ (range = .82 – 1.00), and the mean absolute difference between coders was low (0.92s).

Results

Preliminary analyses revealed no difference between conditions for total looking time.

Habituation phase. All infants habituated. A paired t -test comparing overall looking on the first block and overall looking on the last block confirmed the expected decrease in looking time over habituation, for both the *single* condition ($t(17) = 11.71, p < .001$) and the *multiple* condition ($t(17) = 14.32, p < .001$; all ps are two-tailed).

Related-samples Wilcoxon’s Signed Rank tests demonstrated that infants’ looking also recovered to the post-test stimulus relative to the baseline stimulus in both the *single* condition ($Z = -3.36, p < .01$) and the *multiple* condition ($Z = -3.46, p < .01$), confirming that the decrease in looking times was due to habituation as opposed to fatigue.

It is possible that either single or multiple exemplars were easier to learn, as would be indicated by a difference in looking time during habituation between the *single* condition ($M = 137.09$ seconds, $SD = 61.24$) and the *multiple* condition ($M = 137.77$ seconds, $SD = 100.48$). No such difference was found, ($t(34) = -1.32$, *ns.*, $d = -0.45$). Similarly, the number of trials taken to reach criterion did not differ between the *single* condition ($M = 8.83$ trials, $SD = 3.33$) and the *multiple* condition ($M = 9.83$ trials, $SD = 3.22$), $t(34) = -0.92$, *ns.*, $d = -0.32$). Thus, despite the extra perceptual variation in the *multiple* condition, infants encoded multiple exemplars equally as quickly single exemplars. Further, looking times to the post-test stimulus did not vary between the *single* condition ($M = 24.71$, $SD = 7.44$) and the *multiple* condition ($M = 27.93$, $SD = 6.49$), Kolmogorov-Smirnov $Z = 1.00$, *ns.*, $r = .17$, confirming that infants found single and multiple exemplars equally engaging.

Looking times during test trials. To examine the relative effects of changes in the visual and auditory stimuli on infants' looking times on individual test trials, we submitted looking times to a mixed ANOVA with Category (same, new) and Verb (same, new) as repeated measures and Condition (single, multiple) as a between-subjects factor. Mean looking times during baseline and test trials are depicted in Figure 3. The ANOVA revealed a significant main effect of Verb ($F(1,34) = 4.63$, $p < .05$, $\eta_p^2 = .12$). Planned comparisons (Bonferroni corrected) confirmed that overall, infants discriminated between test trials with the habituated verb “*dacking*” and test trials with the new verb “*keefing*” ($p < .05$), indicating that infants learned the verb presented during habituation.

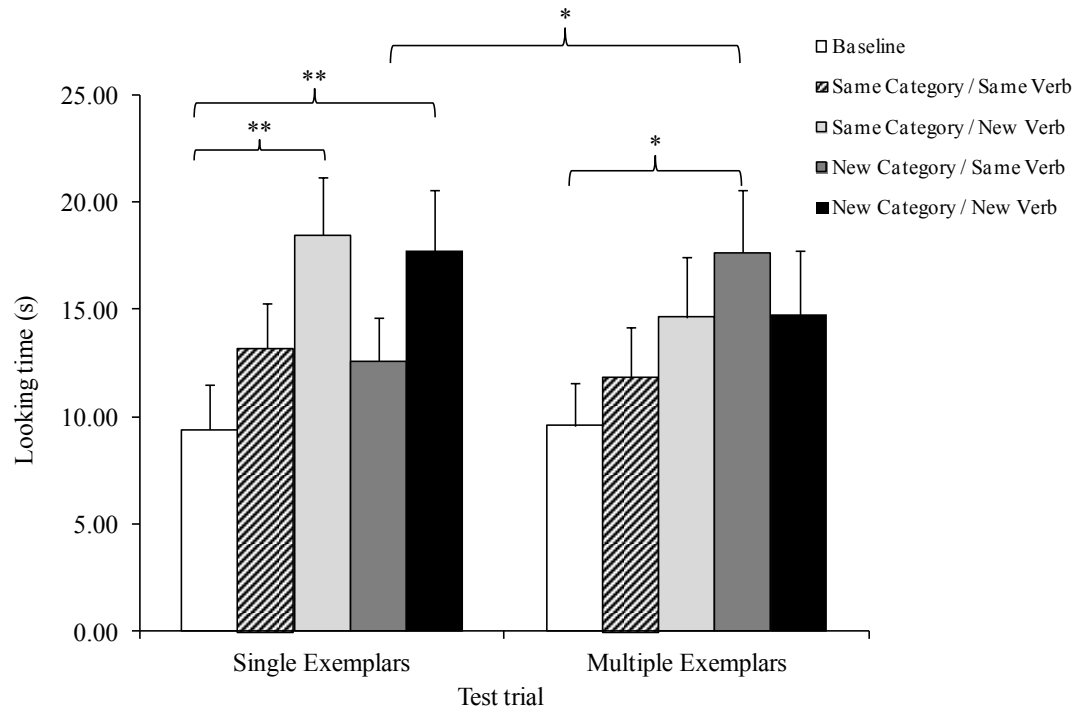


Figure 3. Looking times (seconds) to baseline and test trials. Error bars represent one standard error. ** $p < .01$, * $p < .05$.

The ANOVA also revealed a significant Verb x Condition interaction ($F(1, 34) = 8.07, p < .01, \eta_p^2 = .19$). Two follow-up ANOVAs with Test Trial (SCSV, SCNV, NCSV, NCNV) as the repeated measure were conducted on data from each condition separately. The ANOVAs indicated that the interaction was driven by infants in the *single* condition ($F(3, 51) = 3.24, p < .05, \eta_p^2 = .16$). Planned comparisons confirmed that infants looked longer during NCNV trials than during NCSV trials ($p < .05$): that is, infants in the *single* condition treated the *new-category* stimulus paired with the *new-verb* stimulus as novel relative to the *new-category* stimulus paired with the *same-verb* stimulus. Put differently, infants in the *single* condition learned that “any movement made by the star is labelled *dacking* (and not *keefing*)”. Children in the *multiple* condition looked equivalently at all trial types, $F(3, 51) = 1.92, ns$.

Verb discrimination. Discrimination is indicated by an increase in looking relative to the baseline trial. We therefore compared looking during each test trial to looking during the baseline trial for each condition.

There was no difference in looking times during the baseline trial between the *single* condition ($M = 9.39s$, $SD = 7.23$, range = 1.10s - 29.80s) and the *multiple* condition ($M = 9.60s$, $SD = 7.69$, range = 1.20s – 29.90s), $t(34) = -0.09$, $ns.$, $d = 0.03$). However, planned, paired t -tests comparing looking to each test trial type against the baseline trial for each condition did reveal differences.

In the *single* condition, two comparisons were significant: infants looked longer at SCNV trials than at the baseline ($t(17) = 3.75$, $p < .01$), and longer at NCNV trials than at the baseline ($t(17) = 3.27$, $p < .01$). Paired t -tests confirmed equivalent responding on the same-verb trials ($t(17) = 0.617$, $ns.$) and the new-verb trials ($t(17) = 0.132$, $ns.$). Infants in the *single* condition therefore learned the habituated verb, discriminating the completely novel verb, but did not discriminate changes to visual stimuli.

In the *multiple* condition, infants looked longer at the NCSV trials than at the baseline, $t(17) = 2.38$, $p < .05$. Infants in the *multiple* condition discriminated the novel *new-category* stimulus but did not discriminate any other stimulus. Unlike the children in the *single* condition, these children did not discriminate *new-verb* stimuli.

To compare looking times during the four test trial types between conditions, baseline looking was subtracted from looking during each test trial type for each condition. An independent samples t -test demonstrated that children looked for longer at NCSV trials relative to baseline in the *single* condition ($M = 6.28$ seconds, $SD = 6.29$) than in the *multiple* condition ($M = 13.60$ seconds, $SD = 8.60$), $t(34) = 2.91$, $p = < .01$, $d = 0.99$).

Effect of vocabulary. Based on existing categorisation and noun learning literature (e.g. Samuelson & Smith, 1999), we predicted that infants' prior experience with action categories would affect their verb categorisation. Using amount of looking during habituation as a measure of the ease (or otherwise) of learning the verb-category mapping, we found a strong overall looking/verb vocabulary correlation in the *multiple* condition, Pearson's $r = .53, p < .05$ (two tailed). For infants in the *single* condition, the correlation between verb vocabulary and looking during habituation disappeared (Pearson's $r = .113, ns.$) Thus, infants' experience with verbs and the action categories they label made a difference to looking times during habituation.

Discussion

Twenty-four-month-old infants were habituated to verb-action pairs, encountering either the same pair each time (*single* condition) or three different exemplars from the same action category (*multiple* condition). At test, infants in both conditions showed evidence of having learned the mapping between verb and action. However, encountering multiple versus single exemplars had a marked effect on infants' ability to discriminate new exemplars from the habituated category.

Labels drive discrimination following single exemplars. At test, infants in the *single* condition discriminated test exemplars which included a novel verb. Correspondingly, they did not discriminate test exemplars which included the habituated verb. Thus infants responded equivalently to the two test exemplars that shared the habituated verb, and to the two exemplars that shared the novel verb. One possibility for this behaviour is that in the *single* condition the change in the auditory stimulus was more salient than the change in the visual stimulus.

In Experiment 2 infants were habituated with a single action-verb pairing, as in the *single* condition of Experiment 3. However, these infants discriminated a *same-*

category/same verb exemplar relative to the baseline trial. Why do they not do so in Experiment 3? First, the *single* condition of Experiment 3 differs in a number of ways from Experiment 2. In Experiment 3, children were habituated to a maximum of 18 trials, rather than familiarised with exactly 8 test trials, giving more time for encoding if required, and allowing faster transition to the test phase if not. Further, in Experiment 3 infants encountered four test trials, allowing for comparison between stimuli across trials (an ability present in children as young as 6 months; Oakes & Ribar, 2005). Finally, in Experiment 3 infants additionally encountered a completely novel verb during test trials, introducing between-stimulus contrasts absent from Experiment 2.

Interestingly, recent evidence points to a dynamic interplay between categorisation and noun labelling. For example, 10-month-old infants formed a single category in the presence of a shared label after familiarisation with exemplars that, in the absence of a label, were perceived as forming two distinct categories (Plunkett, Hu, & Cohen, 2008). In Experiment 3, infants encounter a shared *same-verb* label across SCSV and NCSV trials, and a shared *new-verb* label across SCNV and NCNV trials. From this perspective, then, these data raise the possibility that that the shared label prompted infants in the *single* condition encode the *same-category* and *new-category* stimuli as exemplars from the same category. However, the current data do not unequivocally support this interpretation, and future research is required to better understand the interaction between labelling and verb categorisation.

Multiple exemplars facilitate verb categorisation. Overall, infants learned a novel verb category via habituation with both single and multiple exemplars. However, infants who encountered multiple exemplars only discriminated exemplars which included a new action category paired with the habituated verb. It is possible that *lack* of discrimination here indicates verb generalisation. If so, experience with multiple

exemplars prompted formation of a broad category which infants generalised to all stimuli except. This account is in line with existing research suggesting that extra perceptual variability in the visual modality triggers formation of inclusive categories (Quinn, 2005; Twomey & Horst, Paper 1, this thesis). Equally, however, it is possible that other factors, for example fatigue or lack of experience with action categories, could underlie these results (although recovery of looking times to the post-test stimulus relative to the baseline suggest that decreases in looking across trials were not due to lack of attention). Nonetheless, the current study demonstrates that just as in noun learning, exemplar variability affects verb categorisation.

General Discussion

These data provide evidence that verb learning is subject to the same influences as noun learning and, more generally that word learning is subject to the same domain-general processes (perceptual variability, vocabulary) as those that underlie with other forms of learning (e.g., Gogate & Hollich, 2010; Thelen & Smith, 1996). As such, the data support the dynamic systems account of cognitive development, in which behaviour of both infants and adults has been demonstrated to emerge online from nested timescales and the dynamic interaction between cognition, body and environment (Smith & Thelen, 2003). In Experiment 1, we add weight to existing evidence of a developmental change in verb semantics, and contribute evidence to the dynamic-associative account of such these patterns of behaviour emerge. In Experiment 2, we demonstrate that 24-month-old infants can rapidly and robustly form verb categories, over a restricted number of temporally co-occurring visual and auditory stimuli (Gogate, Walker-Andrews, & Bahrick, 2001). In Experiment 3, we demonstrate that infants' verb categories emerge online from the multiple timescales of long-term learning history (here, vocabulary size), in-task experience (single vs. multiple

exemplars) and in-the-moment environmental context (visual and auditory differences in test stimuli).

Taken together, these studies indicate that infants' verb learning is not "special". In fact, verb learning and action categorisation may be subject to the same general cognitive processes that underlie noun learning and object categorisation. The current studies therefore inform our understanding of verb learning, and as well as contribute to accounts of language as underpinned by flexible, domain-general, and dynamically interactive cognitive processes.

Acknowledgements

This research was supported by ESRC Quota Award ES/H014942/1 to K.E.T. We would like to thank Lauren Lush and Ruth Pearce for help in data collection, subject recruitment and coding. We would also like to thank the caregivers and children who participated.

References

- Arterberry, M. E., & Bornstein, M. H. (2001). Three-month-old infants' categorization of animals and vehicles based on static and dynamic attributes. *Journal of Experimental Child Psychology*, 80(4), 333-346. DOI: 10.1006/jecp.2001.2637
- Barner, D., Inagaki, S., & Li, P. (2009). Language, thought, and real nouns. *Cognition*, 111(3), 329-344. DOI: 10.1016/j.cognition.2009.02.008
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M-G., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75(4), 1115–1139. DOI:10.1111/j.1467-8624.2004.00729.x
- Bornstein, M. H., & Mash, C. (2010). Experience-based and on-line categorisation of objects in early infancy. *Child Development*, 81(3), 884-897. DOI: 10.1111/j.1467-8624.2010.01440.x
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Chan, C. C. Y., Tardif, T., Chen, J., Pulverman, R. B., Zhu, L. Q., & Meng, X. Z. (2011). English- and Chinese-learning infants map novel labels to objects and actions differently. *Developmental Psychology*, 47(5), 1459-1471. DOI: 10.1037/a0024049
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967-978. DOI: 10.1037/0012-1649.38.6.967
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of Child Language*, 22(3), 497-529. DOI: 10.1017/S0305000900009934

- Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). *Habit 2000: A new program for testing infant perception and cognition (Version 1.0)*. Austin: University of Texas.
- Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: the self-organisation of skilled noun learning. *Developmental Science*, 11(2), 195-203. DOI: 10.1111/j.1467-7687.2007.00665.x
- Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorise objects. *Developmental Psychology*, 42(6), 1000-1011. DOI: 10.1037/0012-1649.42.6.1000
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular Publishing Group.
- Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory*, 4(2), 161-178.
- Gershkoff-Stowe, L. (2002). Object naming, vocabulary growth, and the development of word retrieval abilities. *Journal of Memory and Language*, 46(4), 665-687. DOI: 10.1006/jmla.2001.2830
- Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech Language and Hearing Research*, 50(3), 682-697. DOI: 10.1044/1092-4388(2007/048)
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098-1114. DOI: 10.1111/j.1467-8624.2004.00728.x
- Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117(2), 496-516. DOI: 10.1037/a0019049

- Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: an ecological-dynamic systems view. *Developmental Science*, 4(1), 1-18. DOI: 10.1111/1467-7687.00143
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23-B34. DOI: 10.1016/S0010-0277(02)00186-5
- Haryu, E., Imai, M., & Okada, H. (2011). Object similarity bootstraps young children to action-based verb extension. *Child Development*, 82(2), 674-686. DOI: 10.1111/j.1467-8624.2010.01567.x
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., et al. (2000). *Breaking the language barrier: An emergentist coalition model for the origins of word learning*. Monographs of the Society for Research in Child Development.
- Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infants categorise. *Child Development*, 76(3), 614-631. DOI: 10.1111/j.1467-8624.2005.00867.x
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI: 10.1080/15250000701795598
- Horst, J. S., Scott, E. J., & Pollard, J. P. (2010). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706-713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Horst, J. S., & Twomey, K. E. (2012). It's taking shape: Shared object features influence novel noun generalizations. *Infant and Child Development*, DOI:10.1002/icd.1768
- Imai, M., Haryu, E., & Okada, H. (2005). Mapping novel nouns and verbs onto dynamic action events: Are verb meanings easier to learn than noun meanings for

- Japanese children? *Child Development*, 76(2), 340-355. DOI: 10.1111/j.1467-8624.2005.00849_a.x.
- Jaswal, V. K. (2007). The effect of vocabulary size on toddlers' receptiveness to unexpected testimony about category membership. *Infancy*, 12(2), 169-187. DOI: 10.1111/j.1532-7078.2007.tb00239.x
- Jones, S. S. (2003). Late talkers show no shape bias in a novel name extension task. *Developmental Science*, 6(5), 477-483. DOI: 10.1111/1467-7687.00304
- Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development*, 73(1), 93-109. DOI: 10.1111/1467-8624.00394
- Klee, T., & Harrison, C. P. (2001). *CDI Words and Sentences: validity and preliminary norms for British English*. Paper presented at the Child Language Seminar, University of Hertfordshire.
- Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behavior on infant categorisation. *Infancy*, 13(4), 285-307. DOI: 10.1080/15250000802189428
- Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorisation. *Journal of Experimental Child Psychology*, 98(2), 69-93. DOI: 10.1016/j.jecp.2007.05.001. DOI: 10.1016/j.jecp.2007.05.001
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. DOI: 10.1016/0885-2014(88)90014-7
- Lipinski, J., Simmering, V. R., Johnson, J. S., & Spencer, J. P. (2010). The role of experience in location estimation: Target distributions shift location memory biases. *Cognition*, 115(1), 147-153. DOI: 10.1016/j.cognition.2009.12.008

- Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). *A Unified Theory of Word Learning: Putting Verb Acquisition In Context*. In K Hirsh-Pasek & R. M Golinkoff (Eds.), (2006) Action meets word: How children learn verbs. (pp. 364-391).
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628-634. DOI 10.1111/j.1467-7687.2008.00707.x
- Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., Imai, M., Haryu, E., Vanegas, S., Okada, H., Pulverman, R., & Sanchez-Davis, B. (2010). A developmental shift from similar to language-specific strategies in verb acquisition: A comparison of English, Spanish, and Japanese. *Cognition*, *114*(3), 299-319. DOI: 10.1016/j.cognition.2009.10.002
- Markman, E. M. (1987). *How children constrain the possible meanings of words*. In Neisser, Ulric (Ed), (1987). Concepts and conceptual development: Ecological and intellectual factors in categorisation (pp. 255-287). NY, US: Cambridge University Press.
- Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Developmental Science*, *11*(2), 204-208. DOI: 10.1111/j.1467-7687.2007.00666.x
- Mayor, J., & Plunkett, K. (2010). A Neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1-31. DOI: 10.1037/a0018130

- McDonough, C., Song, L., Hirsh-Pasek, K., Golinkoff, R. M., & Lannon, R. (2011). An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science*, 14(2), 181-189. PMid:21359165
- Mintz, T. H., & Gleitman, L. R. (2002). Adjectives really do modify nouns: The incremental and restricted nature of early adjective acquisition. *Cognition*, 84(3), 267-293. DOI: 10.1016/S0010-0277(02)00047-1
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3. PMid:22291676
- Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3), 255-268. DOI: 10.1080/15248371003699977
- Oakes, L. M., Coppage, D. J., & Dingel, A. (1997). By land or by sea: The role of perceptual similarity in infants' categorisation of animals. *Developmental Psychology*, 33(3), 396-407. DOI: 10.1037/0012-1649.33.3.396
- Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorisation in paired and successive presentation familiarisation tasks. *Infancy*, 7(1), 85-98. DOI: 10.1207/s15327078in0701_7
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalisation and word learning. *Psychological Science*, 21(12), 1894-1902. DOI: 10.1177/0956797610389189
- Pitchford, N. J., & Mullen, K. T. (2001). Conceptualisation of perceptual attributes: A special case for colour? *Journal of Experimental Child Psychology*, 80(3), 289-314. DOI: 10.1006/jecp.2001.2634

- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665-681. DOI: 10.1016/j.cognition.2007.04.003
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, 27(2), 145-179. DOI: 10.1016/0010-0277(87)90017-5
- Quinn, P. C. (2002). Category representation in young infants. *Current Directions in Psychological Science*, 11(2), 66-70. DOI: 10.1111/1467-8721.00170
- Quinn, P. C. (2005). *Young Infants' Categorisation of Human Versus Nonhuman Animals: Roles for Knowledge Access and Perceptual Process*. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 107-130). Mahwah, NJ: Lawrence Erlbaum Associates.
- Quinn, P. C., & Bhatt, R. S. (2009). Transfer and scaffolding of perceptual grouping occurs across organizing principles in 3-to 7-month-old infants. *Psychological Science*, 20(8), 933-938. DOI: 10.1111/j.1467-9280.2009.02383.x
- Quinn, P. C., & Bhatt, R. S. (2010). Learning perceptual organisation in infancy: The effect of simultaneous versus sequential variability experience. *Perception*, 39(6), 795-806. DOI: 10.1068/P6639
- Quinn, P. C., Bhatt, R. S., & Hayden, A. (2008). What goes with what? Development of perceptual grouping in infancy. *Psychology of Learning and Motivation: Advances in Research and Theory*, 49, 105-146. DOI:10.1016/So079-7421(08)00003-0
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old Infants. *Perception*, 22(4), 463-475. DOI: 10.1068/p220463

- Rakison, D. H. (2005). Developing knowledge of objects' motion properties in infancy. *Cognition*, 96(3), 183-214. DOI 10.1016/j.cognition.2004.07.007
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339-349. DOI: 10.1111/j.1467-7687.2008.00786.x
- Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484-494. DOI: 10.1037/0012-1649.39.3.484
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016-1037. DOI: 10.1037//0012-1649.38.6.1016
- Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalisation: Moment-to-moment interactions shape children's naming biases. *Infancy*, 11(1), 97-110. DOI: 10.1207/s15327078in1101_5
- Samuelson, L. K., & Horst, J. S. (2008). Confronting complexity: insights from the details of behavior over multiple timescales. *Developmental Science*, 11(2), 209-215. DOI: 10.1111/j.1467-7687.2007.00667.x
- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1-33. DOI: 10.1016/S0010-0277(99)00034-7
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, 6(12), DOI: e28095.
- Sandhofer, C. M., & Smith, L. B. (1999). Learning colour words involves learning a system of mappings. *Developmental Psychology*, 35(3), 668-679. DOI: 10.1037/0012-1649.35.3.668

- Schutte, A. R., & Spencer, J. P. (2010). Filling the gap on developmental change: Tests of a dynamic field theory of spatial cognition. *Journal of Cognition and Development, 11*(3), 328-355. DOI: 10.1080/15248371003700007 Pii 925511148
- Seston, R., Golinkoff, R. M., Ma, W. Y., & Hirsh-Pasek, K. (2009). Vacuuming with my mouth? Children's ability to comprehend novel extensions of familiar verbs. *Cognitive Development, 24*(2), 113-124. DOI: 10.1016/j.cogdev.2008.12.001
- Slaughter, V., & Suddendorf, T. (2007). Participant loss due to "fussiness" in infant visual paradigms: A review of the last 20 years. *Infant Behavior & Development, 30*(3), 505-514. DOI: 10.1016/j.infbeh.2006.12.006
- Slobin, D. I. (2003). Language and thought online: *Cognitive Consequences of Linguistic Relativity*. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (Vol. 157, pp. 178). Massachusetts: Massachusetts Institute of Technology.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*(1), 13-19. DOI: 10.1111/1467-9280.00403
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Sciences, 7*(8), 343-348. DOI: 10.1016/S1364-6613(03)00156-6
- Stephen, D. G., & Dixon, J. A. (2009). The self-organisation of insight: Entropy and power laws in problem solving. *The Journal of Problem Solving, 2*(1), 6.
- Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language, 24*(3), 535-565. DOI: 10.1017/S030500099700319X

- Thelen, E., & Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.
- Thom, E. E., & Sandhofer, C. M. (2009). More is more: The relationship between vocabulary size and word extension. *Journal of Experimental Child Psychology*, 104(4), 466-473. DOI: 10.1016/j.jecp.2009.07.004
- Twomey, K. E., & Horst, J. S. (Paper 1, this thesis). That's more like it: More exemplars facilitate word learning.
- Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, 43(3), 217-242. DOI: 10.1006/cogp.2001.0764
- Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition*, 95(3), B59-B68. DOI: 10.1016/j.cognition.2004.09.003
- Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, 34(6), 1289-1309. DOI: 10.1037/0012-1649.34.6.1289
- Woodward, A. L., & Markman, E. M. (1998). Early word learning. Damon, William (Ed), (1998). *Handbook of Child Psychology: Volume 2: Cognition and Language*. Hoboken, New Jersey: John Wiley & Sons.
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47(5), 1220-1229. DOI: 10.1037/a0024023

Yuan, S., & Fisher, C. (2009). "Really? She Blicked the Baby?": Two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20(5), 619-626. DOI: 10.1111/j.1467-9280.2009.02341.x

Running head: BAYESIAN VERSUS EMERGENTIST MODELS

Two Households, Both Alike in Dignity:

Bayesian Versus Emergentist Models of Word Learning and Categorisation

Katherine E. Twomey

University of Sussex

Author note

A version of this paper is in preparation, to be submitted as follows:

Twomey, K. E. (in prep.). A View From Above or Bottoms Up? Bayesian Versus Emergentist Models of Word Learning and Categorisation. *Cognitive Systems Research*.

Abstract

Although computational models offer exciting new insights into cognitive development, at present the field is divided in terms of theoretical accounts of cognition. Some argue that it can be understood as optimal inference under uncertainty and modelled using top-down hierarchical structures – the Bayesian approach. Others contend that cognition emerges from interaction over nested timescales between the brain, the body, and the environment – the emergentist approach. This paper attempts to reconcile the two accounts. First, it situates each approach in its historical and theoretical context, then compares two models of the same word learning phenomenon: the shape bias. Rather than reject one approach in favour of another, this review concludes that although emergentist models currently offer the most readily testable predictions, and therefore deeper insight into cognitive development, an integration of environmentally-grounded Bayesian research with existing emergentist models would be of great benefit not only to research in word learning, but also to our understanding of the unfolding of cognition across the lifespan.

Two Households, Both Alike in Dignity:

Probabilistic Versus Emergentist Models of Cognitive Development

For decades, psychologists have painstakingly explored behavioural and developmental phenomena, using techniques ranging from the psychophysical to neuroimaging (e.g., Aslin & Salapatek, 1975; Cohen & Strauss, 1979; Fava, Hull & Bortfield, 2011; Oakes, Madole, & Cohen, 1991). In recent decades, however, computational models have successfully made the conceptual leap from domains such as neurophysiology (McCulloch & Pitts, 1943), signal processing (Green & Swets, 1966) and thermodynamics (Kelso, Ding, & Schöner, 1993) to join the battery of investigative tools available to psychology providing exciting new insight into the hidden computations underlying human behaviour. Computational models allow the researcher to examine change in a learner's or neural system's behaviour given a certain input, often over time (Elman, 1990), offering an unprecedented ability to pilot new experimental designs and facilitating quick and inexpensive testing of new theories.

However, progress in the field is hampered by a theoretical schism, perhaps best illustrated by the hotly-debated differences between two of the most common approaches to modelling cognitive development: probabilistic, Bayesian models (henceforth “Bayesian models”) and emergent, neural network models (henceforth “emergentist models”). This review explores the theoretical and empirical context of each type of model, presents a comparison of a Bayesian and an emergentist model of a well-known word learning phenomenon in infants – the “shape bias” (e.g., Landau, Smith, & Jones, 1988) – and finally evaluates the contribution of each approach to our understanding of cognition. The evidence presented here suggests that current neurologically-plausible emergentist models offer the most readily-testable predictions, and therefore, at present, contribute the most valuable insights to the field. However,

taken as a whole, the modelling literature points to the benefits of a synergistic approach in which Bayesian and emergentist models are designed with the common goal of paying serious attention to the learner's environment, as well as to the emergence of cognitive structure over time.

The modeller's challenge. Computational models consist of mathematical algorithms that generate an output in response to an input. Inputs to models of human cognition are schematised mathematical formalisms of some relevant aspect of a human learner's environment. Cognitive modellers aim to write algorithms that process input (e.g., visual images, auditory signals) and generate output (e.g., language, categories, motor behaviour), which reflect the inputs and outputs to real-world cognition. For example, inputs to the seminal Rumelhart & McClelland (1985) connectionist model of the acquisition of the English past tense consist simply of strings of phonemes representing English verbs, encoded as patterns of activations across a layer of neurons (for terminology see section: *Emergentist models: Mathematical implementation and structure*). The model processes these inputs to generate an output consisting of another string of phonemes instantiated in a layer of output nodes. Given inputs reflecting the structure of children's early-learned verbs, this model reproduces the patterns of verb past tense overgeneralizations demonstrated by children learning English. Thus, the algorithms governing these processes may reflect some cognitive substrate of children's past tense acquisition.

The first and critical step in modelling is to select a target cognitive phenomenon, bearing in mind that the target data limit the situations to which a model's findings can be generalised. For instance, a model examining infants' looking times to novel stimuli is unlikely to inform the modeller about the amount of time a school-age child will spend looking at printed words on the page of a story book. Equally, a model of adult

decision making will be uninformative about babies' behaviour in habituation experiments. Thus, Hahn & Oaksford's (2007) Bayesian model of adults' ratings of reasoning fallacies must remain agnostic as to low-level perceptual learning in infants. Similarly, Gliozi, Mayor, Hu & Plunkett's (2009) model of infant word learning and object categorisation will not help the researcher understand adults' acceptability judgements of different reasoning strategies. Importantly, then, choosing an approach requires commitment to strong theoretical assumptions about both representational structure and the learner's environment.

Similarly, choice of model type also affects how different models inform our understanding of cognitive development. If a model includes preprogrammed "rules", then the implication is that the cognitive process under examination also depends on some kind of prior knowledge (e.g., Gopnik & Tenenbaum, 2007). However, if a model learns from input alone, this implies that the cognitive process under examination is also learned purely from input (e.g., Elman, 1993). Further, some models may be semi-structured, incorporating assumptions based on existing knowledge from other domains, such as neuroscience (e.g., Chang, Dell & Bock, 2006). The implications of such theoretical and implementational assumptions for both Bayesian and emergentist models are discussed in the following paragraphs.

Bayesian Models

Theoretical background

Bayesian models are inference calculators, computing the probability of all possible hypotheses being true given a particular task environment and observed data. Fundamental to this approach is a commitment to Marr's (1982) "computational" level of analysis. According to Marr, cognitive processes can be described at three levels. At the top of Marr's hierarchy, the "computational" level concerns the goal of the cognitive

process (e.g., picking up an object), and the way in which that goal might be achieved (e.g., a reach and a grasp). Pre-specified constraints operating on the possible outputs of the process (e.g., which objects are within reach) are also defined at the computational level. Moving down a step, the “representation and algorithm” level involves the selection of input/output representations and appropriate rules (i.e., algorithms) to achieve the desired outcome. For example, the representation level might include rules such as “if an OBJECT is CURVED, hold fingers in curled position”, “If an OBJECT is THIN, hold fingers in pincer position”, operating over representational symbols such as OBJECT, CURVED, and THIN). Finally, the lowest, “implementation” level describes the physical realisation of the components required to achieve the predefined goal (e.g., the neural response generating a motor action).

The problem space in Bayesian models is therefore defined top-down, from the highest, computational level. Importantly, lower-level processes are not addressed. For example, an infant chooses to reach towards a one toy rather than another in response to “Where’s your bear?” From a Bayesian perspective, this behaviour results from the probability of *bear* referring to the reached-to toy being greater than the probability of *bear* referring to the other toy. Thus, infants’ behaviour is explained by inference-making at the computational level (e.g., “reach for the toy that looks most like the other items called *bear*”). The implementational and representational levels (e.g., the way in which visual input is processed and compared to existing representations of the BALL category) are not considered. Perfors, Xu, Griffiths and Tenenbaum (2011) explicitly situate Bayesian models at the computational level, as do numerous other influential Bayesians (e.g., Chater, Oaksford, Nakisa, & Redington, 2003; Gopnik, et al., 2004; Gopnik & Tenenbaum, 2007; Kemp, Perfors, & Tenenbaum, 2007; Lee & Sarnecka, 2010; Shultz, 2007; Xu & Tenenbaum, 2007). Importantly, adhering to the

computational level means that all possible outputs in a Bayesian model are determined *a priori* by the modeller (see *Mathematical Implementation and Structure*, this section).

Implicit in a commitment to the computational level is a simultaneous commitment to the “information processing” view of cognition (e.g., Klahr & Wallace, 1976; Lachman & Butterfield, 1979). By this account, mental representations of real-world entities consist of schematic symbols manipulated by *a priori* rules (for example, the “words and rules” account of language development, Berent & Pinker, 2008; Huang & Pinker, 2010; Pinker & Prince, 1988). Understanding cognition, therefore, consists of understanding the computational-level rules that generate observed behaviour, rather than the underlying neurophysiology. In this way, Bayesian modellers openly eschew the study of the brain structure and function (Daw, Courville, & Dayan, 2008). Consequently, Bayesian approaches have been described as the “new Behaviorism” (e.g., Jones & Love, 2011).

Just as information processing has contributed its “rules” to Bayesian approaches, early models of inference under uncertainty have contributed the notion of probabilistic reasoning (Duda, Hart, & Nilsson, 1976; Rousseau, 1968; Weiss, Kulikowski, Amarel, & Safir, 1978). Recent research suggests that cognitive development depends at least in part on learning the probabilistic (or statistical) structure of the perceptual and social environment (Gopnik & Tenenbaum, 2007; Oaksford & Chater, 2003; Yu, Smith, Klein, & Shiffrin, 2007). Instead of “symbols”, then, modern Bayesian models define multiple hypotheses about the learner’s environment (the “hypothesis space”), assign a probability to each hypothesis, and continually update these probabilities on encountering new data.

Mathematical implementation and structure

As noted, Bayesian researchers do not generally address brain structure. Instead, Bayesian models compute the probability of behavioural outcomes by calculating the probability of a hypothesis being true given data collected (Bayes' rule; see below for more detail).

Bayesian models consist of a structured “hypothesis space” (a set of mutually-exclusive hypotheses), a formal definition of the relationship between each hypothesis and the observable data (the “likelihood”), and the set of prior probabilities of each hypothesis in the hypothesis space (or “priors”; Chater, Tenenbaum & Yuille, 2006; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010). The model is then presented with data in vector form and an adapted version of Bayes' rule is used to calculate for each hypothesis the posterior probability given the data, $P(h_i|d)$, as follows:

$$P(h_i|d) = \frac{P(d|h_i)}{\sum_{h_j \in \mathcal{H}} P(d|h_j)P(h_j)}$$

where $P(d|h_i)$ is the likelihood of observing the data given the hypothesis i (“likelihood function”), $P(h_i)$ is the likelihood of prior hypotheses j (“priors”), and

$\sum_{h_j \in \mathcal{H}} P(d|h_j)P(h_j)$ represents the sum over all hypotheses of the product of the priors and likelihoods. Bayes' rule therefore attributes to the hypothesis in question a proportion of the total probability in the model, taking the data observed into account (Perfors, et al., 2011; see also Dienes, 2011). Once the posterior probabilities have been updated in response to the data, the hypothesis with the maximum posterior probability is taken as the model's “choice” or “inference”. In this way, Bayesian models examine (probabilistically) optimal behaviour in a given task environment.

Recent Bayesian models focus on the complexity of the learner's environment by employing complex hypothesis spaces – that is, hypothesis spaces which reflect all possible hypotheses in a given task environment, rather than a smaller number of modeller-selected outcomes. However, increased complexity brings with it problems of combinatorial explosion (Sanborn, Griffiths, & Navarro, 2010). For example, evidence suggests that young infants' categories can be based on clusters of co-occurring features (Baumgartner & Oakes, 2011; Plunkett, Hu, & Cohen, 2008; Younger & Cohen, 1983). Mathematically, the number of categories an object could belong to (however improbable that membership may be) is determined by the number of possible feature clusters. For example, the features FEATHERS and BEAK combine to make three possible categories: FEATHERS, BEAK, and FEATHERS+BEAK. Adding the single feature FLIES adds four new possible clusters, up to a total of seven. Thus, the number of possible categories (and therefore the number of hypotheses) increases exponentially as new features are added. Inevitably, therefore, the scope of Bayesian models is restricted by computational capacity (Perfors, et al., 2011).

To address this, recent models have employed Monte Carlo methods, in which computations are performed using a limited number of samples from a probability distribution, rather than the distribution in its entirety. In Bayesian models, this entails using random samples from the likelihood functions to generate posterior probabilities (Chater, Tenenbaum & Yuille, 2006b; Sanborn, et al., 2010). The consequences of these methods are double-edged. On the one hand, Monte Carlo-based models naturally capture environmental stochasticity. On the other, the algorithms used in Monte Carlo estimation came from nuclear physics (Liu, 2008; von Neumann, Metropolis, & Ulam, 1951), leading some to question the extent to which such models reflect – and therefore

inform our understanding of – human cognition (Jones & Love, 2011; but see also Chater, et al., 2011).

Assumptions

Bayesian models are highly structured and require the modeller to make several decisions. First, the modeller decides *a priori* which hypotheses are relevant to the model and how these hypotheses relate to one another. Thus, the hypothesis space is structurally constrained – and whether these constraints are innate or learned is often left unaddressed (Chater, Tenenbaum & Yuille, 2006a). This top-down structure means that the vast majority of Bayesian models generally cannot generate completely novel predictions (but see Tenenbaum & Griffiths, 2001). Instead, Bayesian models express novelty in terms of choosing a never-before chosen hypothesis (although all hypotheses are modeller-defined).

Second, the modeller hand-picks the likelihood function; that is, the shape of the probability distribution over each possible hypothesis is therefore also chosen in advance. Similar to the hypothesis space, how the modeller defines these likelihoods constrains what the model can do. Likelihood functions are usually based on known probability distributions (e.g., normal, Dirichlet, e.g., Xu & Tenenbaum, 2007).

Finally, the prior probabilities (“priors”) of each hypothesis are also determined by the modeller. Again, the choice of priors fundamentally affects the way in which a model instantiates “cognition”. For example, the careful design of priors to reflect statistical regularities in the environment can contribute to an in-depth and informative model of the learner’s perceptual environment and learning history (“Enlightened Bayes”, Jones & Love, 2011; see also Anderson, 1991). Alternatively, priors can be viewed as the learner’s cognitive environment or personal probability of a hypothesis being true; this type of prior is often based on existing empirical evidence (e.g.,

Oaksford, Chater, & Larkin, 2000). Finally, priors can be deliberately unstructured (that is, non-informative; for a discussion, see Myung & Pitt, 1997), although the extent to which such models reflect cognition and learning history is unclear (Jones & Love, 2011).

The high degree of modeller control and large number of free parameters lends Bayesian approaches great flexibility. Consequently, psychology alone has seen Bayesian models of cognitive structures as diverse as conditional inference spaces (Oaksford & Chater, 2003), causal relations (Gopnik, et al., 2004), hierarchical syntax in language and music (Dawson & Gerken, 2011; Eisner, 2002), and clusters of perceptual features in object categories (Xu & Tenenbaum, 2007) as well as models of complex hierarchical cognitive structure such as compositionality in syntax (Perfors, et al., 2011) – an ability currently outside the scope of connectionist models.

However, flexibility comes at a philosophical and practical price. The emergentist camp have forcibly argued that Bayesian models risk design (or designer) bias (McClelland et al., 2011) due to the degree of influence the modeller has on the structures the model can represent. Further, the abundance of degrees of freedom in Bayesian models allows the modeller to accurately fit almost any dataset obtained in almost any task environment, giving rise to Popperian problems of falsifiability (Bowers & Davis, 2012). Correspondingly, in providing a perfect fit to a given dataset, some Bayesian models suffer from difficulties stemming from overfitting; that is, a model that very accurately reflects a single dataset will not generalise well to another (Gurney, 1997). Compounded by the fact that the hypotheses are mutually exclusive (in order to obey the axioms of probability, that is, total probability in the model cannot exceed 1), some have argued that Bayesian models simply reframe datasets into the

language of probability and reveal little about cognition in any other context than the task environment specific to that model (e.g., Glymour, 2011).

Indeed, Bayesian modellers themselves acknowledge the problem of lack of generalisability, and have tried to overcome it. For example, Tenenbaum & Griffiths (2001) proposed a model in which the hypothesis space is associated to overlapping subregions of notional “psychological space”. This permits a one-to-many relationship between a single subregion and several hypotheses, such that several different psychological computations could account for a single data set, and importantly, generating testable predictions. However, this model also assumes the existence of “natural kinds” of real-world entity, a theory that is itself the source of impassioned debate (Booth, 2009; Colunga & Smith, 2008b; Gelman, 2003; Madole & Oakes, 1999; Mandler & McDonough, 1998; McCarthy, 2008; Rakison, 2007), thereby merely shifting, and not resolving, the debate.

Clearly, the Bayesian research program offers great flexibility, freeing the modeller from the restrictions of faithfully representing neuronal structure and thereby facilitating research into the complex hierarchical structures that permeate our world (Chater, et al., 2011). However, Bayesian detractors argue that probabilistic models are underconstrained, and that the resulting literature is incoherent (McClelland et al., 2011).

Emergentist models

Theoretical background. In contrast to the Bayesian emphasis on probabilistic inference, emergentist models of cognition simulate the function of the human brain. From this perspective, a focus on low-level neural processing is necessary, and, it is argued, sufficient to understand cognition (Crick, 1989; McClelland, et al., 2011).

The emergentist tradition began with neural network models, in which neural processing is simulated using a network of mathematically modelled neurons

(McCulloch & Pitts, 1943). From the outset, influential researchers such as Hebb emphasised the importance of neurophysiology to psychological theory, rejecting the behaviourist “black box” view of cognition (Hebb, 1949/2002; McLeod, Plunkett, & Rolls, 1998). Emergentist approaches thus explicitly eschew the rule- and structure-based computations of Bayesian models, instead favouring flexible, emergent representations distributed over a network of units. Nonetheless, the emergentist program began with probabilistic representations in mind: Rosenblatt, whose perceptron model was one of the earliest and most influential contributions to the field, stated: “The need for a suitable language for the mathematical analysis of events in systems where only the gross organisation can be characterised, and the precise structure is unknown, has led the author to formulate the current model in terms of probability theory” (1958, pp. 387-8). Thus, both emergentist and Bayesian approaches have probability theory at their core – a striking historical similarity between two now distinct paradigms.

Emergentist models proliferated following the paradigm’s resurgence in the 1980s catalysed by the Parallel Distributed Processing group (e.g., Rumelhart & McClelland, 1986). The field made further progress with the development of Simple Recurrent Networks (SRNs, that is, networks with layer(s) of hidden units that serve as memory, such that output at time t serves as input at time $t + 1$; Elman, 1990; Jordan, 1986). These recurrent emergentist models are members of the dynamical system family (Kelso, et al., 1993): complex physical systems of reciprocally coupled, continually interacting components. Such systems may exhibit both stable, predictable behaviour and complex, difficult-to-predict behaviour (for example the behaviour of a weather system). This behaviour can be modelled using mathematical functions; astonishingly, these functions describe the behaviour of an incredibly diverse range of physical system, from the stable, predictable behaviour of an oscillating pendulum, to

the complex, difficult-to-predict behaviour of a weather system. The mathematical functions which describe dynamical systems may converge to “attractor states”, which describe a system’s stable behaviours. Alternatively, functions may be at an “instability”, describing a system’s chaotic behaviours (Kelso et al., 1993).

Neural networks, being complex systems of reciprocally coupled, continually interacting artificial neurons, also exhibit such behaviour. In cognitive science, therefore, attractor states are taken to represent stable cognitive structures, for example categories (Westermann & Mareschal, 2009), motor plans (e.g; Thelen & Spencer, 1998) or location memory (Schutte, Spencer & Schöner, 2003). Because every component of a dynamical system is coupled to every other component, Marr’s (1982) levels of analysis are extraneous to the emergentist approach. In neural networks, stable, lasting structure emerges from interactions between neurons as governed by the strength of the connections which link them: representation, computation and implementation are therefore one and the same (McLeod, et al., 1998).

Outside psychology, neural network models have been implemented in countless domains requiring pattern recognition and prediction, for example finance (e.g., Brabazon & O'Neill, 2008), the military (e.g., H. Chen, Wu, Wang, Lin, & Cai, 2011), demographics (e.g., Dombi, Rosbolt, & Severson, 2010), computer vision (e.g., C. Chen, 2010), and robotics (e.g., Berthouze & Metta, 2005; Twomey, Horst & Morse, paper 6, this thesis). Within psychology, emergentist models have successfully modelled various cognitive processes including aspects of development such as categorisation (e.g., French, Mareschal, Mermillod, & Quinn, 2004), fast mapping (e.g., Mayor & Plunkett, 2010) and word learning (e.g., McMurray, Horst, & Samuelson, in press).

Mathematical implementation and structure

The most widely-researched branch of emergentism is the diverse family of connectionist models (Gurney, 1997; McLeod et al., 1998). The paradigm's commitment to faithful simulation of neurophysiological structure is evident in its terminology: connectionist models are “neural” networks of discrete processing units, in the form of idealised artificial “neurons”. Each neuron is connected to its neighbours by adaptable weights, via which neural “activity” spreads from neuron to neuron.

Arguably the most influential neuron model is the perceptron (Rosenblatt, 1958; see Fig. 1). Artificial neurons sum activation input from neighbouring units. The summed activation is processed by some function chosen by the modeller (commonly sigmoidal), and if the result is above a predetermined threshold, activation is output to neighbouring units. A randomly-initialised network of neurons can be trained to reflect a given input pattern, via a process of error reduction over iterative weight adaptation, for example, “backpropagation” networks (Rumelhart, Hintont, & Williams, 1986).

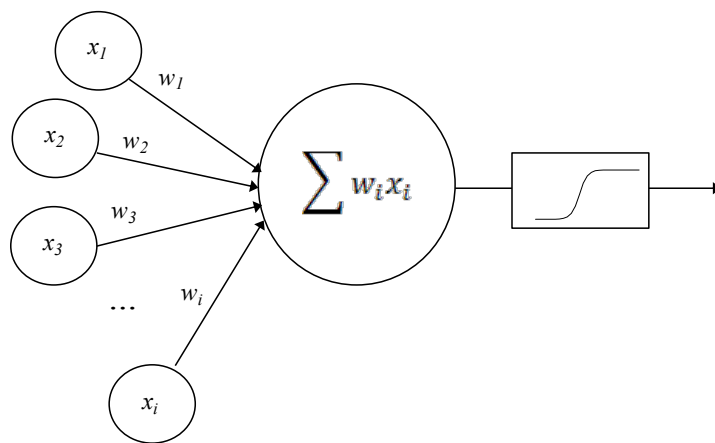


Figure 1. An artificial neuron

More recently, neurologically-based connectionist principles have been adapted to formally embody the dynamic systems emphasis on the emergence of structure over

time in Dynamic Field Theory (DFT; Kelso, et al., 1993; Thelen, Schöner, Scheier, & Smith, 2001). Where connectionist models simulate learning by increasing connection weights between discrete units, DFT employs spatially continuous fields of activation (“Dynamic Neural Fields”, or DNFs) which interact dynamically, feeding above-threshold activation to a separate memory field. Because DNFs are topologically meaningful, they have been successfully used to investigate spatial (amongst other) phenomena such as infants’ perseverative reaching (Schutte & Spencer, 2002), memory for object locations (Schutte & Spencer, 2009; Simmering & Spencer, 2008), and word-object binding via spatial location (Samuelson, Smith, Perry, & Spencer, 2011).

Assumptions

Just as in Bayesian models, the emergentist research program makes several assumptions about cognitive structure. First, and fundamentally, cognitive structure emerges from neuronal and perceptual interaction (Riley, Shockley, & Van Orden, 2012; Samuelson, et al., 2011). However, the structure of the simulated neurons and the nature of their interactions are designed *a priori* by the modeller. Network architectures thus vary greatly, from backpropagation networks in which supervised learning depends on an error-reduction algorithm (e.g., Schlesinger & Parisi, 2004), to self-organising maps which evolve unsupervised to reflect statistical regularities in the input (e.g., Mayor & Plunkett, 2010).

Despite evidence from neuroscience of localised neurological function (although the extent of localisation is disputed, e.g., Kolb & Whishaw, 2003; Molfese & Burger-Judisch, 1991), connectionist models often assume uniformity of neural structure and interactions, and simplify the neurotransmitter system into simple excitatory and inhibitory interactions (Cooper & Shallice, 2010). Some Bayesian modellers argue that the level of abstraction in these mathematical formalisations of neural function is so

great as to render negligible the informative value of emergentist models (Griffiths, et al., 2010; Norris, 2011). In response, however, emergentist modellers emphasise the importance of simplification to the discovery of the components necessary to complex behavioural systems (Simmering, Triesch, Deák, & Spencer, 2010). Nonetheless, early choices with respect to model architecture clearly involve strong assumptions about the structure of the brain.

Second, and again as in Bayesian models, connectionist models make assumptions about environmental structure: for example, which inputs are relevant to the process under investigation and how those inputs are related. In Rogers & McClelland's (2004) Semantic Cognition model of categorisation, information about the properties of object category exemplars (e.g., labels such as "robin", and features such as "feathers" and "sing") is input via discrete units. Relationships between these properties are expressed in terms of the adaptable weights between these properties, such as CAN (linking category-characteristic behaviours to exemplars) and ISA (linking labels to exemplars). The structure of the "robin" category is therefore instantiated in linked semantic propositions such as ROBIN+CAN+SING and ROBIN+ISA+BIRD. Thus, the model assumes that relationships between properties in different modalities are functionally identical, that individual properties are equally salient across perceptual modalities, and that elements of the environment such as BIRD and SING are no more related than elements such as CAT and FLY. However, whether these assumptions accurately represent relationships between real-world entities is disputed (Gliozzi, Mayor, Hu, & Plunkett, 2009).

Third, connectionist models make strong assumptions about representational structure, an issue that Bayesian models, for better or worse, sidestep entirely. For example, representations can be localist with an individual unit representing the visual

appearance of an entire object (e.g., Rogers & McClelland, 2004). Alternatively, representations can be distributed, in which multiple units fire concurrently, forming unique patterns in response to different inputs (e.g., Samuelson, 2002). Localist representations appeal to exemplar-based models of categorisation (e.g., Medin, 1978) and resist problems of catastrophic interference, in which learning new representations brings about “forgetting” of older representations (Mayor & Plunkett, 2010). However, whether complex representations can be (or indeed, need be) represented with localist units is disputed (for a review, see Page, 2000).

Important differences in modelling word learning.

Children’s word learning has long fascinated scientists and philosophers alike. Their ability to quickly and accurately assign referents to words in the absence of feedback is astonishing, given the complexity of perceptual environment (Quine, 1960). Nonetheless, environmental cues regularly co-occur, and children learn these regularities (Smith, 2000). Learned associations between cues then allow the learner to accurately predict future co-occurrences (e.g., Wu, Gopnik, Richardson, & Kirkham, 2011). In the case of learning a new word, for example, a child may see a cylindrical item with a handle, whilst hearing the label *cup*, and form the rough, initial hypothesis that the label refers to the object. This ability is known as “fast mapping” (Carey & Bartlett, 1978; Dollaghan, 1985; Heibeck & Markman, 1987). Over repeated encounters with cylindrical objects with a handle alongside the label *cup*, the child learns that new exemplars of cylindrical objects with a handle are likely to also be called *cup*. However, the cognitive processes that underpin learning are debated (e.g., McClelland et al., 2012). Moreover, this debate is embodied by the contrasting theoretical priorities of the two families of model discussed here.

Bayesian models view learning as inference over existing (often innate; Kemp, et al., 2007) structure. Thus, based on prior knowledge that solid objects' shape indicates their category membership, following several encounters with the label *bird* and a creature with wings and two legs, the probability of a new object with wings and two legs also being called *bird* is high. On the Bayesian assumption that humans behave optimally, the optimal inference is that this new object is indeed called *bird*. Emergentist models, on the other hand, view learning as the strengthening of associations between learned representations stored as patterns of activation over time (McClelland, et al., 2011; Smith, 2000). Thus, when the representation for the label *bird* is repeatedly activated alongside the representation for the winged, two-legged object, the two representations become associated. After several encounters, encountering one item (e.g., object with wings and two legs) activates the representation for the other item (e.g., *bird*), even in the absence of that second item.

The theoretical assumptions discussed earlier have implications for Bayesian and emergentist definitions of learning alike. Because Bayesian models ignore mechanism and focus on behaviour, they cannot separate behavioural changes due to physiological development from behavioural changes due to cognitive development (i.e., learning). For example, when an infant begins to reach for objects, Bayesian models cannot tell us whether the infant has learned to move her arm stably towards a goal in the physical environment (physiological development, e.g., Thelen, Corbetta, & Spencer, 1996; Thelen & Spencer, 1998) or has learned to represent objects as bounded, graspable entities (representational development, e.g., Johnson, 2010; Needham & Baillargeon, 1998). Thus, relative to emergentist models, Bayesian models are uninformative as to the timescale of development. As Jones & Love (2011) argue, "In rational [Bayesian] models...nothing develops" (p.182).

The lack of observable “learning” in Bayesian models, compounded by the assumption of optimality, is also problematic for models of atypical learning. Specifically, it is not possible to “lesion” a Bayesian model to examine various cognitive impairments; further, the assumption of optimality has attracted criticism from researchers who point out that human development is not globally optimal but rather locally optimal, that is, limited by some physiological or developmental variable – for example, in the case of visual short-term memory as the limiting factor in category development (Oakes, Ross-Sheehy, & Luck, 2006). In contrast, the emergentist program has generated several informative connectionist models of atypical development (e.g, Mareschal & Thomas, 2007).

Emergentist models, on the other hand, represent learning explicitly, either in adjustment over time of connection weights, in connectionist networks, or as traces of activation in a separate “long term memory” field, in DNFs. Whether paying attention to the neuronal substrates of learning is always an advantage, however, is not clear. Arguably, Bayesian approaches imply a now-questioned Cartesian dualist perspective on cognition and the brain, in which cognition, or mind, is divorced from neural mechanism, or body. However, emergentism requires equally strong assumptions about the *coupling* of brain structure and cognition. Due to the interconnectedness of a neural network, the presence of a node, even when that node’s activation does not reach threshold, affects the dynamics of the entire system. This is borne out by work with constructivist neural networks, in which an initial network is trained until learning plateaus. At this point, a new node is recruited into the network, allowing new patterns of interaction – and therefore new representations – to emerge (Westermann, Sirois, Shultz, & Mareschal, 2006). The architecture of connectionist neural networks therefore means that effectively, structure *is* representation. Confusingly, this lack of

distinction has been described both as a theoretically limiting by-product of an overemphasis on structure (Griffiths, et al., 2010), and as a theoretically innovative conceptualisation of cognition as fundamentally dynamic and emergent (Colunga & Smith, 2008a).

Modelling the shape bias

This section compares a Bayesian and an emergentist model of the shape bias, a much-studied word learning phenomenon in which English-learning children with approximately 100 nouns in their vocabularies begin fast mapping labels to solid objects on the basis of shape similarity (Gershkoff-Stowe & Smith, 2004; Son, Smith, & Goldstone, 2008). The shape bias is commonly tested using a Novel Noun Generalisation task (NNG; e.g., Horst & Twomey, 2012; Samuelson & Smith, 2000). Specifically, the experimenter presents children with an exemplar from a novel category, and gives it a novel label: “This is my *blicket!*” The experimenter then shows the child several novel test objects (e.g., two), each of which match the exemplar on at least one feature. For example, one test object might be the same shape as the exemplar and the other might be made from the same material. The experimenter then asks the child to choose one of the test objects: “Which one’s your *blicket?*”. English-learning children with sufficient experience of categories and their labels will reliably choose the shape-matching test object, leading some to posit a prelinguistic conceptual understanding that shape is a cue to category membership (“shape-as-cue”, Diesendruck & Bloom, 2003; Diesendruck & Graham, 2010), and others to argue that the shape bias emerges from statistical regularities in young children’s early vocabularies (“attentional learning”, Samuelson & Perone, 2010; Smith, et al., 2002). The two models described here illustrate this conflict between top-down and bottom-up explanation of the shape bias.

A Bayesian shape bias model: Kemp, Perfors & Tenenbaum (2007)

In common with the vast majority of Bayesian models, Kemp, Perfors & Tenenbaum (2007; henceforth KP&T) simulate the shape bias at Marr's (1982) computational level. Specifically, the authors describe the shape bias as a form of inductive constraint learned via the formation of inferences-over-inferences or "overhypotheses".

Figure 2 depicts the architecture and procedure. The model consists of a 3-level hierarchy of hypothesis spaces, level 3 defining the hypotheses possible at level 2, which in turn define the hypothesis possible at level 1. Thus, level 3 assumes *a priori* knowledge of what is possible at level 2, which could be either learned or innate.

During training, the model is presented with categories at level 1 via feature vectors (see panel (a) of Figure 2) sampled from a Dirichlet probability distribution (for a review, see J. Huang, 2005) defined by parameters α and β , that is, α and β describe the space of all possible exemplars. Note, however, that viewed differently, α and β also represent overhypotheses, in that they contain the information the model "knows" about objects in general. It is in this sense that the Bayesian model behaves in a top-down fashion.

Categories consist of two exemplars, each represented by a unique feature vector. Each vector contains a category marker; individual feature values between 1 and 10 for shape, colour, and texture; and a size feature value of "1" or "2"⁷. For category 1, for example, at the second level the model assigns a high probability to the overhypothesis "exemplars from category 1 have shape feature 1, but other features vary" (and a low probability to every other possible overhypothesis). At level 3, the model assigns a high

⁷ KT&P do not discuss why this feature is included, although research suggests that size plays little role in children's categorization of solid object (Landau, Smith & Jones, 1988)

probability to the over-overhypothesis “exemplars with the same shape are members of the same category”. Note that the authors do not explicitly discuss labels, but the category marker performs a comparable function (see “Training Vectors” panel of Figure 2).

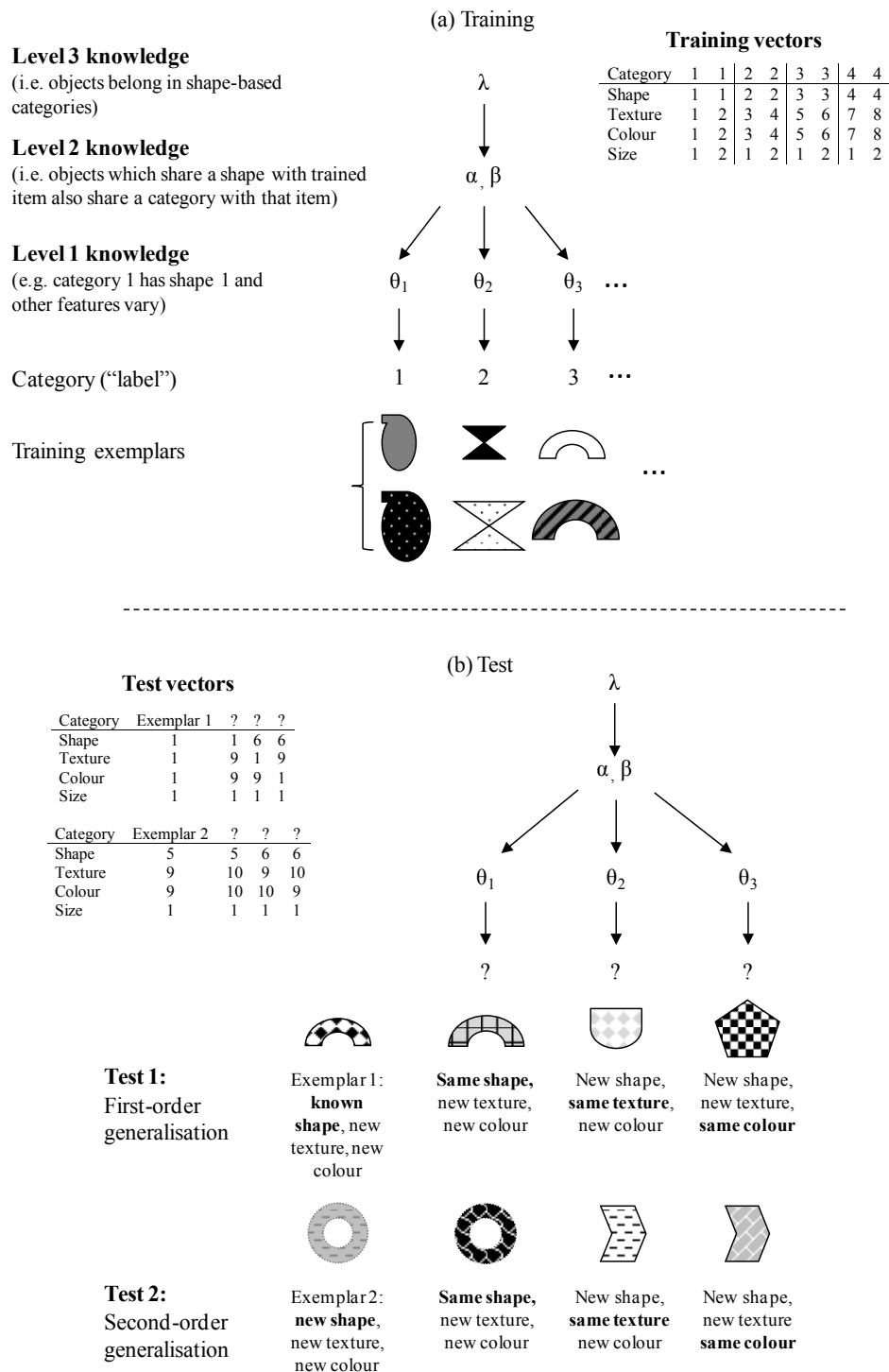


Figure 2. Architecture and procedure used in Kemp, Perfors & Tenenbaum (2007)

Second, the model is presented with a second-order generalisation test, in which an entirely new exemplar is presented. The model is then prompted to choose the most likely category match from three novel test objects – again, a shape match, a texture match and a colour match, but importantly, completely novel in all other features. Again, the probability of each test item sharing a category feature with the exemplar is calculated.

Category	1	1	2	2	3	3	4	4
Shape	1	1	2	2	3	4	5	6
Material	1	2	3	4	5	5	6	6
Size	1	2	1	2	1	2	1	2
Solidity	1	1	1	1	2	2	2	2

Table 1. Feature vectors for the extended model, Kemp, Perfors & Tenenbaum, 2007.

In line with the empirical studies conducted by Smith et al. (2002), the authors found that their model was more likely to categorise objects by shape than would be expected by chance, showing no preference for texture or colour matches. In an extended model, a “solidity” feature replaced colour in the feature vectors, and the training set included “non solid” exemplars in which material and solidity remained constant within categories, but shape and colour varied (see Table 1). This model consistently categorised solid exemplars by shape, and nonsolid exemplars by material, demonstrating not only children’s shape bias and a bias to categorise nonsolid substances by material, but also reflecting their ability to learn about object/substance

ontology (Samuelson & Smith, 2000). Note that the discovery of this distinction was enabled by a preprogrammed assumption that categories *may* be grouped into ontological kinds – that is, the model is told *a priori* that objects and substances may differ.

Although ostensibly a replication of the empirical data, the model's apparent development of a shape bias and a subject-object ontology is perhaps less surprising when viewed in the context of the feature vectors for this second model (see Table 1), which specify that solid objects and nonsolid substances are different: the distinction is specifically encoded in the feature vectors. In terms of modelling the input children receive, the model thus reflects a situation in which children come to the word learning task with an innate substance ontology, in explicit support of the shape-as-cue account (e.g., Diesendruck & Bloom, 2003). Further, the model assumes that solid objects *always* share a shape and a label, that children encounter equally as many words for nonsolid substances as they do for solid object, that object/substance information is equally as influential of categorisation as other featural information, including labels. These assumptions do not have unequivocal empirical support (e.g., Imai & Gentner, 1997; Gary Lupyan, 2012; Samuelson & Smith, 1999). Specifically, solid objects do not always share a shape and a label (for example, a *brush* can be long, thin and cylindrical with narrow bristles, or broad and oval with wide bristles) and some objects share a shape but not a label (for example, a *ball* and an *orange* are both spherical; see also Bloom & Markson, 1998). Further, children do not encounter as many words for nonsolid substances as solid objects, and words for nonsolid substances are less consistent in referring to items that share a material than are words for solid objects in referring to items that share a shape (Samuelson & Smith, 1999). Clearly, then the

inputs to KP&T's model do not accurately reflect the structure of the linguistic environment of English-learning children.

The KP&T model highlights the importance of grounding Bayesian models in the environment: here, the model is weakened by assumptions made not about the learner but about the inputs – that all real-world entities are grouped into objects and substances before learning takes place. This is perhaps most clearly demonstrated by the fact that in the real world, English-speaking children's material bias for nonsolid substances is much more fragile than the bias demonstrated by this model (e.g., Samuelson & Horst, 2007). Unfortunately, whether a more representative result would be generated if the model were given more strongly environmentally-grounded input set is not addressed.

An emergentist shape bias model: Colunga & Smith (2005).

In contrast to the KP&T model, Colunga & Smith (2005; henceforth C&S) model the shape bias using a neurologically-inspired connectionist network, with learning specifically instantiated in a Hebbian algorithm that increases the excitatory weights between neurons that fire simultaneously.

Figure 3 depicts the architecture of the C&S model. The model receives input from feature vectors via a perceptual layer, consisting of three networks representing an object's shape, material (including colour) and solidity. Shape and material inputs consist of patterns of activation distributed across the networks. Solidity is represented locally, by the activation of one of two solidity units. Labels are also represented locally in a separate word layer. The perceptual and the word layers are both coupled reciprocally to a hidden layer⁸, and all layers are recurrently connected to themselves.

⁸ This layer is "hidden" in that it receives input internally via the perceptual and label layers, rather than externally via the perceptual layer. Activation in the hidden layer is, however, available for inspection.

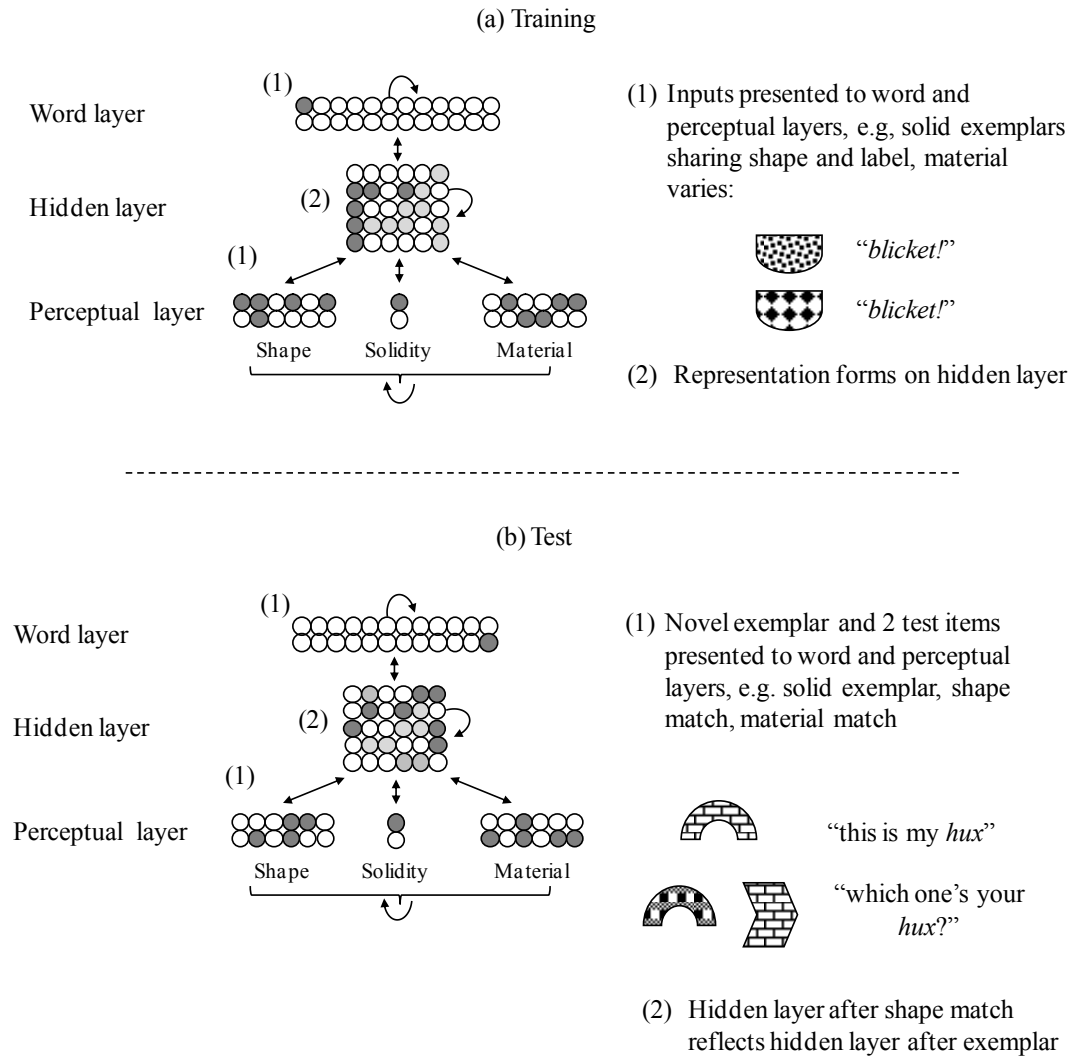


Figure 3. Architecture and procedure used in Colunga & Smith (2005). Arrows denote excitatory connections between layers.

During training, the model is presented with 20 categories, 10 from solid/shape categories and 10 from nonsolid/material categories. For each solid/shape category, shape input for each category is first generated randomly and then held constant for each training exemplar from that category. Material input is generated randomly for each exemplar. Thus, for shape/solid categories, within each category, shape is constant but material is variable. Shape and material inputs are presented simultaneously with

the activation of the single “solid” node as well as the activation of a single label node.

Both of these are also held constant for exemplars from the same category.

Material/nonsolid exemplars are generated in a corresponding manner. That is, material, nonsolid and label inputs are held constant within categories while shape varies.

Training consists of 100 different exemplars for each of the 20 categories.

The model is tested with a simulated NNG task. At test the model is presented with one completely novel exemplar followed by two completely novel test objects, one at a time (test objects could be solid or nonsolid shape matches, or solid or nonsolid material matches). The pattern of activation on the hidden layer for each test item is compared with the previous pattern of activation on the hidden layer for the exemplar, and the absolute difference between the two is calculated. The *probability* of having chosen a shape match is then calculated across a total of 40 exemplars using Luce’s choice rule,

In line with Smith et al.’s (2002) empirical results and KP&T’s Bayesian model, C&S’s model demonstrated systematic patterns of generalisation, matching solid exemplars to test items of the same shape and nonsolid exemplars to test items of the same material. Further, the average absolute difference between internal representations for shape/solid items and material/nonsolid items was smaller than the average absolute difference between representations for shape/nonsolid and material/solid items. Put another way, representations for shape/solid exemplars (and material/nonsolid exemplars) were more similar to one another than representations for shape/nonsolid exemplars (and material/solid exemplars). Thus, after training with a simple idealised vocabulary, the network formed the implicit higher-order generalisation that “solid things belong in shape-based categories” – without being explicitly programmed with the capacity to do so.

In a second experiment the model was presented with a training set that more accurately reflected the statistical regularities in the early vocabularies of English-learning children, as well as the wider perceptual variation in shape of solid objects relative to comparison to shape variation in nonsolid substances (Samuelson & Smith, 1999). The model showed a strong shape bias, and generated the novel prediction that infants should generalise novel names for nonsolid substances by material only when those materials were presented in “natural” formations like smears or heaps, a prediction borne out by a further empirical study.

The authors acknowledge that empirical evidence for differential processing of specific types of category may initially appear to point to dedicated cognitive modules for processing specific perceptual inputs (e.g., “representational” or “conceptual” primitives; Carey, 2011; Mandler, 2012). However, this model learns substance ontology via a simple, unbiased associative mechanism. C&S argue that seemingly propositional knowledge about the world, for example, “solid things are the same type of thing” can emerge from experience rather than *a priori* structure and can be represented in a graded, distributed manner.

Evaluating Bayesian and emergentist shape bias models

Two distinct types of model have been presented that successfully model the same phenomenon. Given an idealised vocabulary, in both cases a schematised learner acquired a shape bias, and even more strikingly, learned an ontological distinction between solid objects and nonsolid substances – an apparent triumph for both approaches. However, the current models exhibit important structural and theoretical differences – with implications for our interpretation of their findings.

Modelling cognitive representation. The KP&T (2007) model does not explicitly represent perceptual input. In this sense, the model simply simulates

children's categorisation *behaviour* whilst leaving to one side issues of representation: categorisation is established via the calculation of the probability of category membership for each new exemplar. From this perspective, the Bayesian take on the shape bias simulates optimal behaviour in the NNG task. In ignoring any other level than the computational, the authors were able to show a systematic behaviour arising purely from statistical regularities in the environment without having to make strong assumptions about the nature of perceptual input.

In contrast, the C&S (2005) model does generate explicit representations, through the activation distributed in the hidden layer. Categorisation is established via comparison of activation on this layer in response to different exemplars: similar activation is taken to reflect the model having recognised the exemplars as belonging to the same category. However as KP&T argue, categorisation is context-dependent, and factors such as taxonomic relations and syntactic frame also affect children's behaviour in NNG tasks (Gelman & Bloom, 2000; Subrahmanyam, Landau, & Gelman, 1999). On this view, the C&S model as described only examines one piece of the word learning puzzle. However, the Bayesian model as reported also ignores these wider issues. Of course, for both models, issues of parsimony and computational capacity would have rendered modelling the entire word learning process unfeasible.

Unlike C&S's model, KP&T's model explicitly represents hierarchical structure. In contrast, C&S's model learned an *implicit* substance ontology, or overhypothesis, by which shape/solid and material/nonsolid representations were more internally coherent than shape/nonsolid and material/solid, in line with children's behaviour in the NNG task (see also Perry & Samuleson, 2011). This representational distinction emerged, unsupervised, from the dynamics of the network coupled with the statistics of the vocabulary input alone. Thus, KP&T's criticism of the C&S model's ability to

represent first- and second-order generalisations holds only on the assumption that such abstractions need be explicitly – that is, innately – represented.

Conversely, although the hierarchical structure of the Bayesian model neatly captures increasingly more abstract levels of category knowledge, this surface clarity comes at a price. First, because the hierarchical structure is designed by the modeller, KP&T's model does not address development over time – the model assumes that there is no structural difference between children's and adults' categories. However, whether children's categories are structured in the same way as adults is by no means established (Rakison, 2000). Similarly, any model based on this assumption neglects the possibility that children's representational taxonomies change qualitatively over development.

Second, in the Bayesian paradigm, the modeller – and not the model – guides what can be learned, top down. As KP&T concede, the inevitable conclusion of current Bayesian approaches is that some cognitive structure must be innate; however, the nature/nurture problem has been lengthily and acrimoniously discussed. Recent calls for a paradigm shift in cognitive science (e.g., Johnson, 2010; Spencer, et al., 2009) in order to move beyond this well-worn debate imply that a rigid adherence to innatism may serve only to limit Bayesian models' ability to generate novel, informative and testable predictions with regard to cognition.

Modelling the environment. Both KP&T and C&S claim their models learn structure from statistical regularities in the environment, in this case, the co-occurrence of labels with perceptual features. To an extent, this is true: both models take input from vectors which analogically represent different features of the things young children encounter, for example, shape, material, and solidity.

However, a close comparison of the inputs to the two models reveals differences between the two models' interpretation of the task environment; and despite the

argument that Bayes offers exceptional scope for modelling environmental regularities (Fernbach & Sloman, 2011; Jones & Love, 2011), of the two models under comparison, inputs to C&S's model are clearly most strongly environmentally-grounded.

First, the two models simulate category labels very differently. C&S's model instantiates words separately from perceptual input, in line with recent categorisation studies, which demonstrate that words may drive categorisation over and above correlations between visual features (Mather & Plunkett, 2010; Plunkett, et al., 2008; but see Gliozzi, et al. for a contrasting account). KP&T, however, employ a "category" property in their feature vectors, but do not clearly state what this represents. The reader must therefore choose: in conjunction with the likelihood function this "category" property must either represent either the learner's knowledge of category labels specifically, or the learner's knowledge of categories in general. If the "category" feature represents labels, then this model assumes that there is a one-to-one mapping between categories and labels, an assumption undermined by a large body of work on taxonomies categorical structure (Belchadha, 1996; Mandler & Bauer, 1988; Mayor & Plunkett, 2010; Pauen, 2002; Quinn & Johnson, 2000). If the "category" feature does not represent labels, then the task environment in which KP&T's model is situated does not reflect the task environment of the empirical study they attempt to replicate; that is, KP&T's task cannot reflect the NNG task described by Smith et al. (2002), because the model includes no "noun" input. Thus, the remainder of this section assumes that the authors do intend to model labelling.

Second, because the structure of the inputs to KP&T's extended model is markedly different to the observed structure of young children's early vocabularies (Samuelson & Smith, 1999), KP&T in fact fail to replicate the empirical studies. Specifically, children in Smith et al., (2002) exhibited a strong shape bias and a much

weaker material bias (see also Samuelson, 2002). KP&T's model demonstrated a strong shape bias, but also an equally strong material bias. This nonrepresentative pattern of responding stems from the perfect correlation between the "category" feature and the "shape" feature (for solid items) or the "material" feature (for nonsolid items; see Table 1) in the vocabulary presented to the KP&T model; that is, shape and material are perfect indicators of category membership. Given the absence of stochasticity in KP&T's feature vectors, *not* categorising solid objects by shape would perhaps be more surprising than the shape and material biases the authors describe.

KP&T acknowledge that further research is needed to establish why their model fails to replicate Smith et al.'s (2002) data but do not attempt to address the question. Indeed, based on their current design it is difficult to see how this could be done without significant changes to the model's architecture (e.g., modelling with labels separately from other perceptual features) – arguably, abandoning the existing model and developing a new one entirely.

C&S's first simulation also employs a vocabulary with perfect correlations between shape/material and solidity, and their first model consequently exhibits the same pattern of generalisation as KP&T's model: a strong shape bias and a strong material bias. However, when the model was trained with a vocabulary closely reflecting the statistical regularities in young children's early vocabularies, the model exhibited a weakened material bias, replicating the Smith et al. (2002) results.

Third, the two models exhibit substantial differences in the task procedure. During training, the C&S model encounters the equivalent of encountering 100 different exemplars of 20 distinct categories, alongside the relevant category label. Although less than the 100-or-so nouns suggested by Gershkoff-Stowe & Smith (2004) as the threshold for the onset of the shape bias, the frequency of labelling goes some way to

reflecting the considerable amount of labelling experience amassed by young children. In stark contrast, KP&T's model is presented with just two exemplars per category for a total of four categories. Although designed to reflect the longitudinal training design employed by Smith et al. (2002) with 17-month-old children, this approach does not reflect children's experience of categories and labels at that age. Indeed, because several studies have demonstrated that English-speaking children only acquire a shape bias after substantial experience with categories and category labels (Gershkoff-Stowe & Smith, 2004; Perry, Samuelson, Malloy, & Schiffer, 2010; Samuelson, 2002; Smith, et al., 2002), KP&T's swift acquisition of a shape bias suggests that the computations it carries out may be more efficient than, and therefore not reflective of, those carried out by young children.

Converging support for the emergentist approach. Overall, both models inform our understanding of children's categorisation, demonstrating that generalisation biases can be learned from regularities in input. However, KP&T fail to replicate their target data, despite allowing their model considerable assumptions about *a priori* structure. Further, the model is untested: the only prediction it makes is the presence of a strong shape bias in the absence of naming, for which, as noted, the empirical evidence is equivocal. On the other hand, C&S successfully replicate the target data by increasing the ecological validity of their model (that is, providing it with a more representative vocabulary), and go on to generate and replicate novel predictions. Thus, C&S present a rigorously tested, environmentally-grounded model of a word learning phenomenon, whilst KP&T, lacking a serious account of the task environment and failing to make any replicable predictions, can only claim with any confidence that they provide a proof-of-concept that hierarchical Bayesian models can make higher-order generalisations.

General Discussion

Bridging the gap between Bayesians and emergentists.

Clearly, the Bayesian approach just described is less informative to our understanding of word learning and categorisation than is the emergentist approach. Dealing with the effect of fine-grained and graded perceptual inputs is not a Bayesian strength, and Bayesian models of implicit cognitive processes such as word learning seem unlikely to contribute anything new to the field. However, Bayesian models do have their place, and in that place, can be highly informative. For example, recent work in the Bayesian tradition has successfully captured adults' reasoning in conditional inference tasks, as well as generated novel predictions, which have subsequently been empirically replicated (Oaksford & Chater, 2003, 2009, 2011). These studies suggest that adults do use probabilistic reasoning for explicit problem-solving, and provide a strong foundation for further research into optimality in cognition. In general, Bayesian models, with their *a priori* assumptions of a structured cognition, are invaluable in investigating the interaction between such structures and behaviour – although not, of course, in exploring from where such structures originate (Chater, Tenenbaum & Yuille, 2006).

So, how to heal the rift between the Bayesians and the emergentists? Borsboom, Wagenmakers and Romeijn (2011) emphasise that Bayesian and emergentist models attack different questions, and that this difference hinges on the subtle difference between process-based and mechanistic models. Process-based models simulate how a system moves from one state to another, just as the boxes and arrows on a flow chart illustrate the different stages in a process. Mechanistic models, in contrast, examine how different parts of a system influence each other in time (e.g. Rogers & McClelland, 2004). Bayesian models, from this perspective, are process models, in that they

describe how an entity moves through a series of states (that is, the repeated updating of probabilities of hypotheses) as a result of encountering new data. Thus, Bayesian models are *not* mechanistic precisely *because* they are process-based: they do not seek to answer how a process moves from one stage to another, but instead illustrate what these stages are, and in what order they can occur. Essentially, the two camps are addressing orthogonal research programs: the emergentists focusing on algorithmic simulations of dynamically emergent cognitive structure, and the Bayesians focusing on inference in a probabilistically-structured environment. Viewed this way, it is no surprise that computational models of cognitive development so often prove controversial; what is surprising is the degree of suspicion between the two camps, given that the techniques themselves need not be contradictory.

It is possible to envisage a future for computational models in which the two approaches reach a type of synergy (Jones & Love 2011). For example, a Bayesian model could usefully describe a hierarchical process, the various stages of which could be modelled individually using emergentist, environmentally-grounded models, thus addressing emergentist models' inability to simulate behaviours ostensibly contingent on hierarchical structure and Bayesian models' lack of mechanism and learning. Indeed, recently researchers have begun to focus on integrating inference with environmentally-grounded emergence (Barsalou, 2009), or generating Bayesian priors via simulated neural activation (Köver & Bao, 2010). Similarly, Jones & Love's (2011) proposed Enlightenment Bayes proposes that the Bayesian research program could indeed inform cognitive research, providing it incorporates traditionally emergentist values such as the difference between representation and environmental input.

Clearly, the goal of a unified model of cognitive development lies some way off. Rather than the current theoretical entrenchment of either school, however, an

atmosphere of collaboration among proponents of Bayesian and emergentist models will undoubtedly generate novel accounts of cognition and development. The development of computational modelling signposted a new era in cognitive science, and the integration of these two great traditions can only serve to deepen our understanding not only of cognition and behaviour, but also of their complex and fascinating interactions over development.

References

- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3), 409. DOI: 10.1037/0033-295X.98.3.409
- Aslin, R., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Attention, Perception, & Psychophysics*, 17(3), 293-302. DOI: 10.3758/BF03203214
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1281-1289. DOI: 10.1098/rstb.2008.0319
- Baumgartner, H. A., & Oakes, L. M. (2011). Infants' developing sensitivity to object function: Attention to features and feature correlations. *Journal of Cognition and Development*, 12(3), 275-298. DOI: 10.1080/15248372.2010.542217
- Behlchadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, 60(2), 105–141. DOI: 10.1016/0010-0277(96)00706-8
- Berent, I., & Pinker, S. (2008). Compound formation is constrained by morphology: A reply to Seidenberg, MacDonald Haskell. *The Mental Lexicon*, 3(2), 176-187. DOI: 10.1075/ml.3.2.02ber
- Berthouze, L., & Metta, G. (2005). Epigenetic robotics: Modelling cognitive development in robotic systems. *Cognitive Systems Research*, 6, 189-192. DOI: 10.1016/j.cogsys.2004.11.002
- Bloom, P., & Markson, L. (1998). Intention and analogy in children's naming of pictorial representations. *Psychol. Sci.*, 9(3), 200–204. DOI: 10.1111/1467-9280.00038
- Booth, A. E. (2009). Causal supports for early word learning. *Child development*, 80(4), 1243-1250. DOI : 10.1111/j.1467-8624.2009.01328.x

- Borsboom, D., Wagenmakers, E. J., & Romeijn, J. W. (2011). Mechanistic curiosity will not kill the Bayesian cat. *Behavioral and Brain Sciences*, 34(4), 192. DOI: 10.1017/S0140525X11000215
- Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138(3), 389. DOI: 10.1037/a0026450
- Brabazon, A., & O'Neill, M. (2008). Natural computing in computational finance. *Studies in Computational Intelligence*. DOI: 10.1007/978-3-540-77477-8
- Carey, S. (2011). Concept innateness, concept continuity, and bootstrapping. *Behavioral and Brain Sciences*, 34(03), 152-162.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Chater, N., Goodman, N., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences*, 34(4), 194-196. DOI: 10.1017/S0140525X11000239
- Chater, N., Oaksford, M., Nakisa, R., & Redington, M. (2003). Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes*, 90(1), 63-86. DOI: 10.1016/S0749-5978(02)00508-3
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006a). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, 10(7), 287-291. DOI: 10.1016/j.tics.2006.05.007
- Chater, N., Tenenbaum, J. B., & Yuille, A. (2006b). Technical introduction: A primer on probabilistic inference. *Trends in Cognitive Sciences*, 10(7), Additional online material.

- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234. DOI: 10.1037/0033-295X.113.2.234
- Chen, C. (2010). *Handbook of pattern recognition and computer vision*: World Scientific.
- Chen, H., Wu, Q. X., Wang, Y., Lin, M. Y., & Cai, R. T. (2011). Identification of vehicle types based on spiking neural network model. *Jisuanji Xitong Yingyong-Computer Systems and Applications*, 20(4), 182-185.
- Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development*, 50(2), 419-424. DOI: 10.2307/1129417
- Colunga, E., & Smith, L. B. (2003). The emergence of abstract ideas: Evidence from networks and babies. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358(1435), 1205-1214. DOI: 10.1098/rstb.2003.1306
- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112(2), 347-382. DOI: 10.1037/0033-295X.112.2.347
- Colunga, E., & Smith, L. B. (2008a). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology*, 26(2), 174-192. DOI: 10.1016/j.newideapsych.2007.07.012
- Colunga, E., & Smith, L. B. (2008b). Knowledge embedded in process: The self-organization of skilled noun learning. *Developmental Science*, 11(2), 195-203. DOI 10.1111/j.1467-7687.2007.00665.x.
- Cooper, R. P., & Shallice, T. (2010). Cognitive neuroscience: The troubled marriage of cognitive science and neuroscience. *Topics in Cognitive Science*, 2(3), 398-406. DOI: 10.1111/j.1756-8765.2010.01090.x

- Crick, F. (1989). The recent excitement about neural networks. *Nature*, 337(6203), 129.
DOI : 10.1038/337129a0
- Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 431-452). Oxford, UK: Oxford University Press. DOI: 10.1093/acprof:oso/9780199216093.003.0019
- Dawson, C., & Gerken, L. (2011). When global structure "Explains Away" local grammar: A Bayesian account of rule-induction in tone sequences. *Cognition*, 120(3), 350-359. DOI 10.1016/j.cognition.2010.11.016
- Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, 6(3), 274. DOI: 10.1177/1745691611406920
- Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development*, 74(1), 168-178. DOI: 10.1111/1467-8624.00528
- Diesendruck, G., & Graham, S. A. (2010). Kind matters: a reply to Samuelson & Perone. *Cognitive Development*, 25, 149-153. DOI: 10.1016/j.cogdev.2010.02.003
- Dollaghan, C. (1985). Child Meets Word Fast Mapping in Preschool-Children. *Journal of Speech and Hearing Research*, 28(3), 449-454. PMid: 4046586
- Dombi, G., Rosbolt, J., & Severson, R. (2010). Neural network analysis of employment history as a risk factor for prostate cancer. *Computers in Biology and Medicine*, 40(9), 751-757. DOI: 10.1016/j.combiomed.2010.07.002
- Duda, R. O., Hart, P. E., & Nilsson, N. J. (1976). Subjective Bayesian methods for rule-based inference systems: DTIC Document.
- Eisner, J. (2002). Discovering syntactic deep structure via Bayesian statistics. *Cognitive Science*, 26(3), 255-268. DOI: 10.1207/s15516709cog2603_3

- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179-211. DOI: 10.1207/s15516709cog1402_1
- Fava, E., Hull, R., & Bortfeld, H. (2011). Linking Behavioural and Neurophysiological Indicators of Perceptual Tuning to Language. *Frontiers in Psychology*, 2. DOI: PMid:21738514
- Fernbach, P. M., & Sloman, S. A. (2011). Don't throw out the Bayes with the bathwater. *Behavioral and Brain Sciences*, 34(4), 198-199. DOI: 10.1017/S0140525X11000264
- French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3-to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology-General*, 133(3), 382-397. DOI: 10.1037/0096-3445.133.3.382.
- Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought*: Oxford University Press, USA.
- Gelman, S. A., & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition*, 76(2), 91-103. DOI: 10.1016/S0010-0277(00)00071-8
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098-1114. DOI: 10.1111/j.1467-8624.2004.00728.x
- Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, 33(4), 709-738. DOI: 10.1111/j.1551-6709.2009.01026.x
- Glymour, C. (2011). Osiander's psychology. *Behavioral and Brain Sciences*, 34(04), 199-200. DOI: 10.1017/S0140525X11000276

- Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004).
A theory of causal learning in children: causal maps and Bayes nets.
Psychological Review, 111(1), 3. DOI: 10.1037/0033-295X.111.1.3
- Gopnik, A., & Tenenbaum, J. B. (2007). Bayesian networks, Bayesian learning and
cognitive development. *Developmental Science*, 10(3), 281-287. DOI:
10.1111/j.1467-7687.2007.00584.x
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol.
1974): Wiley New York.
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010).
Probabilistic models of cognition: exploring representations and inductive biases.
Trends in Cognitive Sciences, 14(8), 357-364. DOI: 10.1016/j.tics.2010.05.004
- Gurney, K. (1997). *An introduction to neural networks*. London: UCL Press. DOI:
10.4324/9780203451519
- Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A
Bayesian approach to reasoning fallacies. *Psychological Review*, 114(3), 704.
DOI: 10.1037/0033-295X.114.3.704
- Hebb, D. O. (2002/1949). *The organization of behavior: A neuropsychological theory*:
Lawrence Erlbaum.
- Heibeck, T. H., & Markman, E. M. (1987). Word learning in children - an examination
of fast mapping. *Child Development*, 58(4), 1021-1034. DOI: 10.2307/1130543
- Horst, J. S., & Twomey, K. E. (2012). It's taking shape: Shared features influence
children's novel noun generalizations. *Infant & Child Development*.
DOI:10.1002/icd.1768
- Huang, J. (2005). Maximum likelihood estimation of Dirichlet distribution parameters.
CMU Technique Report.

- Huang, Y. T., & Pinker, S. (2010). Lexical semantics and irregular inflection. *Language and cognitive processes*, 25(10), 1411-1461. DOI: 10.1080/01690961003589476
- Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, 62(2), 169-200. DOI: 10.1016/S0010-0277(96)00784-6
- Johnson, S. P. (2010). Perceptual completion in infancy. In S. P Johnson (Ed.), *Neoconstructivism: The new science of cognitive development*. Oxford: Oxford University Press.
- Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(04), 169-188. DOI: 10.1017/S0140525X10003134
- Jordan, M. I. (1986). An introduction to linear algebra in parallel distributed processing. *Parallel distributed processing*, 1, 365-422.
- Kelso, J., Ding, M., & Schöner, G. (1993). Dynamic pattern formation: A primer. In L. B. Smith & E. Thelen (Eds.), *A dynamic systems approach to development: Applications*. Cambridge, MA: MIT Press.
- Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, 10(3), 307-321. DOI: <http://dx.doi.org/10.1111/j.1467-7687.2007.00585.x>
- Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information-processing view*. Erlbaum Hillsdale, NJ.
- Kolb, B., & Whishaw, I. Q. (2003). *Fundamentals of Human Neuropsychology* (Fifth ed.). New York: Worth Publishers.

- Köver, H., & Bao, S. (2010). Cortical plasticity as a mechanism for storing Bayesian priors in sensory perception. *PloS one*, 5(5), DOI: 10.1371/journal.pone.0010497
- Lachman, R., & Butterfield, E. C. (1979). *Cognitive psychology and information processing: An introduction*: Lawrence Erlbaum.
- Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language*, 31(6), 807-825. DOI: 10.1016/0749-596X(92)90040-5
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. 10.1016/0885-2014(88)90014-7
- Lee, M. D., & Sarnecka, B. W. (2010). A model of knower-level behavior in number concept development. *Cognitive Science*, 34(1), 51-67. DOI: 10.1111/j.1551-6709.2009.01063.x
- Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*: Springer Verlag.
- Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, 108(2), 566-577. DOI: 10.1016/j.cognition.2008.03.009
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label feedback hypothesis. *Frontiers in Psychology*, 3. DOI: 10.3389/fpsyg.2012.00054
- Madole, K. L., & Oakes, L. M. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, 19(2), 263-296. DOI: 10.1006/drev.1998.0481

- Mandler, J. M. (2012). On the spatial foundations of the conceptual system and its enrichment. *Cognitive Science*. DOI: 10.1111/j.1551-6709.2012.01241.x
- Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic? *Cognitive Development*, 3(3), 247–264. DOI:10.1016/0885-2014(88)90011-1
- Mandler, J. M., & McDonough, L. (1998). Studies in inductive inference in infancy. *Cognitive psychology*, 37(1), 60-96. DO: 10.1006/cogp.1998.0691
- Mareschal, D., & Thomas, M. S. C. (2007). Computational modeling in developmental psychology. *Evolutionary Computation, IEEE Transactions on*, 11(2), 137-150.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*: WH Freeman.
- Mather, E., & Plunkett, K. (2010). Novel labels support 10-month-olds' attention to novel objects. *Journal of Experimental Child Psychology*, 105(3), 232-242. doi: DOI 10.1016/j.jecp.2009.11.004. DOI: 10.1016/j.jecp.2009.11.004
- Mayor, J., & Plunkett, K. (2010). A Neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117(1), 1-31. doi: 10.1037/a0018130. DOI: 10.1037/a0018130
- McCarthy, J. (2008). The well-designed child. *Artificial Intelligence*, 172(18), 2003-2014. DOI : 10.1016/j.artint.2008.10.001
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2011). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14(8), 348-356. DOI: 10.1016/j.tics.2010.06.002
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4), 115-133.

- McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford: Oxford University Press.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (in press). Word learning as the interaction of online referent selection and slow associative learning. *Psychological Review*.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85(3), 207. DOI: 10.1037/0033-295X.85.3.207
- Molfese, D. L., & Burger-Judisch, L. M. (1991). Dynamic temporal-spatial allocation of resources in the human brain. An alternative to the static view of hemispheric differences. *Cerebral laterality: Theory and research*, 71-102.
- Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4(1), 79-95. DOI: 10.3758/BF03210778
- Needham, A., & Baillargeon, R. (1998). Effects of prior experience on 4.5-month-old infants' object segregation. *Infant Behavior & Development*, 21(1), 1-24. DOI: 10.1016/S0163-6383(98)90052-2
- Norris, D. (2011). The illusion of mechanism: Mechanistic fundamentalism or enlightenment? *Behavioral and Brain Sciences*, 34(04), 208-209. DOI: 10.1017/S0140525X11000379
- Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants' object examining: Habituation and categorization. *Cognitive Development*, 6(4), 377-392. DOI: 10.1016/0885-2014(91)90045-F
- Oakes, L. M., Ross-Sheehy, S., & Luck, S. J. (2006). Rapid development of feature binding in visual short-term memory. *Psychological Science*, 17(9), 781-787. DOI: 10.1111/j.1467-9280.2006.01782.x

- Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, 18(4), 359-379. DOI: 10.1111/1468-0017.00232
- Oaksford, M., & Chater, N. (2009). The uncertain reasoner: Bayes, logic, and rationality. *Behavioral and Brain Sciences*, 32(01), 105-120. DOI: 10.1017/S0140525X0900051X
- Oaksford, M., & Chater, N. (2011). Causation and conditionals in the cognitive science of human reasoning. *Open Psychology Journal*, 3, 105-118. DOI: 10.2174/1874350101003020105
- Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(4), 883. DOI: 10.1037/0278-7393.26.4.883
- Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, 23(04), 443-467. DOI: 10.1017/S0140525X00003356
- Pauen, S. (2002). The global-to-basic level shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioral Development*, 26(6), 492-499. DOI: 10.1080/01650250143000445
- Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, 120(3), 302-321. DOI: 10.1016/j.cognition.2010.11.015. DOI: 10.1016/j.cognition.2010.11.015
- Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, 2, 345. DOI: 10.3389/fpsyg.2011.00345
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization

and Word Learning. *Psychological Science*, 21(12), 1894-1902. DOI:
10.1177/0956797610389189

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28(1-2), 73-193. DOI: 10.1016/0010-0277(88)90032-7

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665-681. DOI:
10.1016/j.cognition.2007.04.003

Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge: MIT Press.

Quinn, P. C., & Johnson, M. H. (2000). Global-Before-Basic Object Categorization in Connectionist Networks and 2-Month-Old Infants. *Infancy*, 1(1), 31–46. DOI:
10.1207/S15327078in0101_04

Rakison, D. H. (2000). When a Rose Is Just a Rose: The Illusion of Taxonomies in Infant Categorization. *Infancy*, 1(1), 77-90. DOI: 10.1207/S15327078in0101_07

Rakison, D. H. (2007). Is consciousness in its infancy in infancy? *Journal of Consciousness Studies*, 14(9-10), 66-89.

Riley, M. A., Shockley, K., & Van Orden, G. (2012). Learning from the body about the mind. *Topics in Cognitive Science*.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386. DOI:
10.1037/h0042519

Rousseau, W. F. (1968). A method for computing probabilities in complex situations:
DTIC Document.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by
back-propagating errors. *Nature*, 323(6088), 533-536. DOI: 10.1038/323533a0

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English
verbs. In J. L. McClelland, D. E. Rumelhart, & and the P. research group (Eds.),
Parallel Distributed Processing: Explorations in the Microstructure of
Cognition (pp. 216–271). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing:*
Psychological and biological models (Vol. 2): The MIT press.

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language
acquisition in connectionist models and 15-20-month-olds. *Developmental*
Psychology, 38(6), 1016-1037. DOI: 10.1037//0012-1649.38.6.1016

Samuelson, L. K., Horst, J. S., Schutte, A. R., & Dobberty, B. N. (2008). Rigid
thinking about deformables: do children sometimes overgeneralize the shape
bias? *Journal of Child Language*, 35(3), 559-589. DOI:
10.1017/S0305000908008672

Samuelson, L. K., & Perone, S. (2010). Rethinking conceptually-based inference -
Grounding representation in task and behavioral dynamics: Commentary on
"Fifteen-month-old infants attend to shape over other perceptual properties in an
induction task," by S. Graham and G. Diesendruck, and "Form follows function:
Learning about function helps children learn about shape," by E. Ware and A.
Booth. *Cognitive Development*, 25(2), 138-148. DOI:
10.1016/j.cogdev.2010.02.002

- Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1-33. DOI: 10.1016/S0010-0277(99)00034-7
- Samuelson, L. K., & Smith, L. B. (2000). Children's attention to rigid and deformable shape in naming and non-naming tasks. *Child Development*, 71(6), 1555-1570. DOI: 10.1111/1467-8624.00248
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, 6(12), e28095. DOI: 10.1371/journal.pone.0028095
- Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, 117(4), 1144. DOI: 10.1037/a0020511
- Schlesinger, M., & Parisi, D. (2004). Beyond backprop: emerging trends in connectionist models of development: an introduction. *Developmental Science*, 7(2), 131-132. DOI: 10.1111/j.1467-7687.2004.00329.x
- Schutte, A. R., & Spencer, J. P. (2002). Generalizing the dynamic field theory of the A-not-B error beyond infancy: Three-year-olds' delay- and experience-dependent location memory biases. *Child Development*, 73(2), 377-404. DOI: 10.1111/1467-8624.00413
- Schutte, A. R., & Spencer, J. P. (2009). Tests of the Dynamic Field Theory and the Spatial Precision Hypothesis: Capturing a Qualitative Developmental Transition in Spatial Working Memory. *Journal of Experimental Psychology-Human Perception and Performance*, 35(6), 1698-1725. DOI: 10.1037/A0015794.
- Schutte, A. R., Spencer, J. P., & Schöner, G. (2003). Testing the dynamic field theory: Working memory for locations becomes more spatially precise over

development. *Child Development*, 74(5), 1393–1417. DOI: 10.1111/1467-8624.00614

Shultz, T. R. (2007). The Bayesian revolution approaches psychological development. *Developmental Science*, 10(3), 357-364. DOI: 10.1111/j.1467-7687.2007.00588.x

Simmering, V. R., & Spencer, J. P. (2008). Generality with specificity: the dynamic field theory generalizes across tasks and time scales. *Developmental Science*, 11(4), 541-555. DOI: 10.1111/j.1467-7687.2008.00700.x.

Simmering, V. R., Triesch, J., Deák, G. O., & Spencer, J. P. (2010). A dialogue on the role of computational modeling in developmental Science. *Child Development Perspectives*, 4(2), 152-158. DOI: 10.1111/j.1750-8606.2010.00134.x

Smith, L. B. (2000). Learning How to Learn Words: An Associative Crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a Word Learner: A Debate on Lexical Acquisition* (pp. 51-80). New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780195130324.003.003

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13-19. DOI: 10.1111/1467-9280.00403

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, 108(3), 626-638. DOI: 10.1016/j.cognition.2008.05.002

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer

abide the nativist-empiricist debate. *Child Development Perspectives*, 3(2), 79-87. DOI: 10.1111/j.1750-8606.2009.00081.x

Spencer, J. P., & Schöner, G. (2000). A dynamic field model of location memory.

Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society, 924–929 1075.

Subrahmanyam, K., Landau, B., & Gelman, R. (1999). Shape, material, and syntax:

Interacting forces in children's learning in novel words for objects and substances. *Language & Cognitive Processes*, 14(3), 249-281. DOI: 10.1080/016909699386301

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, 24(4), 629-640. DOI:

10.1017/S0140525X01000061

Thelen, E., Corbetta, D., & Spencer, J. P. (1996). Development of reaching during the first year: Role of movement speed. *Journal of Experimental Psychology-*

Human Perception and Performance, 22(5), 1059-1076. DOI: 10.1037/0096-1523.22.5.1059

Thelen, E., Schöner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of

embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24(1). DOI: 10.1017/S0140525X01003910

Thelen, E., & Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.

Thelen, E., & Spencer, J. P. (1998). Postural control during reaching in young infants: A dynamic systems approach. *Neuroscience and Biobehavioral Reviews*, 22(4),

507-514. DOI: 10.1016/S0149-7634(97)00037-7

- Twomey, K. E., Morse, A., & Horst, J. S. (in press). An Embodied Model of Young Children's Categorization and Word Learning. In L. J. Gogate & G. Hollich (Eds.), *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence*: IGI Global.
- von Neumann, J., Metropolis, N., & Ulam, S. (1951). Monte Carlo Method. *National Bureau of Standards/Applied Math. Series, 12*, 36-38.
- Weiss, S. M., Kulikowski, C. A., Amarel, S., & Safir, A. (1978). A model-based method for computer-aided medical decision-making. *Artificial Intelligence, 11*(1), 145-172. DOI: 10.1016/0004-3702(78)90015-2
- Westermann, G., & Mareschal, D. (2009). Modelling the Transition from Perceptual to Conceptual Organization. In J. Mayor, N. Ruh, & K. Plunkett (Eds.), *Connectionist Models of Behaviour and Cognition*, Progress in Neural Processing (pp. 153–164).
- Westermann, G., Sirois, S., Shultz, T. R., & Mareschal, D. (2006). Modeling developmental cognitive neuroscience. *Trends in Cognitive Sciences, 10*(5), 227-232. DOI: 10.1016/j.tics.2006.03.009
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology, 47*(5), 1220-1229. DOI: 10.1037/a0024023
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272. DOI: 10.1037/0033-295X.114.2.245
- Younger, B. A., & Cohen, L. B. (1983). Infant Perception of Correlations among Attributes. *Child Development, 54*(4), 858-867. DOI: 10.2307/1129890

Yu, C., Smith, L. B., Klein, K. A., & Shiffrin, R. M. (2007). Hypothesis testing and associative learning in cross-statistical word learning: Are they one and the same? *Proceedings of the 29th Annual Meeting of the Cognitive Science Society*.

Running head: A DFT MODEL OF CHILDREN'S CATEGORY LABEL LEARNING

All Things Considered: Dynamic Field Theory Captures the Effect of
Category Variability on Young Children's Word Learning

Katherine E. Twomey & Jessica S. Horst

University of Sussex

Author note

A version of this paper is published as:

Twomey, K. E. & Horst, J. S. (2011). All things considered: Dynamic Field Theory captures effect of categories on children's word learning. *Proceedings of the 2011 Society of Artificial Intelligence and the Simulation of Behaviour Convention*

Abstract

Recent research demonstrates that within-category variability profoundly influences young children's word learning and categorisation (Twomey & Horst, Paper 1, this thesis). This paper describes a Dynamic Field Theory (DFT) neural network simulation of these data. The model was introduced to multiple category exemplars that were either moderately variable (*narrow* condition) or highly variable (*broad* condition). The model captured the empirical data. Like the children, the model was better able to learn category labels in the *narrow* condition and better able to extend category labels in the *broad* condition. Novel insights into the cognitive processes underlying children's ability to map novel words to novel objects are discussed. Overall, these findings form the groundwork for future empirical research and add weight to dynamic-associative accounts of categorisation, word learning, and cognitive development, more generally.

All Things Considered: Dynamic Field Theory Captures the Effect of
Category Variability on Young Children's Word Learning

Young children learn to categorise their environment remarkably easily, despite its well-documented complexity (Mervis & Rosch, 1981). Children start using categories to make sense of their world from as early as three months (Quinn, Eimas, & Rosenkrantz, 1993), and by 12 months children begin to learn labels for their categories (Fenson, et al., 1993). However, word learning is no trivial task: as Quine (1960) famously noted, children must determine the referent of a label from a seemingly infinite array of candidates. Nonetheless, children rapidly and reliably associate novel labels with novel referents ("fast mapping"; Carey & Bartlett, 1978; Gershkoff-Stowe & Hahn, 2007; Heibeck & Markman, 1987). However, fast mapping is only the beginning of the word learning process (Munro, Baker, McGregor, Docking, & Arciuli, 2012). Full word learning requires repeated, cross-situational exposures, during which label-category associations are strengthened (Smith & Yu, 2008), such that children can use labels in new contexts or after a delay (Horst & Samuelson, 2008; Kucker & Samuelson, 2011; Smith & Yu, 2008; Twomey & Horst, Paper 1, this thesis).

New insights into the word learning puzzle

Given this complexity, how do children learn to categorise and label with such proficiency? Despite decades of behavioral research and numerous proposed explanations (e.g., Gentner, 1983; Mandler, Fivush, & Reznick, 1987; Nelson, 1973; Rakison & Lupyan, 2008; Spelke & Kinzler, 2007), the answer is not straightforward. Indeed, because empirical studies cannot reveal underlying cognitive processes, the same datasets have been presented as evidence for conflicting accounts of word learning (Booth & Ware, 2010; Diesendruck & Graham, 2010; Samuelson & Perone, 2010). Recently, however, neural network models, with their explicitly-defined associative

mechanisms, have offered new insight into children's label generalisations (Samuelson, 2002; Samuelson & Horst, 2007) and category development (Mareschal, French, & Quinn, 2000; Westermann & Mareschal, 2009). Nonetheless, the relationship between in-task experience, categorisation and word learning remains unclear.

Research in domains as diverse as motor development (Thelen & Spencer, 1998), phonological acquisition (Rost & McMurray, 2009), and visual categorisation (Younger & Cohen, 1983) as well as word learning (Mather & Plunkett, 2009; Perry, Samuelson, Malloy, & Schiffer, 2010) suggests that multiple, variable experiences facilitate learning. Recently, Twomey & Horst (Paper 1, this thesis) demonstrated that variability between category exemplars strongly influences word learning. However, as with any empirical study, the data only tell us so much, and further work is required to elucidate the cognitive mechanisms underlying the relationship between word learning and category variability. The following section briefly summarises the empirical study (see also Twomey & Horst, Paper 1, this thesis), which were simulated using the neural network model described below (see section *Dynamic Neural Field Simulation*)

Supporting Empirical Data

Method

Participants. Twenty-four typically developing, monolingual, English speaking 30-month-old children participated. Children were randomly assigned to either the *narrow* or *broad* condition.

Stimuli. Known stimuli for all conditions consisted of 18 toy objects known to 30-month-old children. Novel stimuli in each condition consisted of 12 toy objects unfamiliar to 30-month-old children. Examples of novel stimuli are depicted in Figure 1.

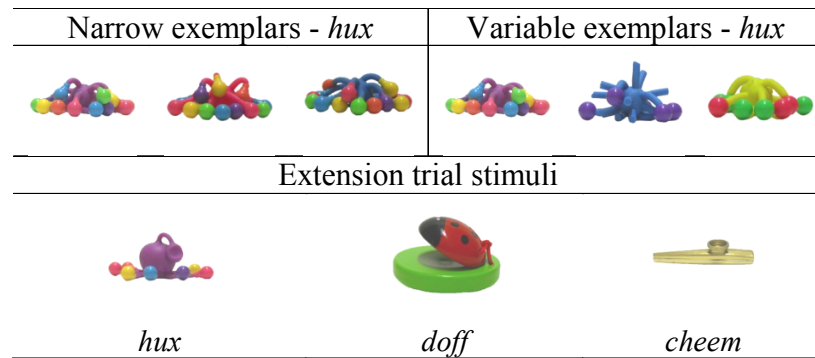


Figure 1. Example novel stimuli used in Twomey & Horst (Paper 1, this thesis)

Procedure and design. The experiment consisted of two phases: referent selection (18 trials) and test (six trials). During referent selection children encountered three novel categories (see Fig. 1). On each referent selection trial children saw an array of three toys (two known, one novel) and were asked to point to one of the objects (e.g., “Can you show me the *hux*?”). Each category was encountered in the context of one novel label (*i.e.*, *hux*, *doff*, or *cheem*), and each novel word was heard three times. Overall, children saw six trials for each category (three known, three novel) and received nine known and nine novel label trials (three per category for both trial types). Children either saw novel categories with moderately variable exemplars (*narrow* condition) or with highly variable exemplars (*broad* condition; see Figure 1).

After a five-minute delay the test phase began. On each retention trial children saw an array of three previously-seen toys (one from each novel category) and were asked to point to one of the objects. Extension trials immediately followed and were identical to retention trials except that new items from the novel category were shown (see Figure 1).

Results

Referent selection. Results are depicted in the left panel of Figure 2. All children succeeded during referent selection. Children in both conditions chose the target object at levels significantly greater than chance (0.33) on both known label trials

(*narrow*: $t(11) = 10.56, p < .0001, d = 3.05$; *broad*: $t(11) = 17.51, p < .0001, d = 5.05$) and novel label trials (*narrow*: $t(11) = 5.99, p < .0001, d = 1.7$; *broad*: $t(11) = 15.67, p < .0001, d = 4.52$). Unpaired t -tests revealed no difference between conditions for either known or novel referent selection trials (known: $t(22) = 0.28, ns.$; novel: $t(22) = 0.63, ns.$). Within-category variability therefore did not affect referent selection.

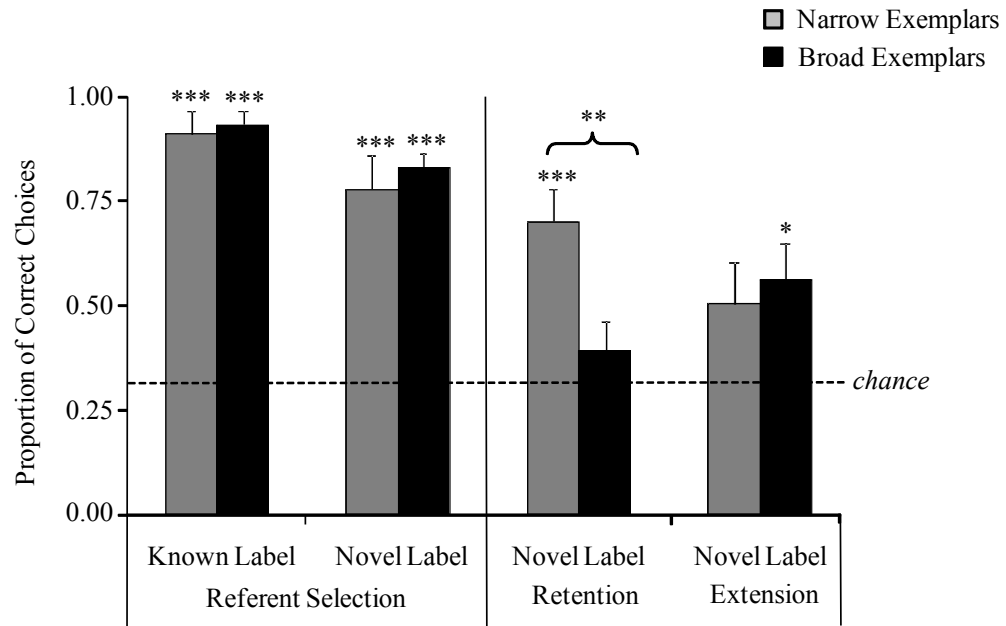


Figure 2. Experimental results: Children's proportion of correct choices. Dotted line represents chance (.33). Error bars represent one standard error. *** $p < .001$, ** $p < .01$, * $p < .05$.

Test trials. Results are depicted in the right panel of Figure 2. Only children in the *narrow* condition retained novel labels at levels significantly above chance (0.33), $t(11) = 4.76, p < .001, d = 1.38$. These children retained novel labels better than children in the *broad* condition, $t(22) = 2.98, p < .01, \eta^2 = 0.29$. In contrast, only children in the *broad* condition extended the novel labels at levels greater than chance, $t(11) = 2.63, p < .05, d = 0.76$; however no difference between groups was found for extension trials.

Discussion. Only children in the *narrow* condition retained novel category labels; however, these children did not extend these labels to completely novel category exemplars. In contrast, children in the *broad* condition did not retain novel category labels, but did extend them to completely novel category exemplars. Thus, moderate – but not high – variability facilitates retention of novel category labels, and high – but not moderate – variability facilitates extension (cf. Perry, et al., 2010; Quinn, et al., 1993).

These empirical results pose several questions. First, what real-time processes underlie children’s ability to infer the correct referent of the novel labels in this task? Second, why does low variability help retention but hinder extension? Finally, why does high variability hinder retention but help extension? We address these questions using a neural network model of children’s behaviour in the empirical task. The model demonstrates that apparently complex reasoning can emerge from the low-level associative processes that drive learning in our model. Further, the use of a computational simulation allows us to examine category formation moment-by-moment, shedding light on the interplay between variability and categorisation.

Dynamic Neural Field Simulation

Dynamic Field Theory (DFT) is a formalisation of Dynamic Systems Theory (DST; Thelen & Smith, 1996) which has successfully captured data from both motor and perceptual tasks (Samuelson & Horst, 2008; Simmering, Schutte, & Spencer, 2008). According to DST, stable behaviours self-organise from interactions between the body and the physical environment taking place within nested timescales. Cognition and sensorimotor input are inextricably coupled and embedded in real-time environmental input, as well as just-past experience and longer-term learning history. Consequently, according to DST, “not all knowledge must be stored in the brain” (Spencer & Schöner,

2003, p. 406). DST has been applied in a variety of domains to parsimoniously explain hitherto puzzling phenomena; for example, the sudden disappearance of young children's stepping reflex, perseverative reaching in A-not-B tasks (Thelen & Ulrich, 1991) and U-shaped behavior in goal-directed reaching (Thelen & Spencer, 1998).

Dynamic Field Theory (DFT) is a mathematical dynamical systems framework in which stable behaviours self-organise in-the-moment (Kelso, Ding, & Schöner, 1993; Spencer & Schöner, 2003). A class of neural network with neurological plausibility and subsymbolic representation at heart, DFT has much in common with connectionism (Spencer, Thomas, & McClelland, 2009). However, unlike connectionism, which focuses on simulating stimulus-response behaviour (McLeod, Plunkett, & Rolls, 1998) by iteratively updating association strengths between discrete processing units (Amari, 1977), DFT focuses on developmental change as mediated *in time* by attractor states (that is, points of behavioural stability) and instabilities (that is, points of behavioural change).

DFT is implemented computationally in the Dynamic Neural Field (DNF). At the mathematical level, DNFs model continuous neural populations and explicitly represent their input metrics (e.g., if a DNF represents colour input, “pink” will be nearer “red” on the input axis than will “green”). Further, neuronal activation is updated continuously and asynchronously rather than iteratively and synchronously, reflecting the asynchronous update of neurons in the brain (Schneegans & Schöner, 2008). Thus, DNFs simulate online cognitive processes, from which representations emerge in real-time. DNF models have successfully captured experimental data from looking tasks (Perone, Spencer, & Schöner, 2007), dimensional change card-sorting tasks (Buss & Spencer, 2008), spatial recall tasks (Lipinski, Simmering, Johnson, & Spencer, 2010) and novel noun generalisation tasks (Samuelson & Horst, 2007).

The goal of this simulation was to replicate data from Twomey & Horst (E2a; Paper 1, this thesis) to examine the effect of exemplar variability on retention and extension of novel names. We extend Faubel & Schöner's (2008) embodied DNF model of object recognition to a word learning context and demonstrate that the apparently complex reasoning required for word learning via fast mapping can in fact be accounted for by low-level dynamic-associative mechanisms.

Architecture. Activation in DNFs is governed by local excitation, lateral inhibition and global inhibition. Thus, locations close to an activated location also become activated via local excitation, while activations at more distant locations is suppressed via lateral inhibition. These interactions generate localised, self-sustaining peaks of activation (Spencer, Simmering, Schutte, & Schöner, 2007). Architecture of the model is depicted in Figure 3. The current model consists of two 2-dimensional layers; specifically, an input DNF, which receives inputs representing labels and objects, coupled reciprocally to a Hebbian layer, which stores slow-decaying activation traces. Activation in the perceptual layer is generated by input along the label and object axes. Activation is governed by the general equation below:

$$(1) \tau u_{o,l}(x,t) = -u(x,t) + h + S(x,t) + \int w(x-x')\sigma(u(x',t))dx'$$

where $u_{o,l}(x,t)$ is the activation level along the object (o) and label (l) dimensions at location x , as a function of time (t), mediated by the timescale of the dynamics, τ .

Current activation in the perceptual layer, $-u(x,t)$, receives external input, $S(x,t)$, and is subject to excitatory and inhibitory interaction defined by a Gaussian kernel with weight w , and width σ . The resting level of the system is defined by h .

The formation of a peak in the perceptual layer represents an association between given locations along the object and label axes; that is, the child's decision to map a label to an object. Importantly, neighbouring locations are mutually excitatory,

whereas more distant locations are mutually inhibitory. Peaks leave corresponding activation traces in the Hebbian layer, which in turn facilitate future object-label mappings at this location.

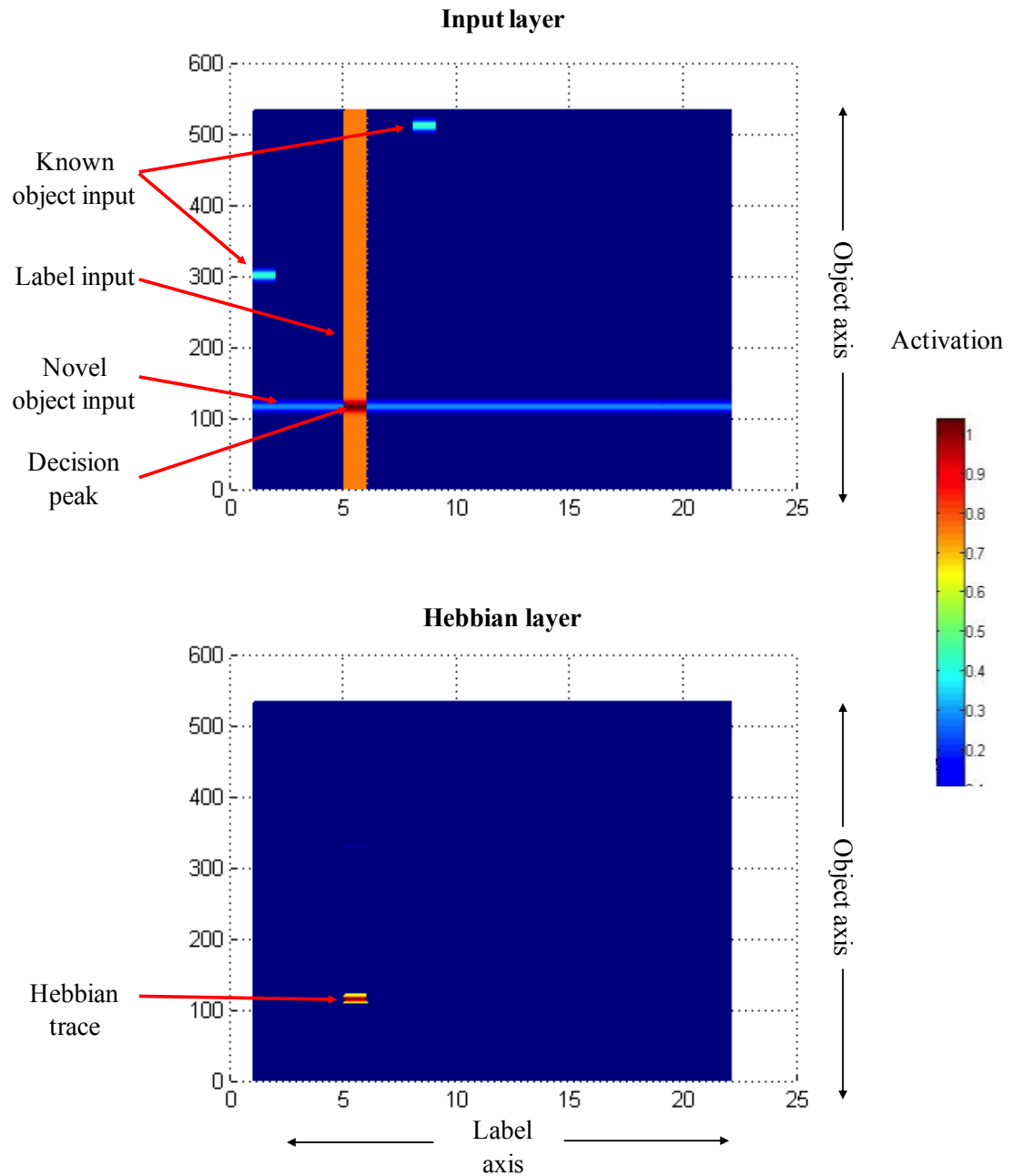


Figure 3. Architecture of the DNF model. Depicts input and Hebbian layers after a single novel referent selection trial. NB: “Perceptual layer” panel depicts total input for a single trial

Stimuli and procedure. Variability in input along the object axis reflects variability in category structure in the experimental stimuli. Specifically, the model is either presented with *narrow* or *broad* categories. *Narrow* category input consists of a central exemplar at a given location along the object axis, and two further inputs at nearby locations. *Broad* category input consists of the same central exemplar and two inputs at more distant locations.

Importantly, to ensure that the model stimuli accurately reflected the novel stimuli used by Twomey & Horst (Paper 1, this thesis), we asked 18 adults to rate the each category of novel exemplars from the empirical task for within-category similarity. In the empirical task, one “central” exemplar from each category appeared in both the *narrow* and *broad* conditions. Participants used a Likert scale (1 = highly similar, 11 = highly dissimilar) to rate stimuli from each category relative to the “central” exemplar from that category.

These ratings generated the input values for the object stimuli presented to the model. Central exemplars are located equidistantly along the object axis (total length = 522 units) in order to avoid biasing the model’s categorisation. Thus, central exemplars are positioned at locations 115, 265 and 415. The location of the first additional novel exemplar was determined by subtracting the mean similarity rating for one of the remaining objects to the value of the central exemplar. The location of the second additional novel exemplar was determined by adding the mean similarity rating for the remaining object from the value of the central exemplar. For example, one *narrow* category consists of a central input at location 115_{object} , with the other two exemplars located at 112_{object} (i.e., $115 - 3$) and 117_{object} (i.e., $115 + 2$), while the corresponding *broad* category consists of a central input at location 115_{object} , with the other two exemplars located at 107_{object} (i.e., $115 - 8$) and 122_{object} (i.e., $115 + 9$).

Thus, for all trials, inputs reflect the experimental design employed by Twomey & Horst (E2a; Paper 1, this thesis). Number of trials and trial order are identical to the empirical task. The model was run 12 times per condition.

Referent selection. On each referent selection trial the model receives two “known” inputs and one “novel” input. Known inputs consist of a Gaussian hump of activation at predetermined locations along the object and label axes, reflecting the fact that children come to the experiment already having learned the names of the known objects. For example, a known input at location $(80_{object}, 3_{label})$ represents a learned association between location 80 along the object dimension and location 3 along the label dimension. From an empirical perspective, this input might represent a learned association between the label “car” and the object CAR. Simultaneously, the model receives a ridge of novel input at a specific location along the object axis but generic along the label axis (see Figure 3). For example, a ridge of input at (115_{object}) represents a specific novel object (as depicted in Figure 1) but could correspond to any label. These inputs are presented to the model for the equivalent of three seconds. This reflects the pause at the beginning of each trial during which children could look at, but not interact with, the objects, and allows self-stabilising peaks of activation to form at the known object locations.

Next, the model is presented with a label via a ridge of input along the label axis which intersects either with one of the known object inputs (reflecting “Can you show me the *car*?”) or with the novel object input (reflecting “Can you show me the *hux*?”).

During referent selection the model encounters three novel categories over 18 referent selection trials. Formation of a peak at any location is taken as the model’s response to the label stimulus.

Test phase. After 11 “empty” trials corresponding to the five-minute delay in the experiment⁹, in which no stimuli are presented, the test trials begin. Object inputs consist of three generic ridges of activation at the previously encountered novel object locations along the object axis. The model receives label input in the same manner as in referent selection. The three subsequent extension trials are identical to retention trials except that the initial novel object inputs are given at locations close to but not identical to previously encountered locations. Thus, during extension trials the model associates novel labels with completely new novel objects.

Results. Simulation data are depicted in Figure 4. The model is very accurate on referent selection trials, for both known referent selection (*narrow*: $t(11) = 24.28, p < .001$; *broad*: $t(11) = 14.37, p < .001, d = 8.67$) and novel referent selection (*narrow*: $t(11) = 24.28, p < .001$; *broad*: $t(11) = 32.55, p < .001, d = 14.64$). When presented with *narrow* categories, the model correctly associated novel category exemplars with previously-encountered novel labels on retention trials, $t(11) = 3.47, p < .01, d = 2.09$. However, on extension trials the model did not associate completely novel exemplars with previously-encountered labels, $t(11) = 1.04, ns., d = 0.62$. In contrast, when presented with *broad* categories, the model did not associate novel exemplars with the appropriate label on retention trials, $t(11) = 1.80, ns., d = 1.09$. Finally, on extension trials the model associated completely novel exemplars with previously-encountered labels only when presented with *broad* categories $t(11) = 2.63, p < .05, d = 1.59$ ¹⁰.

⁹ 11 trials corresponded approximately to the overall proportion of the experiment made up of the 5-minute break and allowed activation in the Hebbian field to decay. This figure generated the strongest results, however smaller numbers of “empty” trials yielded the same overall pattern.

¹⁰ Simulation data are statistically equivalent to the empirical data, as confirmed by a mixed ANOVA with trial type (known, novel, retention, extension) as a repeated measure and Data Type (empirical, simulation) and Condition (narrow, broad) as between-subjects factors. No significant main effect or interaction involving Data Type was revealed.

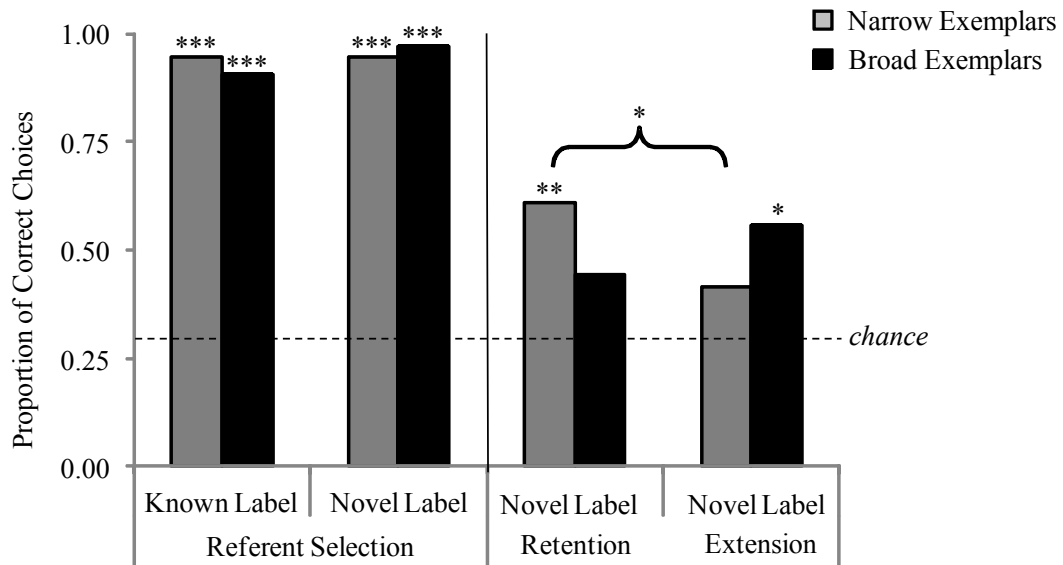


Figure 3. Model results: proportion of correct responses. Dotted line represents chance (.33). *** $p < .001$, ** $p < .01$, * $p < .05$.

Discussion. Empirical data provided by Twomey & Horst (Paper 1, E2a, this thesis) were replicated using a 2-layer DFT model with stimuli closely based on those used in the original task and an identical experimental design.

How, then, do these computational data address the questions posed by the empirical study? First, children’s remarkable ability to correctly infer the referent of a novel word when presented with an array of several known objects and a single novel object has previously been explained with constraint-based accounts. For example, children may use a process of elimination to reason that because the two known objects have names, the novel word must refer to the novel object (“dysjunctive syllogism” or “mutual exclusivity”, e.g., Halberda, 2006). Alternatively, children may assume that novel names refer to novel objects, without paying attention to any known objects in the array (the “Novel Name-Nameless Category” or N3C principle, Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992). Clearly, the model does not come pre-programmed with any high-level reasoning ability, or *a priori* lexical principles, yet it succeeds in accurately

mapping novel labels to novel referents. But without any such higher-level biases, how does the model succeed in the referent selection task?

Recall that activation in the DNF is governed by local excitation and lateral inhibition. Therefore, peaks at a given location (x_{object}, y_{label}) will inhibit the formation of new peaks at other locations $(x \pm a_{object}, y_{label})$ or $(x_{object}, y \pm a_{label})$. It is this mechanism, which supports fast mapping via “mutual exclusivity”. Referring again to Figure 3, a trial might begin with known object peaks at $(300_{object}, 1_{label})$ and $(510_{object}, 8_{label})$ and a novel object at (115_{object}) . A novel label ridge is presented at (5_{label}) which intersects with every location along the object dimension. At this point in the trial, if activation along the object axis were equal at all points, then a new peak could form at the intersection between the novel label and any object location. However, the known peaks suppress formation of new peaks at the intersection between the novel label and the known object locations. That is, the known object peak at $(300_{object}, 1_{label})$ inhibits formation of a new peak at $(300_{object}, 5_{label})$. Correspondingly, the known object peak at $(510_{object}, 8_{label})$ inhibits formation of a new peak at $(510_{object}, 5_{label})$.

The only location along the object dimension with above-threshold activation which is not already associated with a novel label is the location corresponding to the novel object, $(115_{object}, 5_{label})$. The most likely location for a peak to form, then, is at the intersection of the novel object and the novel label (but note that due to the noise parameter, the model, in line with children, is not 100% accurate in matching novel objects to novel labels). In the current model, then, behaviour that some have argued is contingent on complex metacognitive reasoning (Halberda, 2006; Markman, 1994; Markman, Wasow & Hansen, 2003) emerges in fact from simple associative processes interacting with longer-term learning history (known label-object associations), in-task learning, and online input. This also suggests that other neural networks with similar

excitatory and inhibitory mechanisms should also be able to achieve fast mapping via mutual exclusivity without recourse to higher-level reasoning (see Twomey, Horst & Morse, Paper 6, this thesis).

Second and third, the model also sheds light on how exemplar variability influences both retention and extension. There is a direct relationship between perceptual similarity and category breadth (as demonstrated in the categorisation literature, e.g., Quinn, 1993), which gives rise to differences in label extensions dependent on perceptual similarity of novel exemplars to just-seen exemplars. This prediction derives from the structure of the memory trace for novel objects formed during referent selection. Figure 4 depicts an example memory trace for the “265” novel object category after the referent selection phase – that is, the after the model has encountered six exemplars from that category – for both the *narrow* condition (left panel) and the *broad* condition (right panel). As can be seen in the figure, the memory traces for two conditions differ in activation strength and range.

Specifically, the *narrow* memory trace has greater maximum activation (0.20) than the broad memory trace (0.12). At retention, the model is presented with a previously-seen novel object input at location 265. On presentation of a label (e.g., at location 11 on the label dimension) the memory trace in the *narrow* condition is sufficient to facilitate the formation of a peak at $(265_{object}, 11_{label})$, as one would expect if the label were learned. The memory trace in the *broad* condition, however, is not sufficient to facilitate correct peak formation.

In contrast, the *narrow* memory trace has a smaller range (18 units, from location 256 to location 274) than the *broad* memory trace (31 units, from location 248 to location 279). At extension, the model is presented with a completely novel object input at location 274. In the *narrow* condition, location 274 is at the edge of the

memory trace – with zero stored activation. On presentation of a label (again at location 11), there is insufficient activation in the *narrow* memory trace to facilitate formation of a peak at the correct location of $(274_{object}, 11_{label})$. In the *broad* condition, however, location 274 falls within the memory trace, providing sufficient activation to facilitate formation of a peak at the correct location. Thus, the model predicts that even small differences in relative exemplar similarity will determine whether or not children extend the label to the novel exemplar.

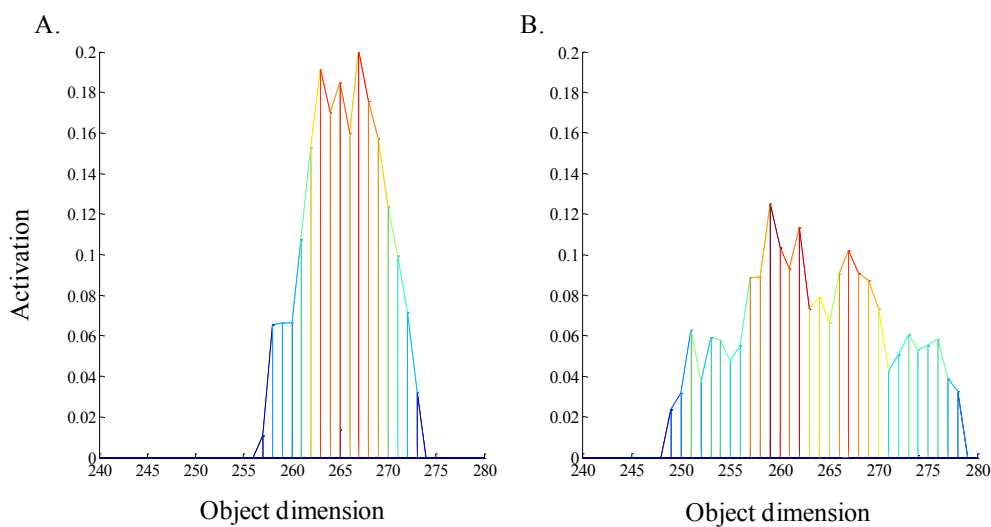


Figure 4. Activation in the Hebbian layer after 18 referent selection trials. Panel A depicts the *narrow* condition, and Panel B depicts the *broad* condition.

General Discussion

The model presented here successfully captures data from an empirical word learning task using simple, associative mechanisms. However, before we generalise these mechanisms from word learning to the wider context of cognitive development, future work must focus on testing the model's novel predictions (see Twomey & Horst, Paper 5, this thesis). Nonetheless, the current data constitute the first DFT model of fast mapping via mutual exclusivity and as such provide proof-of-concept for the application of DFT to fast mapping paradigms. Further, the data point firmly away from accounts of categorisation, word learning and cognitive development more generally which posit

fixed or specialised innate structure (e.g., Diesendruck & Bloom, 2003; Carey, 2011; Hauser, Chomsky & Fitch, 2002; Mandler, 2012). Rather, this simulation supports a view of word learning and categorisation as flexible, emerging from simple associations made over time between label and object, in the context of cognition developing bottom-up from the dynamic coupling between brain, environment, and time.

Acknowledgements

This research was supported by ESRC Quota Award ES/H014942/1 to K.E.T and an ESRC Overseas Institutional Visit Grant to K.E.T. We would like to thank Prof. John Spencer, Dr. Larissa Samuelson, and all members of the DelTA Center, University of Iowa, USA, for guidance and useful discussions in building the DFT model.

References

- Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2), 77-87. DOI: 10.1007/BF00337259
- Booth, A. E., & Ware, E. A. (2010). Categories, concepts and causality: A reply to Samuelson and Perone. *Cognitive Development*, 25(2), 154-157. DOI: 10.1016/j.cogdev.2010.02.004
- Buss, A. T., & Spencer, J. P. (2008, July). *The emergence of rule-use: a DNF model of the DCCS*. Paper presented at the 30th Annual Meeting of the Cognitive Science Society, Washington, DC.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & A. Miller (Eds.), *Linguistic Theory and Psychological Reality* (pp. 264–293). Cambridge, MA: MIT Press.
- Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Dev.*, 74(1), 168–178. DOI: 10.1111/1467-8624.00528
- Diesendruck, G., & Graham, S. A. (2010). Kind matters: a reply to Samuelson & Perone. *Cognitive Development*, 25, 149-153. DOI: 10.1016/j.cogdev.2010.02.003
- Faubel, C., & Schöner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks*, 21(4), 562-576. DOI: 10.1016/j.neunet.2008.03.007. DOI: 10.1016/j.neunet.2008.03.007
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular Publishing Group.

- Gentner, D. (1983). Structure-mapping - A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170. DOI: 10.1207/s15516709cog0702_3
- Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech Language and Hearing Research*, 50(3), 682-697. DOI: 10.1044/1092-4388(2007/048)
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28(1), 99-108. DOI: 10.1037/0012-1649.28.1.99
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), 310-344. DOI: S0010-0285(06)00025-9 [pii] 10.1016/j.cogpsych.2006.04.003
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579. DOI: 10.1126/science.298.5598.1569
- Heibeck, T. H., & Markman, E. M. (1987). Word learning in children - an examination of fast mapping. *Child Development*, 58(4), 1021-1034. DOI: 10.2307/1130543
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI: 10.1080/15250000701795598
- Kelso, J., Ding, M., & Schöner, G. (1993). Dynamic pattern formation: A primer.
- Kucker, S. C., & Samuelson, L. K. (2011). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*.
- Lipinski, J., Simmering, V. R., Johnson, J. S., & Spencer, J. P. (2010). The role of experience in location estimation: Target distributions shift location memory biases. *Cognition*, 115(1), 147-153. DOI 10.1016/j.cognition.2009.12.008

- Mandler, J. M. (2012). On the Spatial Foundations of the Conceptual System and Its Enrichment. *Cognitive Science*.
- Mandler, J. M., Fivush, R., & Reznick, J. S. (1987). The development of contextual categories. *Cognitive Development*, 2(4), 339-354. DOI: 10.1111/j.1551-6709.2012.01241.x
- Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, 36(5), 635. DOI: 10.1037/0012-1649.36.5.635
- Markman, E. M. (1994). Constraints on Word Meaning in Early Language-Acquisition. *Lingua*, 92(1-4), 199–227. DOI: 10.1016/0024-3841(94)90342-5
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47(3), 241–275. DOI: 10.1016/S0010-0285(03)00034-3
- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, 14(1), 60-76. DOI: 10.1080/15250000802569702
- McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford: Oxford University Press.
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32(1), 89-115.
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3. DOI: 10.3389/fpsyg.2012.00041
- Nelson, K. (1973). Some evidence for the cognitive primacy of categorization and its functional basis. *Merrill-Palmer Quarterly*, 19(1), 21-39.

- Perone, S., Spencer, J., & Schöner, G. (2007). *A dynamic field theory of visual recognition in infant looking tasks*. Paper presented at the Cognitive Science Society.
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, 21(12), 1894-1902. DOI: 10.1177/0956797610389189
- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge: MIT Press.
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually aimilar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22(4), 463-475. DOI: 10.1068/p220463
- Rakison, D. H., & Lupyan, G. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development*, 73(1), 1-110.
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339-349. DOI: 10.1111/j.1467-7687.2008.00786.x
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016-1037. DOI: 10.1037//0012-1649.38.6.1016. DOI: 10.1037//0012-1649.38.6.1016
- Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: Moment-to-moment interactions shape children's naming biases. *Infancy*, 11(1), 97-110. DOI: 10.1207/s15327078in1101_5

- Samuelson, L. K., & Horst, J. S. (2008). Confronting complexity: insights from the details of behavior over multiple timescales. *Developmental Science*, 11(2), 209-215. doi: DOI 10.1111/j.1467-7687.2007.00667.x. DOI: 0.1111/j.1467-7687.2007.00667.x
- Samuelson, L. K., & Perone, S. (2010). Rethinking conceptually-based inference - grounding representation in task and behavioral dynamics: Commentary on "Fifteen-month-old infants attend to shape over other perceptual properties in an induction task," by S. Graham and G. Diesendruck, and "Form follows function: Learning about function helps children learn about shape," by E. Ware and A. Booth. *Cognitive Development*, 25(2), 138-148. DOI: 10.1016/j.cogdev.2010.02.002
- Schneegans, S., & Schöner, G. (2008). Dynamic field theory as a framework for understanding embodied cognition. *Handbook of Cognitive Science: An embodied approach*, 241–271.
- Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, 1202, 68-86. DOI: 10.1016/j.brainres.2007.06.081
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558-1568. DOI: 10.1016/j.cognition.2007.06.010
- Spelke, E. S., & Kinzler. (2007). Core knowledge. *American Psychologist*, 55(11), 1233-1243. DOI: 10.1037/0003-066X.55.11.1233
- Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, 6(4), 392-412. DOI: 10.1111/1467-7687.00295

- Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioural development? Insights from a dynamic field theory of spatial cognition. In J. M. Plumert & J. P. Spencer (Eds.), *The Emerging Spatial Mind*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780195189223.003.0014
- Spencer, J. P., Thomas, M., & McClelland, J. (2009). *Toward A Unified Theory of Development: Connectionism and Dynamic Systems Theory Reconsidered*. New York, NY: Oxford University Press. DOI: 10.1093/acprof:oso/9780195300598.001.0001
- Thelen, E., & Smith, L. B. (1996). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.
- Thelen, E., & Spencer, J. P. (1998). Postural control during reaching in young infants: A dynamic systems approach. *Neuroscience and Biobehavioral Reviews*, 22(4), 507-514. DOI: 10.1016/S0149-7634(97)00037-7
- Thelen, E., & Ulrich, B. D. (1991). Hidden skills - a dynamic-systems analysis of treadmill stepping during the 1st year. *Monographs of the Society for Research in Child Development*, 56(1). DOI: 10.2307/1166099
- Twomey, K. E & Horst, J. S. (paper 1, this thesis). That's more like it: More exemplars facilitate word learning.
- Twomey, K. E. & Horst, J. S. (paper 5, this theses). Testing a Dynamic Neural Field model of children's category labelling.
- Westermann, G., & Mareschal, D. (2009). Modelling the Transition from Perceptual to Conceptual Organization. In J. Mayor, N. Ruh & K. Plunkett (Eds.), *Connectionist Models of Behaviour and Cognition* (Vol. 18, pp. 153-164).

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, 54(4), 858-867. DOI: 10.2307/1129890

Running Head: TESTING A DNF MODEL OF CATEGORISATION

Testing a Dynamic Neural Field Model of Children's Category Labelling

Katherine E. Twomey & Jessica S. Horst

University of Sussex

Author note

A version of this paper is in preparation, to appear in:

Twomey, K. E. & Horst, J. S. (2013). Testing a Dynamic Neural Field model of children's category labelling. *Proceedings of the 13th Neural Computation and Psychology Workshop, Basque Center on Cognition, Brain and Language, San Sebastian, Spain.*

Abstract

Despite the complexity of their environment, children spontaneously learn to categorise and label. Empirical research has created debate as to whether children achieve this through high-level reasoning or low-level associative processes. Recently, computational simulations of cognitive development have generated novel insight into categorisation and labelling. Here we present an empirical test of a Dynamic Neural Field (DNF) model of categorisation. The model predicted that children will extend novel labels to new category-central exemplars, but not to new category-peripheral exemplars. In the empirical task, 44 30-month-old children were asked to fast map novel labels to three novel categories over 18 fast mapping trials. After a short delay, children were presented with retention trials with previously-encountered novel objects, and extension trials with completely novel which were either category central or category peripheral (Rosch, 1975). In line with the model, only children who saw category central exemplars extended novel names, and children who saw category peripheral exemplars did not. The empirical replication suggests that DNFs provide an informative model of infant categorisation and word learning. The current study adds further weight to the dynamic systems view of development as the emergent product of low-level, dynamic interactions between perception, and the environment.

Testing a Dynamic Neural Field Model of Children's Category Labelling

The puzzle of how young children learn to categorise and label objects is well-known. Countless empirical studies have provoked often impassioned debate as to the cognitive underpinnings of this impressive skill. Born into an enormously rich perceptual environment, from an early age children parse objects into categories and treat exemplars from a single category equivalently. By 18 months, children have begun to label these categories (e.g., Houston-Price, Plunkett, & Harris, 2005); that is, they can reliably infer the referent of a novel word despite the complexity of the space of potential referents (Quine, 1960). This ability to form a quick, initial hypothesis about a word's meaning is known as fast mapping (Carey & Bartlett, 1978). Several theoretical accounts of categorisation and word learning have been offered, from low-level associative learning (e.g., Smith, 2000) to *a priori* conceptual primitives (e.g., Carey, 2011), and recent research demonstrates that the two phenomena are intimately linked (Gliozzi, Mayor, Hu, & Plunkett, 2009; Plunkett, Hu, & Cohen, 2008).

Fast mapping and word learning have latterly been the focus of computational research, in which the potential underlying cognitive and/or neural structures are described mathematically. These simulations use schematised inputs designed to reflect a particular experimental environment or developmental stage, and process them using a mathematical algorithm intended to be a simplified representation of some cognitive operation. The output is taken to represent the outcome of the cognitive process in question, in the given environment. For example, the relationships between numerical input vectors might reflect, for example, the relationship between objects and labels in an experimental situation before a learning phase (e.g., Gliozzi, Mayor, Hu & Plunkett, 2009). Over repeated iterations of the model (the learning phase), these relationships change, such that the numerical output vectors may reflect, for example children's

developing associations between object categories and labels. Importantly, however, models must be tested, by generating novel predictions about behaviour given a new environment or task. Further, these predictions must also be replicated experimentally for findings to be generalised with any confidence. For example, if providing the model with different label stimuli alters the categories formed by the model during the learning phase, this prediction can be tested by rerunning the original experiment and changing the label stimuli to reflect those given to the model. If replication is successful, it is appropriate to conclude that the mechanisms driving the model reflect in some way the mechanisms driving cognition (see Colunga & Smith, 2005 for examples of successful empirical replications of model predictions).

One of the great benefits of computational models is that unlike in biological organisms cognitive structure is available for inspection: for example, modellers can watch associations between categories and words develop in-situ (that is, in the simulated neural/cognitive structure), and measure and compare these categories directly. Thus, when rigorously tested by empirical replication, computational models greatly add to our understanding of cognition generally, and categorisation and word learning specifically.

The current paper presents just such an empirical test of a model's explanation for a behaviour. Twomey & Horst (Paper 4, this thesis) describe a Dynamic Neural Field model (for a review, see Spencer, Thomas, & McClelland, 2009) which has successfully replicated data from an empirical study examining the effect of category variability on 30-month-old children's label learning. The current paper presents a novel prediction generated by the model (Simulation) and an empirical replication of that prediction (Experiment).

Simulation

Dynamic Neural Field models (DNFs). DNFs (Spencer & Schöner, 2003) are emergentist simulations of changes in neural activation in response to external stimuli. In contrast to their more widely-applied connectionist cousins (McClelland, et al., 2011), DNFs model neural structure and time continuously; all representations are therefore distributed across activity in the input layer and a Hebbian “memory” layer. They are also topologically functional, such that similar inputs are represented as close together along a given axis.

Dynamic Neural Field models consist of one or more input layers, which initially receive input in the form of a modeller-defined increase in activation at a certain location in the layers. These inputs represent neural responses to external stimuli. Over time, the dynamics of the DNF allow peaks of activation to emerge in the input field thanks to locally excitatory and laterally and globally inhibitory neural interactions; that is, activation spreads from a given location to its neighbours, whilst activation at more distant locations is suppressed. These peaks represent associations between stimuli.

The input layer is coupled reciprocally to its “memory” – known as a Hebbian layer – such that when a peak forms in the input field, activation spreads to the Hebbian layer, where it is stored and decays slowly. The term “Hebbian” thus derives from the fact that a memory trace corresponding to an association between features made in the input layer at time t facilitates the reappearance of that association at time $t + 1$ (Hebb, 1949; Munakata & Pfaffly, 2004). The Hebbian layer therefore performs a “fire together, wire together” function similar to connection weights in connectionist models, simulating learning over time and repeated experience.

Categorisation by correlated features. Existing empirical and behavioural research demonstrates that children may form categories based on co-occurrence of perceptual features, and that the “best” or most category-central exemplars are those that share most features with other category exemplars (Rosch, 1975). For example, in a series of seminal studies, Younger and colleagues demonstrated that infants become increasingly adept across development at perceiving feature correlations between stimuli (Younger & Cohen, 1983, 1986; Younger & Fearing, 1998; see also Quinn, Eimas & Rosenkrantz, 1993). More recently, connectionist models have simulated the developmental differentiation of children’s categories based on the assumption that categories are scaffolded from covariation of shared perceptual features (Rogers & McClelland, 2004). For example, because SPARROW and PIGEON share some features (e.g., wings, beak, feathers), these shared features exhibit coherent covariation, in that things with wings often also have beaks and feathers. Representations for exemplars with many coherently covarying features are close in representational space. Clusters of such items in representational space constitute categories. Taken together, these and other studies suggest that exemplars of a given category share perceptual features and are represented close together in representational space (see also, Sloutsky & Fisher, 2004).

That children can extend known category labels to novel exemplars is not in dispute (Diesendruck, Markson, & Bloom, 2003; Landau, Smith, & Jones, 1988; Samuelson & Horst, 2007; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). However, whether even perceptually very similar exemplars are category central or category peripheral seems likely to affect label extension. We therefore hypothesised that, given two perceptually-similar novel exemplars, both of which are potential members of a previously-encountered category, the model would extend labels to category central novel exemplars, but not to category-peripheral novel exemplars.

Method

Architecture. The model consists of a two-dimensional input layer, representing number of shared features on one axis and labels on the other, as depicted in Figure 1 (see also Twomey & Horst, paper 4, this thesis). Pictured this way, feature inputs are orthogonal to label inputs, and simultaneous label and feature inputs therefore intersect. The formation of a peak at this intersection represents an association between these two inputs. In behavioural terms, peaks represent an association between a label and an object. Peaks of activation in the input layer feed activation to a coupled Hebbian layer, which stores a slow-decaying “memory trace” of activation across trials. Thus, during the familiarisation phase, learning of associations between objects and labels is manifested in the slow-decaying Hebbian memory trace. Stored activation at a given location in the Hebbian layer supports formation of peaks at that location later during the simulation.

Stimuli. The simulation represents the category centrality of novel stimuli via their proximity along the feature axis. “Novel” object stimuli consist of input ridges along the feature axis, but generic along the label axis; that is, on the first presentation of a novel object, a peak could form at the intersection of that input and any location on the label axis, with equal probability, reflecting the fact that children in fast mapping tasks do not know the name of the novel objects they encounter. “Known” object stimuli consist of Gaussian humps of activation at locations in the input field representing children’s previously learned associations between known labels and known objects. Label stimuli consist of a ridge of activation along the label axis. Label stimuli could therefore be associated with any position along the feature axis. Locations for novel and known object stimuli, as well as labels are given in Table 1.

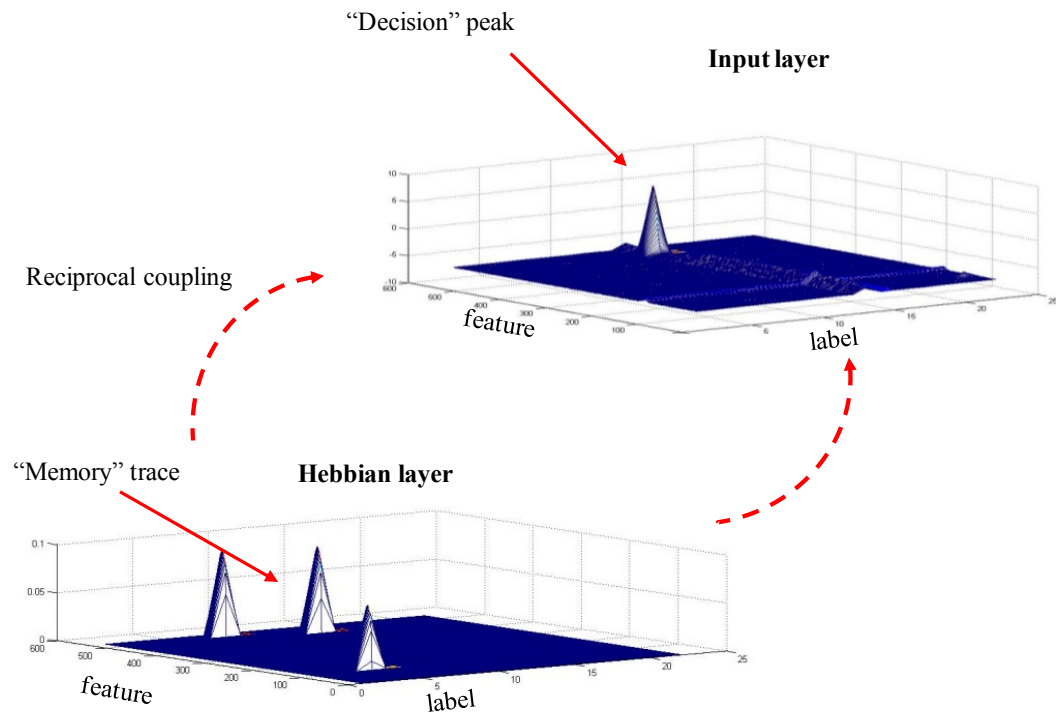


Figure 1. Architecture of the Dynamic Neural Field model. The input layer has locally excitatory and laterally and globally inhibitory dynamics, allowing activation “decision” peaks to emerge over time. Activation in peaks spreads to the Hebbian layer, where it acts as a slow-decaying memory trace.

Procedure and design. During the referent selection phase the model is familiarised with three novel categories (each consisting of three exemplars) and three novel labels, presented in blocks of six trials per category. Each block consists of three known and three novel trials. Each novel exemplar serves once as the target (during a novel trial), and once as a competitor (on a known trial). The model is therefore presented with a total of 18 referent selection trials.

A single referent selection trial consists of an initial presentation of two known object humps and a single novel object ridge. This reflects the general empirical procedure used to test fast mapping of labels to 3D objects, in which an array of two or more known objects and one novel object are presented to children simultaneously (e.g., Axelsson, Churchley, & Horst, 2012; Horst & Samuelson, 2008; Horst, Scott, & Pollard,

2010). Then, the model is given a label input ridge, reflecting the experimenter's request for the child to "get the [novel label]". This ridge intersects with a known object peak and/or the novel object ridge. A peak of activation may develop at one of these intersections, reflecting the child's choice, which may or may not be correct.

Sets of stimuli presented during Referent Selection						Extension			
Known object 1		Known object 2		Novel category exemplars		<i>Peripheral</i>		<i>Central</i>	
Object	Label	Object	Label	Object	Label	Object	Label	Object	Label
40	21	470	13	113	5				
30	1	510	8	115	5	119	5	114	5
450	16	230	7	117	5				
150	9	60	4	263	11				
80	18	360	2	265	11	269	11	264	11
380	15	170	14	267	11				
210	3	320	12	413	17				
340	20	20	6	415	17	419	17	414	17
490	19	190	10	419	17				

Table 1. Locations along feature and label axes of inputs to the model. Inputs representing category-central extension exemplars are closer to exemplars seen during referent selection than inputs representing category-peripheral extension exemplars.

Immediately following referent selection the model is given eleven trials in which no stimuli are presented, designed to reflect the five-minute delay between referent selection and test in the empirical task (see also, Twomey & Horst, paper 1, this

thesis). Then, the model receives three retention trials. These proceed in an identical manner to referent selection trials, except that object stimuli consist of three novel object ridges, one from each previously-encountered category. The model receives a different novel label on each retention trial. Thus, the model can only accurately respond if the memory trace associating novel objects to novel labels is sufficiently robust. Finally, the model is presented with three generalisation trials. At this stage, the model either receives inputs that reflect either completely novel, category-central exemplars or completely novel category-peripheral exemplars (see Table 1). The model was run 24 times per condition.

Results and Discussion

Results from the simulation are depicted in Figure 2. During referent selection the model mapped known labels to the correct referent at levels greater than expected by chance (0.33, all reported tests two-tailed), *central*: $t(23) = 56.46, p < .001, d = 11.44$; *peripheral* $t(23) = 49.52, p < .001, d = 10.09$. Similarly, the model mapped novel labels to the correct referents at above-chance levels, *central*: $t(23) = 28.58, p < .001, d = 5.81$, *peripheral*: $t(23) = 28.16, p < .001, d = 5.73$. At test, the model retained novel labels at above-chance levels in both conditions, *central*: $t(23) = 5.46, p < .001, d = 1.12$, *peripheral*: $t(23) = 4.76, p < .001, d = 0.97$ (note that no difference was anticipated between conditions for retention, as stimuli presented during referent selection and retention are identical across conditions). In contrast, however, the model extended novel labels in the *central* condition, $t(23) = 5.44, p < .001, d = 1.12$; but did not extend novel labels in the *peripheral* condition, $t(23) = -0.45, ns., d = -0.09$. A planned independent samples *t*-test confirmed a significant difference between conditions for extension, $t(46) = 4.17, p < .001, d = 1.23$. Thus, as predicted, the model extended novel

names only to objects that shared many shared features with the categories encountered during referent selection.

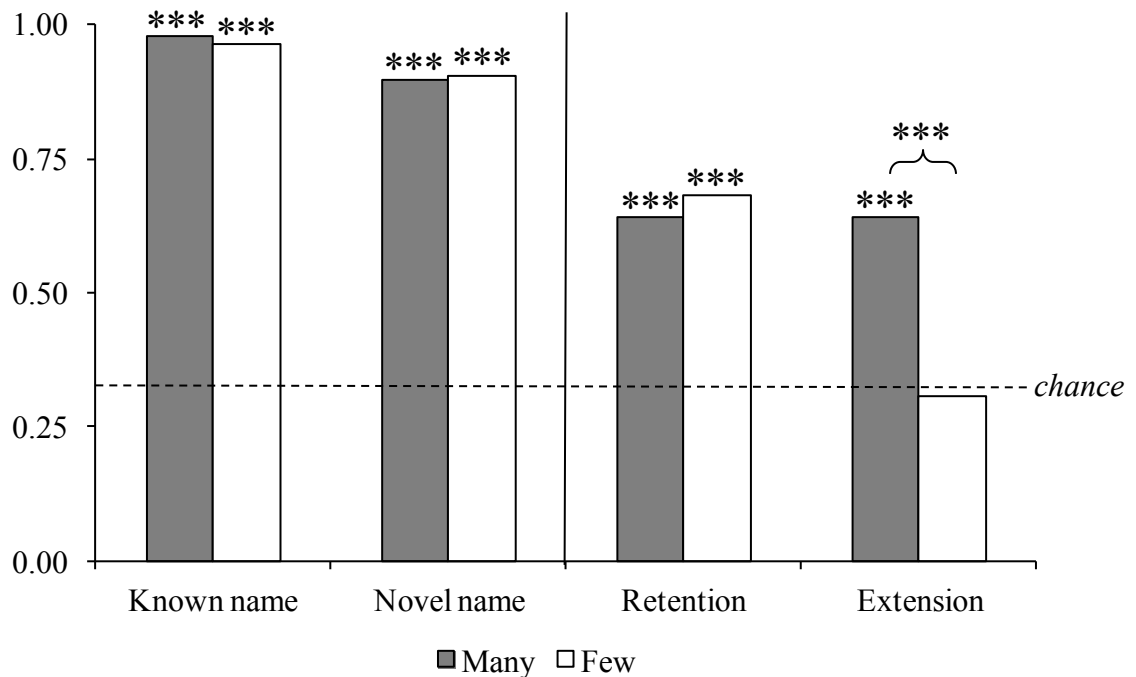


Figure 2. Simulation results. *** $p < .001$. Chance = 0.33, all tests two-tailed.

Empirical task

Using the same architecture and procedure as a previous, successful DNF simulation of 30-month-old children's behaviour in a fast mapping task (see Twomey & Horst, paper 4, this thesis), the DNF model predicts that children will extend previously fast-mapped novel names to completely category entral – but not category peripheral – exemplars. The current experiment tests this prediction empirically with 30-month-old children using a design identical to that used in the simulation just described.

Importantly, the stimuli used during referent selection were identical across conditions until the extension trials when children were presented with category central or category peripheral exemplars. This empirical study provides a robust test of the model, and a

successful replication serves as proof-of-concept of this architecture as an informative model of infant category label learning via fast mapping.

Method

Participants. 44 typically developing, monolingual, English-speaking 30-month-old children (23 girls, $M = 29\text{m}, 2\text{d}$, $SD = 45.56$ days; range = 24m, 0d - 32m, 29d) with a mean productive vocabulary of 521 words ($SD = 128.92$ words, range = 263 - 662 words) and no family history of colourblindness participated. Children were from predominantly middle class homes. Half of the children were randomly assigned to the *central* condition, and the other half were randomly assigned to the *peripheral* condition. Children's ages and productive vocabularies did not differ between conditions. Data from 10 additional children were excluded from analyses due to fussiness (7), experimenter error (2) and illness (1). Caregivers were reimbursed for travel expenses and children received a small gift for participating.

Stimuli. Known objects consisted of eighteen toys from categories familiar to 2-year-old children: an apple, a banana, a bus, a bunch of carrots, a cow, a cup, an elephant, a fish, a fork, a frog, a bunch of keys, a phone, a plane, a pig, a sheep, a shoe, a pair of sunglasses and a train.

Novel objects are depicted in Figure 3 and consisted of fifteen toys from three categories not familiar to 2-year-old children. In order to test the simulation's prediction, we designed novel exemplars that shared different numbers of perceptual features with exemplars from the same category. We did so in two ways. First, we kept the core shape of the objects the same across category members. Preschool children and infants are able to differentiate shape components in 3D objects (or "geons"; Abecassis, Sera, Yonas, & Schwade, 2001; Fulkerson & Haaf, 2003; Haaf, Lundy, & Coldren, 1996), and categorise solid objects on the basis of shared shape (the "shape bias",

Colunga & Smith, 2008; Landau, et al., 1988). Second, we introduced minor variations in colour and shape. Colour variation has been shown to facilitate novel label retention and, alongside small amounts of shape variation, extension (Twomey & Horst, Paper 1, this thesis). Thus, we reasoned that variation in number of geons as well as number of colours shared between category exemplars would facilitate children's categorisation and label extension. Specifically, based on the DNF simulation, we predicted that children would retain novel labels after familiarisation with a novel category during referent selection, but would only extend novel labels to completely novel objects which shared many features with exemplars from the familiarised category. Thus, in each novel category, each exemplar shared more or fewer geons and colours with every other exemplar in that category. Based on sharing more or fewer features with other exemplars, then, each exemplar from a given category exhibited graded category centrality.

Specifically, the *hux* category consisted of a rigid string of small, coloured wooden blocks attached to a coloured, circular base. Between exemplars, the number and length of the horizontal and vertical components of the string of blocks varied, as did the colour of the base (either blue or yellow). The *doff* category consisted of an ovoid, natural wooden base with two coloured branches on one side. Between exemplars, the number of branches varied, as did the colours of the branch segments (i.e., the first segments were red, the second orange, the third yellow and the fourth red). The *cheem* category consisted of a natural wooden bolt made up of a screw-thread with a large ball on one end, and coloured wooden blocks attached to the screw-thread. Although the five novel exemplars per category clearly did vary in number of shared features, it is important to bear in mind that, during referent selection, children in fact encountered six novel objects per category (that is, each novel object was encountered

twice: once on a known label trial, once on a novel label trial). Across exemplars, number, colour and shape of the attached blocks varied. Objects were approximately the same size (height = 30-79mm, width = 75-154mm, depth = 30-77mm).













Label	Referent Selection		Extension	
			<i>Few shared features</i>	<i>Many features</i>
<i>Hux</i>				
<i>Doff</i>				
<i>Cheem</i>				

Figure 3. Novel objects used in the empirical study.

With the goal of ensuring the extension objects were located either near (*central*) or far (*peripheral*) in representational space relative to the familiarised exemplars, we analysed the structure of each novel category as encountered during referent selection (i.e., three novel exemplars presented twice each, once labelled and once unlabelled). Multidimensional Scaling Analysis (or MSA; specifically the PROXCAL toolkit available in SPSS19) was performed for each novel category separately. MSA allows object dissimilarity information to be projected onto two-dimensional “psychological” space such that representational distances between objects can be visualised on a pair of

axes. MSA has been used to assess perceptual similarity of stimuli in both adult (e.g., Cohen, Nosofsky, & Zaki, 2001) and infant experiments (e.g., Abecassis, et al., 2001).

For each category, every exemplar was compared to every other exemplar along three separate “shared feature” metrics: number of shared geons, a binary shared label value (that is, “labelled” or “unlabelled”), and number of shared colours. Importantly, the three exemplars seen twice during referent selection were compared to every other exemplar twice, once as labelled and once as unlabelled. For each novel category, three matrices of difference values (shared geons, shared labels and shared colours) were entered into the analysis as interval data. Matrices were weighted 4, 4 and 1, respectively. Matrix weights reflected existing research on the relative influence on categorisation of shape, labels and colour. For example, Gliozzi et al.’s (2009) model of object categorisation demonstrated that labels and shape features influence object categorisation equally strongly; further, research demonstrates that shape influences object categorisation over and above colour (Horst & Twomey, 2012). Analyses of the fit of a range of models confirmed that these matrix weightings offered the best fit to the data (see Appendix).

The model provided a good fit to the data in three dimensions (mean normalised raw stress = 0.11, mean dispersion accounted for = 0.89, mean Tucker’s Coefficient = 0.94; note that an excellent model fit is indicated by mean normalised raw stress of around 0.1, dispersion accounted for approaching 1, and Tucker’s coefficient approaching 1, Dugard, Todman, & Staines, 2009). The MSAs confirmed that extension exemplars designed to share many shared features with the familiarised exemplars (see right panel of Figure 2) were closer in representational space to those exemplars than were extension exemplars designed to share few shared features (see Table 2). Importantly, however, these distances were relatively small, indicating that for

each category, the two extension objects were overall perceptually similar to the exemplars seen during referent selection and were therefore candidate members of those categories.

Procedure and design. Before the experiment began the caregiver was asked to complete the Macarthur-Bates Communicative Development Inventory (British Adaptation, Klee & Harrison, 2001). Caregivers were also shown colour photographs of all stimuli to ensure that they were appropriately familiar (in the case of the known objects) or novel (in the case of the novel objects). All children were familiar with all known objects, and no children were familiar with any of the novel objects.

	Average distance			
	<i>hux</i>	<i>doff</i>	<i>cheem</i>	<i>Mean</i>
<i>Central</i>	0.529	0.522	0.526	0.526
<i>Peripheral</i>	0.539	0.542	0.564	0.548

Table 2. Average distances between extension objects and referent selection exemplars, for each novel category, according to Multidimensional Scale Analysis based on shared geons, labels and colours.

During the experiment, children were seated in a booster seat across from the experimenter at a white table. Caregivers sat next to their child and continued to complete the MCDI (Klee & Harrison, 2001). Caregivers were instructed not to label any object or influence or give feedback on their children's responses.

The experiment began with three warm-up trials to familiarise children with the task. Stimuli were presented on a transparent plastic tray divided into three equal

sections. Three known objects, chosen at random from the known objects used during the referent selection phase, were presented to the child on the tray in pseudorandomly-determined position (i.e., left, middle or right). First, the experimenter held the tray stationary on the table and silently counted for three seconds to allow the child to look at the objects (see Horst & Samuelson, 2008). Then, the experimenter asked the child to select one of the objects (“Which one’s the cow? Can you show me the cow?”). All objects were labelled twice, with up to two more labelling instances when children needed encouragement. No object was labelled more than four times. The experimenter then slid the tray towards the child and allowed the child to point to or hand her one of the objects. All warm-up trials proceeded in an identical manner, using the same objects, except that a different object was asked for on each trial. Children were heavily praised for correct responses, and prompted to choose again for incorrect responses. 100% of the children successfully chose the target on each warm-up trial.

Referent selection trials immediately followed the warm-up trials and proceeded in an identical manner, except that children were given no feedback following their choices, that is, the experimenter only said either “OK” or “thank you.”. Each child was presented with 18 referent selection trials across three blocks of six trials (one block per novel category). In each block children saw three different sets of objects twice each: once on a trial with a novel target (“novel label trial”) and once on a trial with a known target (“known label trial”). Each set included one novel and two known objects (e.g., cow, bus, novel object). Each novel exemplar appeared once as a target object and once as a competitor; each novel word was therefore heard a minimum of three times. Within each block, trial order was pseudorandomised such that no more than two trials of the same type would be presented successively. Block order was counterbalanced across participants, and referent selection trials were identical across

conditions. Referent selection trials were immediately followed by a five-minute delay, during which time children remained at the table and coloured pictures from a colouring book. A digital kitchen timer was used to time the five-minute delay.

After the delay, children were presented with a new warm-up trial to re-engage them with the task. 100% of children chose the target in this trial. Three retention trials immediately followed the warm-up trial. Retention trials proceeded in an identical manner to referent selection trials, except that children were presented with three novel exemplars on each trial: one previously-encountered exemplar from each novel category. Which exemplar was presented was counterbalanced across participants using a Latin Square design. Thus, referent selection trials were identical across conditions.

The three extension trials proceeded in an identical manner to the retention trials. Children were presented with three completely novel exemplars, one from each novel category. In the *central* condition, children were presented with the exemplars that shared many perceptual features with those encountered during referent selection. In the *peripheral* condition, children were presented with the exemplars that shared few shared features with those encountered during referent selection.

Coding. Children's responses were coded offline from DVD. A second coder blind to the experimental hypotheses coded 20% of the sessions for reliability. Inter-coder agreement was high, $M = 96.43\%$, $SD = 0.04\%$ (range = 89.29% – 1.00%).

Results and Discussion

Results from the empirical study are depicted in Figure 4. During referent selection children mapped known labels to the correct referent at levels greater than expected by chance (0.33, all reported tests two-tailed), *central*: $t(21) = 50.56$, $p < .001$, $d = 10.78$; *peripheral*: $t(21) = 19.41$, $p < .001$, $d = 4.14$. Children also mapped novel

labels to the correct referent at levels greater than expected by chance (*central*: $t(21) = 20.75$, $p < .001$, $d = 9.06$; *peripheral*: $t(21) = 11.76$, $p < .001$, $d = 5.13$). At test, children retained novel labels at above-chance levels in both conditions, *central*: $t(21) = 2.46$, $p < .05$, $d = 0.53$, *peripheral*: $t(21) = 2.40$, $p < .05$, $d = 0.52$ (as in the simulation, no difference was anticipated between conditions for retention, as stimuli presented during referent selection and retention were identical across conditions). In contrast, however, children extended novel labels in the *central* condition, $t(21) = 3.38$, $p < .01$, $d = 0.73$; but did not extend novel labels in the *peripheral* condition, $t(21) = 1.45$, $ns.$, $d = 0.31$. However, children's proportion of correct choices on extension trials did not differ between conditions $t(42) = 1.00$, $ns.$, $d = 0.31$. Thus, children's overall pattern of responding replicated the overall pattern generated by the model; the moderate effect size suggests that the lack of between-condition difference for extension in the child data is likely due to greater variance in the child data than in the model data.

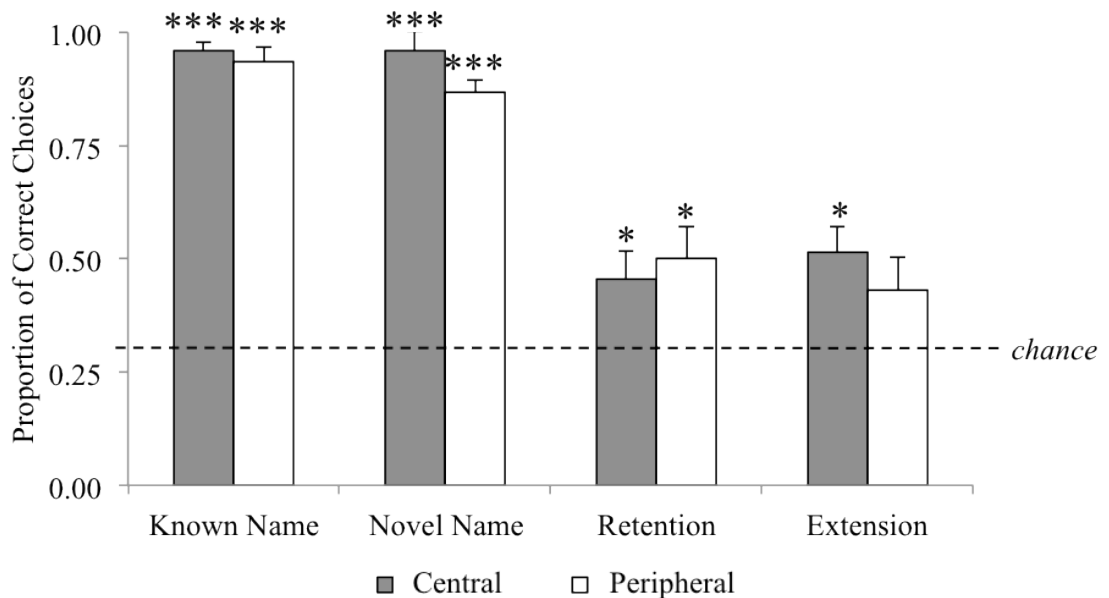


Figure 4. Children's proportion of correct choices in Experiment 1. Dotted line represents chance (.33). Error bars represent one standard error. *** $p < .001$, * $p < .05$.

These data support “correlated features” accounts of categorisation, for example the classic Younger & Cohen (1983; 1986) studies. Here, 10-month-old infants were sensitive to correlations between configural and perceptual attributes in novel 2D animal stimuli (see also Plunkett, et al., 2008; Rakison & Cohen, 1999; Younger & Cohen, 1986; Younger, Hollich, & Furrer, 2004). The current study demonstrates that older children can also generalise labels systematically based on correlations between perceptual features such as geons and colour.

These data also contribute to the debate concerning the status of labels in categorisation. Specifically, some argue for a “label-first” account, in which labels influence categorization more strongly than do perceptual features (e.g., Waxman & Braun, 2005; Lupyan, Rakison & McClelland, 2007). Recall that although extension exemplars shared different numbers of features with other exemplars from their category, the overall perceptual similarity between extension and referent selection objects was still high, both visually (see Figure 3 and the Multidimensional Scaling Analysis reported in Table 2). Thus, in the current study, the “labels-first” view predicts that children may extend novel labels in both conditions, because labels invite categorisation of perceptually-similar objects (Graham, Kilbreath & Welder, 2004; Jaswal, 2004; see Paper 2, this thesis, for similar results in action categorization).

In contrast, others argue that labels function as perceptual features of category exemplars, and influence categorisation in a similar way to other perceptual features such as geons (e.g. Gliozi, et al., 2009; Sloutsky & Fisher, 2011). Our data support this latter argument: if labels strongly drive categorisation, we would expect children to extend novel labels in both conditions, rather than solely in the *central* condition. In the current study, category centrality of extension exemplars drove children’s categorisation. Moreover, the Multidimensional Scale Analysis of the novel stimuli also supports this

view. The best-fitting model is generated when labels and geons are equally weighted, that is, when labels and geons influence category structure equally. Re-running the MSA with labels weights lesser or greater than geon weights decreases model fit (see Appendix), suggesting that in this study at least, labels and geons have an equal impact on categorisation. However, because there was no difference between groups for extension trials, this conclusion should be treated with caution; further research is required to resolve the debate as to the relative influence of labels and perceptual features on categorisation (see Sloutsky & Fisher, 2012). More generally, these and other conflicting data suggest that the relationship between labels and categories is likely to be contingent on task and experimental design (e.g., Jaswal, 2004).

Finally, the results of Experiment 1 – that small differences in perceptual features can lead to significant differences in label extension – provide novel insight into existing tests of word learning, and counsel caution in interpretation. Specifically, in analysing experiments using label generalisation/extension trials of any kind, it is critical to be aware that *lack of generalisation may not indicate lack of word learning* – children may simply be unwilling to generalise a learned label to an exemplar that is too perceptually different from trained exemplar(s).

General Discussion

This paper presents an experimental replication of predictions generated by a computational model of young children's word learning and categorisation. Based on a Dynamic Neural Field model of infants' word learning via mutual exclusivity (Twomey & Horst, Paper 4, this thesis), we simulated children's behaviour in a fast mapping task to examine the nature of children's noun extensions after familiarisation with an object category. The model predicted that children would retain previously fast-mapped labels,

but would extend labels only to category central novel exemplars. To test this prediction, 30-month-old children were familiarised with three novel object categories via 18 referent selection trials. At test, children were able to retain previously fast-mapped novel labels. However, during extension trials, only children who encountered category central completely novel objects were able to extend novel labels (see right hand panel, Figure 4). Children who encountered category peripheral completely novel objects did not generalise novel labels.

Along with the empirical data and computational replication presented in Twomey & Horst (2011), the computational data and empirical replication in the current study constitute the one of the first Dynamic Neural Field models to successfully simulate young children's unsupervised fast mapping, word learning and categorisation. In line with earlier applications of DNFs to developmental phenomena such as the A-not-B error (Simmering, Schutte, & Spencer, 2008), spatial binding of objects to labels (Samuelson, Smith, Perry, & Spencer, 2011) and the shape bias (Samuelson, Schutte, & Horst, 2009), this model successfully simulates apparently complex behaviour using simple low-level associative processes. Importantly, during referent selection the model is able to map novel words to novel referents without any preprogrammed "reasoning" module.

Dynamic Neural Field models are theoretically situated in Dynamic Systems theory, in which complex yet stable behavioural and cognitive structures emerge ad hoc from the interaction between components available at a given time (for example, the body, perceptual input, and the task environment) in the context of nested timescales of learning (for example, lifetime experience with categories and labels, exemplars and labels encountered earlier in the experiment, and the exemplar and label present on a given trial). Thus, these data add weight to the growing body of work demonstrating

that cognition, behaviour and the environment are inextricably coupled and inseparable from their temporal context. As such, the current study contributes not only to our understanding of young children's fast mapping and categorisation, but also to a new conception of developing cognition as an emergent dynamic system.

Acknowledgements

This research was supported by ESRC Quota Award ES/H014942/1 and an ESRC Overseas Institutional Visit Grant to K.E.T. The authors would like to thank Lucy Metcalfe and Nicole Jones for data collection and coding. We are grateful for Plan Toys for the kind donation of stimuli. Finally, we would like to thank children and caregivers for taking part.

References

- Abecassis, M., Sera, M. D., Yonas, A., & Schwade, J. (2001). What's in a shape? Children represent shape variability differently than adults when naming objects. *Journal of Experimental Child Psychology*, 78(3), 213-239. DOI: 10.1006/jecp.2000.2573
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3. PMid:22291676
- Carey, S. (2011). Concept innateness, concept continuity, and bootstrapping. *Behavioral and Brain Sciences*, 34(03), 152-162. DOI: 10.1017/S0140525X10003092
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29. DOI: 10.3758/BF03206386
- Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, 29(8), 1165-1175.
- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112(2), 347-382. DOI: 10.1037/0033-295X.112.2.347
- Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: the self-organization of skilled noun learning. *Developmental Science*, 11(2), 195-203. DOI: 10.1111/j.1467-7687.2007.00665.x. DOI: 10.1111/j.1467-7687.2007.00665.x
- Diesendruck, G., Markson, L., & Bloom, P. (2003). Children's reliance on creator's intent in extending names for artifacts. *Psychological Science*, 14(2), 164-168. DOI: 10.1111/1467-9280.t01-1-01436

- Dugard, P., Todman, J., & Staines, H. (2009). *Approaching multivariate analysis: a practical introduction* (Vol. 67). New York: Psychology Press.
- Ellis, A. E., & Oakes, L. M. (2006). Infants flexibly use different dimensions to categorize objects. *Developmental Psychology*, 42(6), 1000–1011. DOI: 10.1037/0012-1649.42.6.1000
- Fulkerson, A. L., & Haaf, R. A. (2003). The influence of labels, non-labeling sounds, and source of auditory input on 9-and 15-month-olds' object categorization. *Infancy*, 4(3), 349-369. DOI: 10.1207/S15327078IN0403_03
- Gliozzi, V., Mayor, J., Hu, J. F., & Plunkett, K. (2009). Labels as features (not names) for infant categorization: A neurocomputational approach. *Cognitive Science*, 33(4), 709-738. DOI: 10.1111/j.1551-6709.2009.01026.x
- Graham, S. A., Kilbreath, C. S., & Welder, A. N. (2004). Thirteen-month-olds rely on shared labels and shape similarity for inductive inferences. *Child Development*, 75(2), 409–427. DOI:10.1111/j.1467-8624.2004.00683.x. DOI: 10.1111/j.1467-8624.2004.00683.x
- Haaf, R. A., Lundy, B. L., & Coldren, J. T. (1996). Attention, recognition, and the effects of stimulus context in 6-month-old infants. *Infant Behavior & Development*, 19(1), 93-106. DOI: 10.1016/S0163-6383(96)90047-8
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cogn Psychol*, 53(4), 310-344. DOI: S0010-0285(06)00025-9.
- Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

- Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infants categorize. *Child Dev.*, 76(3), 614–631. DOI: 10.1111/j.1467-8624.2005.00867.x
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI: 10.1080/15250000701795598
- Horst, J. S., & Twomey, K. E. (2012). It's Taking Shape: Shared Object Features Influence Novel Noun Generalizations. *Infant and Child Development*. DOI:10.1002/icd.1768
- Horst, J. S., Scott, E. J., & Pollard, J. P. (2010). The Role of Competition in Word Learning Via Referent Selection. *Developmental Science*, 13(5), 706-713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Houston-Price, C., Plunkett, K., & Harris, P. (2005). 'Word-learning wizardry' at 1;6. *Journal of Child Language*, 32(1), 175-189. DOI: 10.1017/S0305000904006610
- Jaswal, V. K. (2004). Don't Believe Everything You Hear: Preschoolers' Sensitivity to Speaker Intent in Category Induction. *Child Development*, 75(6), 1871–1885. doi:10.1111/j.1467-8624.2004.00822.x. DOI: 10.1111/j.1467-8624.2004.00822.x
- Klee, T., & Harrison, C. (2001). *CDI Words and Sentences: validity and preliminary norms for British English*. Paper presented at the Child Language Seminar, University of Hertfordshire.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. DOI: 10.1016/0885-2014(88)90014-7

- 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. DOI : 10.1111/j.1467-9280.2007.02028.x
- McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2011). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences*, 14(8), 348-356. DOI: 10.1016/j.tics.2010.06.002
- Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science*, 7(2), 141-148. DOI: 10.1111/j.1467-7687.2004.00331.x
- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665-681. DOI: 10.1016/j.cognition.2007.04.003
- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge: MIT Press.
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, 27(2), 145-179. DOI: 10.1016/0010-0277(87)90017-5
- Rakison, D. H., & Cohen, L. B. (1999). Infants' use of functional parts in basic-like categorization. *Developmental Science*, 2(4), 423-431. DOI: 10.1111/1467-7687.00086
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology - General*, 104(3), 192–233.

- Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: Moment-to-moment interactions shape children's naming biases. *Infancy, 11*(1), 97-110. DOI: 10.1207/s15327078in1101_5
- Samuelson, L. K., & Horst, J. S. (2008). Confronting complexity: insights from the details of behavior over multiple timescales. *Developmental Science, 11*(2), 209–215. DOI 10.1111/j.1467-7687.2007.00667.x
- Samuelson, L. K., Schutte, A. R., & Horst, J. S. (2009). The dynamic nature of knowledge: Insights from a dynamic field model of children's novel noun generalization. *Cognition, 110*(3), 322-345. DOI: 10.1016/j.cognition.2008.10.017
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding Word Learning in Space. *PloS one, 6*(12), e28095. DOI: 10.1371/journal.pone.0028095
- Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research, 1202*, 68-86. DOI: 10.1016/j.brainres.2007.06.081
- Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General, 133*(2), 166. DOI: 10.1037/0096-3445.133.2.166
- Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology, 111*(1), 65–86. DOI:10.1016/j.jecp.2011.07.007
- Smith, L. B. (2000). Learning How to Learn Words: An Associative Crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a Word Learner: A Debate on*

Lexical Acquisition (pp. 51-80). New York: Oxford University Press. DOI:
10.1093/acprof:oso/9780195130324.003.003

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features? *Journal of Experimental Child Psychology*, 111(1), 65–86. DOI:10.1016/j.jecp.2011.07.007

Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13-19. DOI: 10.1111/1467-9280.00403

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, 6(4), 392-412. DOI: 10.1111/1467-7687.00295

Spencer, J. P., Thomas, M. S. C., & McClelland, J. L. (2009). *Toward a Unified Theory of Development*. Oxford University Press New York.

Twomey, K. E. & Horst, J. S. (Paper 1, this thesis). That's more like it: More exemplars facilitate word learning.

Twomey, K. E. & Horst, J. S. (Paper 2, this thesis). Twinkle twinkle little star: Perceptual variability facilitates early verb learning

Twomey, K. E & Horst, J. S. (Paper 4, this thesis). All things considered: Dynamic Field Theory captures the effect of category variability on young children's word learning.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: new evidence from 12-month-old infants. *Cognition*, 95(3), B59-B68. DOI: 10.1016/j.cognition.2004.09.003

Younger, B. A., & Cohen, L. B. (1983). Infant Perception of Correlations among Attributes. *Child Development*, 54(4), 858-867. DOI: 10.2307/1129890

- Younger, B. A., & Cohen, L. B. (1986). Developmental-Change in Infants Perception of Correlations among Attributes. *Child Development*, 57(3), 803-815. DOI: 10.2307/1130356
- Younger, B. A., & Fearing, D. D. (1998). Detecting correlations among form attributes: An object-examining test with infants. *Infant Behavior & Development*, 21(2), 289-297. DOI: 10.1016/S0163-6383(98)90007-8
- Younger, B. A., Hollich, G., & Furrer, S. D. (2004). An emerging consensus: Younger and Cohen revisited. *Infancy*, 5(2), 209-216. DOI: 10.1207/s15327078in0502_6

Appendix

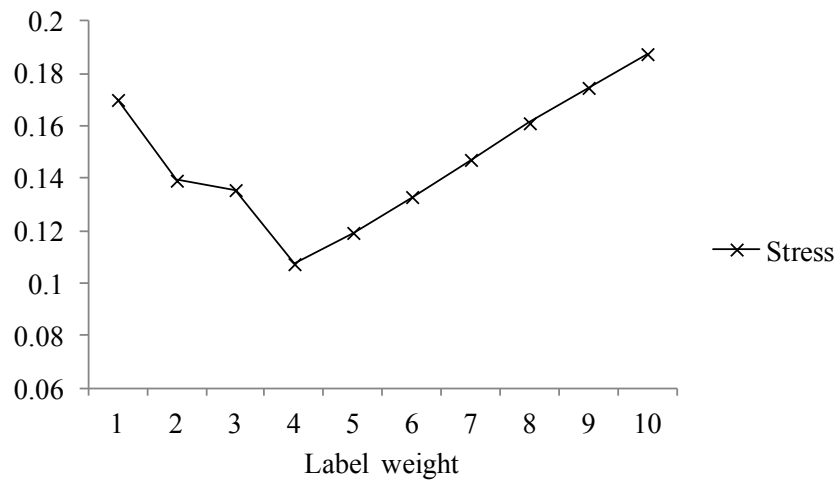


Figure A5. Increase in normalised stress as a function of label weight in the Multidimensional Scale Analysis. Normalised stress lower than 0.15 indicates a good model fit (Dugard, et al., 2009). Best fit occurs when labels are equally weighted with geons (label weight = 4)

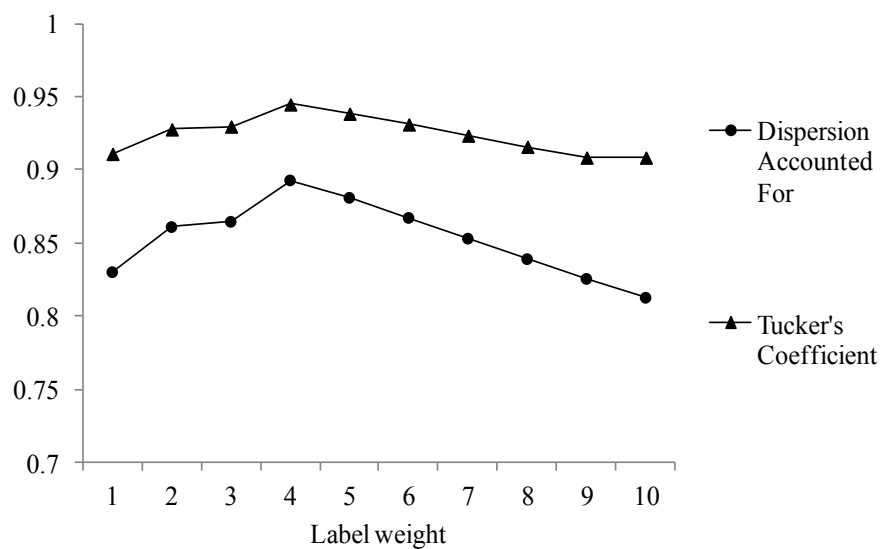


Figure A6. Decrease in Dispersion Accounted For and Tucker's Coefficient as a function of label weight in the Multidimensional Scale Analysis. For both measures, values close to 1 indicate a good model fit (Dugard, et al., 2009). Best fit occurs when labels are equally weighted with geons (label weight = 4)

Running head: AN EMBODIED MODEL OF WORD LEARNING

An embodied model of young children's categorisation and word learning

Katherine E. Twomey¹, Jessica S. Horst¹, Anthony Morse²

¹University of Sussex. ²University of Plymouth

Author note

A version of this paper appears in:

Twomey, K. E., Morse, A. F. & Horst, J. S. *An embodied model of young children's categorization and word learning* (2013). In Gogate, L. & Hollich, G. *Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence*. IGI Global. DOI: 10.4018/978-1-4666-2973-8

Note to examiners

This paper presents the results of a collaboration with the Centre for Robotics and Neural Systems at the University of Plymouth, UK. The first author instigated the collaboration, worked with the third author in adapting the robotic system to the word learning/categorization context, designed the experimental task, collected and analysed all data, and wrote this paper. The third author, however, designed the Epigenetic Robotic Architecture (ERA) which was adapted in this project to control the iCub robot (Morse, de Greeff, Belpeame, Cangelosi, 2010a), and contributed the section describing the equations which govern the neural networks in the ERA (page 20, paragraph 2 to page 22). This contribution was approved by Dr. Susan Ayers, the University of Sussex School of Psychology Director of Doctoral Studies, when this paper was written.

Abstract

Children learn words with remarkable speed and flexibility. However, the cognitive basis of young children's word learning is disputed. Further, although research demonstrates that children's categories and category labels are interdependent, how children learn category labels is also a matter of debate. Recently, biologically plausible, computational simulations of children's behaviour in experimental tasks have investigated the cognitive processes that underlie learning. The ecological validity of such models has been successfully tested by deploying them in robotic systems (Morse, Belpaeme, Cangelosi, & Smith, 2010). We present a simulation of children's behaviour in a word learning task (Twomey & Horst, Paper 1, this thesis) via an embodied system (iCub; Metta, et al., 2010), which points to associative learning and dynamic systems accounts of children's categorisation. Finally, we discuss the benefits of integrating computational and robotic approaches with developmental science for a deeper understanding of cognition.

An embodied model of young children's categorisation and word learning

From birth – indeed, even before birth (James, 2010; Shahidullah & Hepper, 1994) – infants encode a myriad of complex perceptual stimuli. The extent of this complexity cannot be overestimated: in the visual domain alone, the myopic newborn must segment the visual scene, distinguish between figure and ground, group surfaces into objects, represent temporal and spatial continuity of objects, and infer the physical characteristics of partially-occluded objects (Johnson, 2010a). However, very young infants can make sense of the intricacies of their environment. Even neonates can group aspects of their perceptual environment into early categories (Johnson, 2010b), systematically treating discriminably different exemplars as equivalent. A few hours after birth, infants are able to discriminate their mothers' faces from those of strangers (Field, Cohen, Garcia, & Greenberg, 1984) and by three months, infants discriminate male versus female and same- versus own-race faces (Slater et al., 2010).

By the end of their first year, infants have developed an impressive ability to categorise in multiple domains, and use a variety of criteria to do so. For example, infants can use relative luminance to categorise patterns of horizontal or vertical black bars after familiarisation with arrays of light or dark shapes (3-4 months; Quinn, Burke, & Rush, 1993); head information to categorise pictures of animals (3 months; Quinn, Eimas, & Rosenkrantz, 1993); auditory statistical cues to categorise phonemes in the speech stream (6 months; Grieser & Kuhl, 1989); and visual spatiotemporal information to categorise event types (7.5 months; see Baillargeon & Wang, 2002 for a review).

Children's remarkably early ability to detect patterns in their environment is not in dispute (Gogate & Hollich, 2010). However, the processes underpinning children's categorisation and the structure of the categories themselves are less clear-cut. The

current chapter presents novel insights into the interplay between young children's categorisation and word learning from an embodied computational model.

A bidirectional relationship between categories and labels

The astonishing speed and ease with which very young infants form adult-like categories in “noisy learning environments” (Wu, Gopnik, Richardson, & Kirkham, 2011) has led some to suggest that categorisation operates in a top-down fashion, based on innate biases, or core principles, which guide domain-specific developmental processes such as early face and object perception, and imitation (Carey & Spelke, 1994; Meltzoff & Moore, 1977; Slater et al., 1998; Spelke & Kinzler, 2007). Others suggest that perceptual constraints may be co-opted from other domains, but rapidly become domain-specific (Markman, 1994; Waxman & Booth, 2001). Still others argue that categorisation is a fundamentally associative, consistently domain-general process that emerges across development from dynamic interactions between environment, body and cognition (Kovack-Lesh, Horst, & Oakes, 2008; Rakison & Yermolayeva, 2011; Smith, Colunga, & Yoshida, 2010; see also Gogate & Hollich, 2010). Moreover, Rakison (2000) has argued that infant categories are not (and need not be) adult-like.

By the onset of word production at approximately one year children are experienced categorisers. However, learning labels for categories is still no easy task. For each new word that they encounter, children are faced with a dizzying array of possible referents (Quine, 1960). When a child hears a new word for the first time, for example in the context of an object, the child must determine whether that word refers to the whole object, one of its parts, its texture, its colour, its position, its function, an event that the object is involved in, and so on across an infinite number of possibilities.

Echoing the debate surrounding categorisation, theories abound as to how children determine the referent of a novel word. Again, at one end of the spectrum,

some accounts propose innate cognitive faculties (Bloom, 2000; Markman, 1994; Woodward & Markman, 1991). Diesendruck and Bloom (2003), for example, argue that children's *a priori* knowledge of object kinds guides them to use object shape as a guide to category membership and label extension when they encounter novel objects. From this perspective, a static, abstract and extralinguistic object ontology determines children's categorisation behaviour.

In contrast, at the other end of the spectrum, proponents of associative accounts argue that language learning is contingent on domain-general cognitive processes. From this perspective, linguistic structure emerges from statistical regularities in the perceptual environment (Colunga & Smith, 2008; Rogers, Rakison, & McClelland, 2004; Samuelson, 2002). This dynamic systems account of categorisation is in sharp contrast to the nativist stance. Here, behaviour and cognition are dynamically coupled and emerge out of interactions between the agent, the environment, and nested timescales: from long-term learning, to just-past experience to in-the-moment input (Thelen & Smith, 1996). For example, Gershkoff-Stowe and Smith (2004) provide evidence that English-learning children's emerging bias to categorise solid objects by shape reflects statistical regularities in their early-learned vocabulary. They argue that children's categorisation is assembled online as a product of their long-term linguistic experience, experience with object categories and labels, and the demands of the task in hand (see also Samuelson & Horst, 2008). Thus, according to the dynamic systems perspective, understanding development is impossible without viewing cognition as embodied, interactive, and emergent.

Evidence is mounting for an intimate link between categories and their labels. Early in the word learning process children use previous experience with categories to extend newly-learned labels to new category exemplars (Smith, Jones, Landau,

Gershkoff-Stowe, & Samuelson, 2002). For example, a child might learn that her large, furry, brown toy is called a “bear.” Then, experience with further exemplars reinforces that category: for example, learning that the TV animation of a large, brown animal called Yogi is called a “bear,” that the huge, white animal at the zoo is called a “bear,” and so on. Existing knowledge about these categories (in this case, *bear-shaped things are called “bear”*) then influences future categorisation of novel objects, and in turn, new exemplars enrich existing category representations (Smith, 2000). Thus, early in development, existing category knowledge affects children’s generalisation of labels to potentially new category members (Gershkoff-Stowe & Smith, 2004).

Recent empirical research, at different levels of analysis, suggests that the relationship between categories and labels is bidirectional; that is, that category structure and category labels interact dynamically. For example, at the neurological level, hemispheric localisation of children’s categorical perception for colour changes alongside an increase in linguistic experience (Franklin, Drivonikou, Clifford, Kay, Regier, & Davies, 2008; see also Travis, et al. (2011) for neural correlates of object label processing). At the behavioural level, novel labels have also been shown to influence online categorisation, specifically of novel objects displayed to 10-month-old infants in a novelty preference task (Plunkett, Hu, & Cohen, 2008).

Conversely, category structure affects children’s ability to learn category labels. In a longitudinal training study, Perry, Samuelson, Malloy and Schiffer (2010) demonstrated that experience with variable categories facilitated 18-month-old children’s noun generalisations and accelerated vocabulary development. Further, Twomey & Horst (Paper 1, this thesis) demonstrated that in-task category variability directly affects 30-month-old children’s ability to both recall and generalise novel category labels. Children who encountered novel exemplars of low-variability

categories in a referent selection task (see Horst & Samuelson, 2008) learned labels for these categories but did not extend these labels to new category exemplars, while children who encountered high-variability categories did extend labels to new category exemplars. Together these studies offer converging evidence that categorisation need not be contingent on core (or innate) structure; rather, it is a dynamic process in which new cognitive structure is softly assembled as a product of online input and previous experience (for similar arguments see Kovack-Lesh, Horst & Oakes, 2008; Ribar, Oakes & Spalding, 2004).

A step forward in understanding cognitive development

The 1970s saw a boom in interest in cognitive development (e.g., Fantz & Fagan, 1975). Over the past four decades, a wide variety of ingenious experimental paradigms have been developed to investigate infant categorisation. There exists a rich library of data from psychophysical measures such as habituation (Cohen & Strauss, 1979; for a review, see Oakes, 2010), preferential looking (Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987) and eye tracking (Aslin & Salapatek, 1975; see also Gredebäck, Johnson, & von Hofsten, 2009); behavioural studies of manual object-examining (Oakes, Madole, & Cohen, 1991) and deferred imitation (Meltzoff & Moore, 1977); and neuroimaging such as EEG (Samuel, 1978), fMRI (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002), and, more recently, NIRS (Aslin & Mehler, 2005; Fava, Hull & Bortfield, 2011).

Despite exciting progress toward understanding categorisation, however, empirical data can only take us so far: statistical models reveal much about the relationships between variables, but even longitudinal studies only provide a temporally cross-sectional view of those relationships (Simmering, Triesch, Deák, & Spencer, 2010). Hypotheses about the cognitive structures underlying a behaviour can be tested with varying degrees of success, but convincing process-based accounts of both infant

and adult categorisation remain comparatively few. Recently, however, computational models have successfully simulated children's learning in a variety of linguistic and nonlinguistic domains, lifting the metaphorical lid on previously inaccessible cognitive organisation. From probabilistic, Bayesian networks modelling optimal reasoning via hypothesis elimination (e.g., Perfors, Tenenbaum, Griffiths, & Xu, 2011; Xu & Tenenbaum, 2007), to networks modelling associative learning (e.g., Colunga & Smith, 2003; Johnson, Spencer, & Schöner, 2009; Mayor & Plunkett, 2010; Munakata & McClelland, 2003; Rogers & McClelland, 2004; Westermann & Mareschal, 2009), and hybrid models combining probabilistic function with neural plausibility (e.g., Feldman & Bailey, 2000; Rao, 2004) among a variety of formal models of psychological processes (e.g., Gaussian models of synchrony detection, Prince & Hollich, 2005; exemplar models of speeded classification, Nosofsky & Stanton, 2006), computational models offer much-needed insights into categorisation and linguistic phenomena. The current chapter describes a step forward in this computational trend in developmental psychology: a demonstration of categorisation using a neural network model in an embodied robotic system.

Empirical basis of the current project

Since Carey (1978) coined the term “fast mapping,” scores of studies have demonstrated children's ability to quickly link a novel noun to a novel referent (e.g., Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Jaswal & Markman, 2001; and see e.g., Childers & Tomasello, 2002; Heibeck & Markman, 1987; Namy, 2001 for fast mapping in other domains). However, more recently, it has become clear that fast mapping alone does not constitute word learning (Horst & Samuelson, 2008; Munro, Baker, McGregor, Docking & Arciuli, 2012). For example, in the laboratory setting, although 24-month-old children reliably determine the referent of a novel label from an

array of several objects, they do not reliably recall this label after a five-minute delay unless ostensive labelling is provided by the experimenter (Horst & Samuelson, 2008, see also Axelsson, Churchley & Horst, 2012). Robust word learning occurs when the child is able to either use the label appropriately after a delay without scaffolding category judgment from other information in the environment, or to generalise the novel label to previously-unseen category exemplars (Horst & Samuelson, 2008; Munro et al., 2012; Riches, Tomasello, & Conti-Ramsden, 2005).

The simulation presented in this paper is based on a recent study investigating the interaction between category variability and labelling (Twomey & Horst, Paper 1, this thesis) and focuses on two facets of word learning. First, it is well-established that young children are able to infer the correct referent for a novel category label without supervision or feedback (Akhtar, Jipson & Callanan, 2001; Merriman & Bowman, 1989). This ability is readily demonstrated by presenting the child with an array consisting of one or two exemplars from categories for which the child knows a label (e.g., COW and SPOON and one exemplar from a novel (unlabelled) category. Even very young children can reliably map a novel label to the novel exemplar and have been shown to have a bias to map labels to the most novel object in a given array (Halberda, 2003, 2006; Horst, Samuelson, Kucker & McMurray, 2011; Mervis & Bertrand, 1994). However, this ability is susceptible to task, developmental and linguistic factors (Au & Glusman, 1990; Houston-Price, Caloghiris, & Raviglione, 2010; Markman & Wachtel, 1988). Children in the empirical task described by Twomey & Horst (2011) reliably recalled newly-fast-mapped novel category labels without supervision or feedback, despite increases in variability between novel category exemplars. In this chapter we present computational evidence of word learning without supervision in an identical task context, demonstrating that this apparently inference-based behaviour can be

achieved from perceptual input alone.

Second, empirical studies indicate that children's ability to retain novel category labels depends on both in-task and longer-term factors, such as the number of competitor objects the child sees alongside the novel target (Horst, Scott, & Pollard, 2010), frequency of repetition (Mather & Plunkett, 2009), degree of prior familiarity with the exemplars (Kucker & Samuelson, 2011), presence or absence of ostensive labelling cues (Axelsson, Churchley, & Horst, 2012) or trial order (Childers & Tomasello, 2002; Vlach, Sandhofer & Kornell, 2008). Clearly, multiple factors can affect whether children can learn words via fast mapping. Twomey & Horst (2011) asked if children could generalise novel labels to never-before-seen exemplars, after encountering either moderately or highly variable categories in a fast mapping task.

Empirical task.

Procedure. Twenty-four 30-month-old children were familiarised with three novel object categories and their labels. The experiment consisted of three phases: first, a referent selection phase consisting of three known and three novel trials per category (18 trials in total), second, a recall test consisting of one trial per category (three trials in total) and finally, a generalisation test consisting of one trial per category (three trials in total). Children were seated opposite the experimenter across a white table; stimuli were presented equidistantly on a transparent tray. Children were assigned to either the *narrow* ($n = 12$) or *broad* ($n = 12$) conditions.

On each referent selection trial the child was presented with two exemplars from categories for which that child knew a label (known exemplars) and one exemplar of a novel category for which that child did not know a label (novel exemplar). After approximately three seconds during which the child was allowed to look at the objects, the child was then asked to select either the novel or one of the known exemplars

(e.g., “can you show me the *cheem*?”). Children heard each novel label three times. Following a five-minute delay, children were given three recall trials with three novel exemplars (one exemplar from each of the three novel categories). Finally, children were given three generalisation trials with three completely novel exemplars (one never-before-seen exemplar from each of the three categories) to further explore the robustness of the categories learned during familiarisation task. No feedback, positive or negative, was given during or after referent selection or test trials.

Importantly, the design and procedure were held constant between conditions: within-category variability was the only difference between conditions. Previous research indicates that, particularly in the context of a novel label, children will categorise discriminably different stimuli that vary in perceptual features (Plunkett, Hu, & Cohen, 2008; Quinn, Eimas, & Rosenkrantz, 1993; Younger & Cohen, 1986). Decades of research have also demonstrated that English-learning children categorise objects by shape rather than size or colour (e.g., Gershkoff-Stowe & Smith, 2004; Landau, Smith, & Jones, 1988; Smith, et al., 2002; Soja, Carey, & Spelke, 1991). Thus, to ensure minimal variability in the *narrow* condition, exemplars varied in colour alone. To introduce additional variability in the *broad* condition, exemplars varied in colour, size and texture. However, to facilitate categorisation, variation in the *broad* condition was kept low within each category; for example, the castanets varied only in colour and base-shape.

Results. Table 1 depicts the results from the empirical task. All children were able to recognise known exemplars at above-chance levels and were able to reliably map novel labels to novel exemplars. In the recall test, children were able to reliably recall previously fast-mapped category labels when category exemplars varied in colour alone. Children in the *broad* condition, however, were unable to recall previously

mapped category labels when exemplars varied in shape, texture and colour. Children in the *narrow* condition were able to recall significantly more category labels than children in the *broad* condition. In the generalisation test, children in the *narrow* condition did not generalise previously mapped labels to never-before-seen exemplars; however children in the *broad* condition were able to generalise previously mapped labels. However, no difference was found between these two groups.

	Referent selection (fast mapping)		Test Phase (word learning)	
	Known exemplar	Novel exemplar	Recall	Generalisation
Narrow	0.91***	0.77***	0.69***	0.50
	(0.19)	(0.26)	(0.26)	(0.33)
Broad	0.93***	0.83***	0.39	0.56*
	(0.12)	(0.11)	(0.24)	(0.30)

Table 1. Proportion of correct choices in the empirical task. Standard deviations are shown in brackets. ** $p < .01$, *** $p < .001$

Flexible categorisation

The empirical data illustrate children's flexible, online categorisation. The data also provide new evidence about the interaction between categorisation and category labels: clearly, some variability in category structure (e.g., exemplars of different colours) helps children recall category labels, but too much variability impairs this ability. Thus, variability appears to help children learn words but, crucially, only up to a point.

Children's behaviour in the generalisation test is consistent with findings from the categorisation literature. Recall that lack of generalisation in the *narrow* condition is in line with findings demonstrating that young children familiarised with perceptually similar exemplars form narrow categories that exclude variable exemplars (e.g., Eimas, Quinn, & Cowan, 1994; Quinn, Eimas & Rosenkrantz, 1993). Similarly, lack of recall alongside reliable generalisation in the *broad* condition is particularly consistent with exemplar-based models of categorisation (e.g., Medin & Schaffer, 1978; for a discussion, see J. D. Smith & Minda, 1998) according to which a delay between familiarisation and recall allows individual representations to degrade such that the memory trace may be more similar to a novel exemplar than to the original exemplar itself (for a discussion of similar findings see Murphy, 2004; Quinn, 2005; Quinn, 1987).

These markedly different categories emerged solely due to differences in variability of stimuli: across experiments, other sources of variation were held constant. In line with a dynamic systems account of categorisation, in which categories emerge as the product of the dynamic interplay between environment, body, cognition and timescales, these data lend support to the view of categorisation and cognitive development as an interactive, online process.

iCub Implementation

Computational models developed under the umbrella of developmental psychology have successfully simulated categorisation and produced valuable novel insights into the mechanisms by which categories develop. However, such models often (and by their nature) address a single area of interest, for example, modelling a single type of experiment (Mareschal, Quinn, & French, 2002; Twomey & Horst, 2011), or a single domain (Regier, Kay, & Khetarpal, 2007; Samuelson, 2002), often at a single level of analysis (French, Mareschal, Mermillod, & Quinn, 2004; but see McMurray,

Horst, Toscano, & Samuelson, 2009; Schutte & Spencer, 2010). While focusing on a specific area undoubtedly reveals links between a model and its behavioural substrate, the extent to which some models simulate processes *general* to cognition is not clear.

Although computational models have replicated empirical data from several domains, they have been criticised for lacking ecological validity (Cowan, 2003; Diesendruck & Graham, 2010; Murphy, 2003; Hollich & Prince, 2009). However, a large body of work suggests that action and perception, and therefore cognition and behaviour, are fundamentally embodied (e.g., Iiada, Pfeifer, Steels & Kuniyoshi, 2004; Pfeifer & Bongard, 2007; Samuelson, Smith, Perry & Spencer, 2011; Thelen & Smith, 1996; Ziemke, 2003). Embedding a computational model in a humanoid system, however, allows the modeller to directly address issues of embodiment (Morse, de Greeff, Belpaeme, & Cangelosi, 2010).

Robotic implementations of computational models are more than showcases for sophisticated engineering skills, then. Rather, they are indispensable research tools that offer unprecedented insight into the real-time interactions between body and cognition. For example, humanoid robotic systems directly address the embodiment issue (Berthouze & Metta, 2005). Further, a system in which perceptual input really does come from the environment ensures a focus on moment-by-moment perceptual processing as well as minute-by-minute and longer-term learning, forcing the integration of perception, cognition, and action. Embodied models therefore have the potential to offer rigorous tests of theories of domain general cognition in an ecologically (that is, physically, environmentally, and temporally) valid context, allowing us to examine the emergence of complex cognitive processes across both time and experience (see also Morse, de Greeff, Belpaeme, & Cangelosi, 2010).

The iCub

Building on the successful replication of the data described by Twomey & Horst (2011) with a Dynamic Neural Field model (cf. Spencer & Schöner, 2003), and in the context of recent advances in developmental robotics (for a review, see Vernon, Metta, & Sandini, 2007), we asked if the findings could be further explored using a domain-general hybrid neural network model in an embodied, robotic system. To this end, we extended the experimental and theoretical scope of an existing architecture (Morse, et al., 2010b) recently used in the iCub (Metta, et al., 2010).

The iCub is depicted in Figure 1. The iCub is an open-source, humanoid robot of approximately the same size and physical proportions as a three-year-old child with 53 degrees of freedom. Its sensors provide auditory, visual, tactile, force and proprioceptive (force, torque and joint angle) input. The iCub's sensory environment therefore provides some of the richness of that of a young child, though how to make use of this sensory information is left to the modeller. The iCub enables the integration of crossmodal inputs and provides various ways to coordinate its own movements to produce a range of behaviours; for example, to bind haptic and visual information to grasp new objects; auditory, visual and spatial information to recognise and reach for objects; auditory, visual and proprioceptive information to imitate human actions, and so on. As such, the iCub is the focus of numerous research directions, from language as an embodied system (Zeschel & Tuci, 2011) to the dynamics of human-robot interaction (Cangelosi, et al., 2008). An iCub simulator and much of the software developed to control the iCub are freely available under open source licensing from the iCub repository (see <http://eris.liralab.it/>) including the models reported herein as part of the

Aquila cognitive robotics tool kit (Peniak, Morse, Larcombe, Ramirez-Contla, & Cangelosi, 2011).

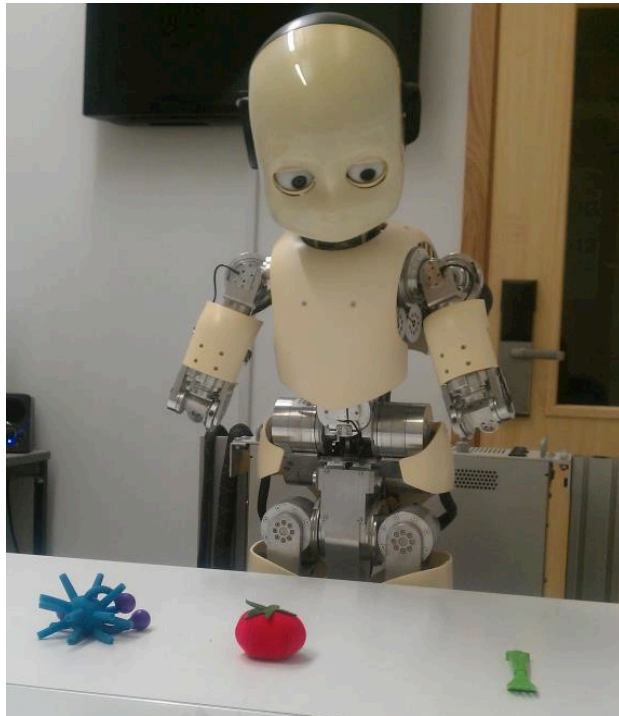


Figure 1. The iCub humanoid robot looking at a *fork*, a *tomato*, and a novel object

Neural architecture and inputs

The architecture employed in the current project was based on Morse et al.’s (2010a) ERA architecture previously used in a successful replication of a word-learning experiment (Morse, Belpaeme, Cangelosi & Smith, 2010; see also (Samuelson, Smith, Perry, & Spencer, 2011; Smith & Samuelson, 2010). This experiment employed the iCub robot to investigate the central role of body in the orchestration of early cognitive development. In line with children’s behaviour in the “*modi*” task (Smith & Samuelson, 2010), the robot was able to learn labels for objects only when label and object were spatially correlated.

The current project exploited a similar architecture (see Fig. 2), in which perceptual input is processed by Self-Organising Maps (SOMs, Kohonen, 1998). SOMs

are neural networks that self-organise over time via a winner-takes-all mechanism, such that the final organisation of output neurons reflects the topology of the input, with neighbouring neurons responding to similar input patterns. SOMs therefore provide a classification mechanism, which lends itself to the categorisation of complex perceptual inputs. It should be noted that the model is not simply a SOM, rather SOMs are used to adapt the model to whatever input space it is applied. The model is a structured network of associative connections providing constant and dynamic spreading of activation and inhibition, resulting in the behaviour discussed herein.

Visual input to the network first passes through two SOMs representing colour, and height/width and edge complexity (hereafter, “shape”), which are each bi-directionally coupled to a central connectionist “hub” of 36 neurons. Visual input is pre-processed from the images provided by the iCub’s cameras.

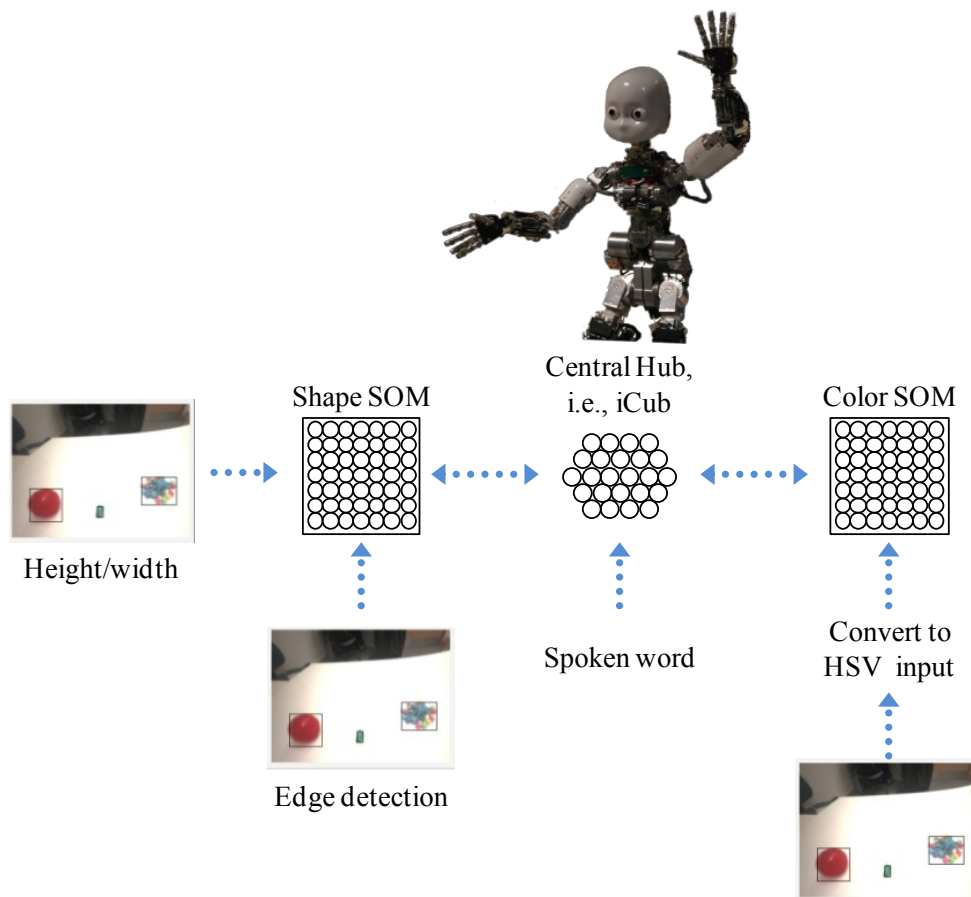


Figure 2. Model architecture.

Colour information for an object in a particular region is extracted by determining the location in HSV colour space (Alvy Ray, 1978) of each pixel in that region. All pixels with a saturation value greater than a threshold of 0.2 (thereby ignoring the white background of the table; henceforth “coloured pixels”) are allocated to one of 36 “bins” each representing 10 degrees of the 360 degree HSV hue continuum, generating a histogram-like colour profile for each object. As each profile is unique, and based on the entire range of colours visible in each object, the model takes into account differences between uniformly and multicoloured objects. The resulting colour profile values provide input to the colour SOM.

Inputs to the shape SOM take into account both height/width and edge complexity. Height and width are calculated for a particular object in a particular region by locating the first and last coloured pixels along the vertical and horizontal axes, giving an approximation of an object’s aspect ratio. Note that these values are susceptible to variation from placement of the objects across trials; this reflects the fact that size alone has been found to be a poor primary indicator of category membership in English-speaking children (Landau, Smith, & Jones, 1988).

Edge complexity is an additive measure calculated by applying an edge-detecting Laplacian filter to the camera image, a technique commonly used for edge detection in computer vision. The resulting filtered image is then thresholded to reduce noise and non-black pixels are counted for each region, generating a value representing the overall edge complexity for each object. Thus, smooth objects with few edges give rise to a lower value than complex objects with many edges. Both the height/width and edge information provide the input to the shape SOM. In contrast to existing models of

categorisation, then, the current model's visual inputs were taken directly from the real-world referents encountered by children during the empirical task.

Similarly, auditory input consists of words spoken by the experimenter. Dragon Dictate speech recognition software is used during an initial training session to learn to recognise labels for known objects over several repetitions. For each taught label, a pool of nodes (in place of a SOM) responds uniquely. As new words are presented to the network, additional nodes are recruited. Critically, the model is not exposed to novel words before the experiments commence; thus, degree of the model's familiarity with labels reflects the degree of children's familiarity with labels. Every effort was made to ensure differences between external inputs to the iCub and the auditory and visual inputs encountered by children in Twomey & Horst (Paper 1, this thesis) were as similar as possible; however, obvious differences between the robot's sensors and actuators and children's perceptual and motor systems inevitably moderate results in any cognitive robotic study.

Input SOMs were initialised with random connection weights as per Equation 1:

Equation 1.

$$A_j = \sqrt{\sum_{i=0}^n (v_i - w_{ij})^2}$$

where A_j is the activity of a given unit after each iteration, v_i is input to that neuron, and w_{ij} is the connection weight between the input unit and the current (output) unit. The SOMs were initialised with random inputs in same range as the real input objects. For example, as mentioned above, colour input from the robot is processed into 36 bins (each representing a 10 degree section of HSV colour space) containing the relative proportion of pixels from an object with colours in that section of HSV space. Thus in

pre-training the SOM is shaped by generating random sequences of 36 numbers, which are then normalised. Each SOM is pre-trained to a neighbourhood size of 1 but remains plastic in subsequent use. On each pass, the Euclidean distance between a given input vector and each weight vector is calculated and its weights are modified to be more similar to that input vector. The output unit associated with the weight vector closest to the input vector is then activated. The weights of all output units are updated as follows:

Equation 2.

$$w_j = w_j + \theta(i, t)\alpha(t)(x - w_i)$$

where w_j is the weight vector of output neuron j , $\alpha(t)$ is a learning rate that decreases monotonically over time, and $\theta(j, t)$ defines the neighbourhood size (neighbourhood is a term defined by Kohonen referring to an area of the neural population physically surrounding j to which Equation 2 is applied). Note that the neighbourhood size also decreases over time to produce a winner-takes-all selection of a single output unit.

The three SOMs are linked to a central “hub” via Hebbian learning as in Equation 3:

Equation 3.

$$\Delta w_{ij} = \lambda x_i x_j (1 - w_{ij})$$

The normalised Hebbian update function where λ is a constant learning rate (0.05), x_i and x_j are two different unit values and Δw_{ij} is the change in the strength of the connection between them. Note that adaptive connections exist only between colour SOM units and word pool units, and between shape SOM units and word pool units. Inhibitory connections within the word pool and within each SOM are not adaptable.

Nodes within each SOM are connected by inhibitory weights with fixed weight values of -0.8. The spread of activation between the SOM units is governed by the

following equations:

Equation 4. The summation of internal and external input.

$$net_{input} = \alpha \sum (w_{ij}x_j) + (\varepsilon e_j)$$

Equation 5. The positive update rule, if $net_{input} > 0$.

$$\Delta x_i = (max - x_i)net_{input} - decay(x_i - rest)$$

Equation 6. The negative update rule, if $net_{input} < 0$

$$\Delta x_i = (x_i - min)net_{input} - decay(x_i - rest)$$

where α is the internal bias (0.1), and ε is the external bias (1.0), e_i is the external input to the j^{th} unit, max is a constant maximum (positive) level of activation for any unit (1.0), min is a constant minimum (negative) level of activation for any unit (-0.2), $decay$ is the rate of decay relative to the difference from rest (0.5), $rest$ is the resting level of activity for any unit (-0.1), x_i and x_j are two different unit values and w_{ij} is the connection weight between them. While the network is fairly robust to parameter variation, the values used here were chosen to be consistent with earlier work using similar, though hand designed rather than autonomously learned, structures (Burton, Bruce & Hancock, 1999; Burton, Bruce & Johnston, 1990; Burton, Young, Bruce, Johnston & Ellis, 1991).

Procedure, robot task

Prior to the experiment, to simulate the productive vocabulary of 30-month-old children, the SOMs were initially taught a label-exemplar pair for each known exemplar the robot would encounter during the experiment. Just as the children in the empirical task came to the experiment knowing that the COW was called “cow” and the SPOON was called “spoon,” the model was able to activate the correct label in response to presentation of the known exemplars it would see on each trial.

To this end, each known exemplar was placed individually in the center of

robot's field of vision on a white tabletop. The SOMs were allowed to settle, forming a unique "object" profile of winning neurons from the colour and shape SOMs for that object (equivalent to allowing children to look at the objects before asking "can you show me the *blicket*?") With the object still in view, the label SOM was presented with the appropriate label input for that exemplar. Associations between the visual and label input were formed and reinforced for each known exemplar. The amount of training given on each known object approximated a child hearing that object labeled 10,000 times in that context.

Figure 3 depicts the procedure for the robot task. The procedure was kept as close as practicably possible to that used in the empirical task, with the exception that for the purposes of this preliminary study, the robot encountered a single category (the *cheem* category, see Figure 3). In each condition, the robot was familiarised with the same stimuli encountered by children. Similarly, the robot task consisted of three phases: referent selection (six trials), recall test (three trials) and generalisation test (one trial). The robot received three known label trials and three novel label trials during referent selection.

On each referent selection trial (Trials 1 – 6, Fig. 3) the robot was presented with two known exemplars and one novel exemplar. Each known object activated the corresponding node in the label SOM (and no label node was activated for the novel object). With all three exemplars in view, the experimenter then presented the robot with either a known label (that is, pretrained, e.g., *fork*) or a novel label (that is, not trained, i.e. *cheem*.) Note that novel labels were completely novel; that is, the first time the robot was presented with a novel label was during referent selection. The robot's response was then determined by restricting the robot's field of vision to the target exemplar and examining the label node activated in response to the target. On any

given trial, the robot's response was considered correct if the activated label node matched the label previously given by the experimenter (e.g., if the “cow” node was activated when the cow was the target, or if the “*cheem*” node was activated when the novel exemplar was the target). When the target was novel, this in-the-moment linking of the novel label to the novel object was considered referent selection.









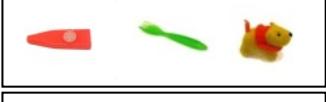
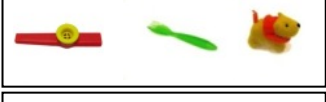









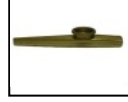
Trial	Narrow	Broad	Target
1.			<i>lion</i>
2.			<i>cheem</i>
3.			<i>cheem</i>
4.			<i>block</i>
5.			<i>cheem</i>
6.			<i>bus</i>
7.			<i>cheem</i>
8.			<i>cheem</i>
9.			<i>cheem</i>
10.			<i>cheem</i>

Figure 3. Procedure for robot task

Test trials in the robot task differed slightly from test trials presented to children. Because the robot encountered a single category, rather than three categories, it was not possible to give trials including three exemplars, one from each familiarised category. However, as noted above, the architecture allows the experimenter to query which label node is activated in response to a given stimulus. Thus, on the three recall test trials (trials 7-9, Fig. 3), the robot was presented with each familiarised novel exemplar individually, and the activated label node was recorded. Proportion of correct responses was then calculated. On the single generalisation trial (trial 10, Fig. 3), the generalisation object was presented individually, and again, the activated label node was recorded.

Novel category exemplars used in the robot task were identical to novel category exemplars used in the empirical task. Manipulations of category structure therefore reflected those in the empirical task. That is, in the *narrow* condition the robot was given three exemplars that differed in colour alone, and in the *broad* condition the robot was given exemplars that varied in colour, shape and size (see Fig. 4).

As in the child experiment, no feedback was given during or after referent selection and test trials. Thus, any learning that occurred did so in a non-ostensive context. Finally, the robot was run through each condition 12 times, to reflect the 12 children in each condition in the empirical task. In some models, in order to obtain a robust result, the simulation of an experiment is run many more times than the experiment itself, or similarly, the simulation is presented on with stimuli many more times on a single run than is a single participant in the experiment. Importantly, the robot did not require more exposures to the stimuli than the children in order for it to exhibit comparable behaviour. That is, both children and the robot encountered three novel exemplars per category.









	Referent Selection / Retention			Extension
<i>Narrow</i>	 small, yellow, plastic, white funnel	 small, red, plastic, white funnel	 small, blue, plastic, white funnel	 large, metal, brass funnel
<i>Broad</i>	 small, yellow, plastic, white funnel	 large, red, plastic, yellow funnel	 large, blue, plastic, spots, blue funnel	 large, metal, brass funnel

Figure 4. Novel stimuli used in the robot task

Results

Results from the robot task are depicted in Figure 5. Data from the referent selection phase of the robot task reflect data obtained by Twomey and Horst (2011). Specifically, during referent selection, in both *narrow* and *broad* conditions, the robot was able to reliably select the correct exemplar on known category trials (*narrow*: $t(11) = 16.40$, $p < .001$, $d = 9.89$; same results for *broad*: $t(11) = 16.40$, $p < .001$, $d = 9.89$). The robot also reliably selected the correct exemplar on novel category trials (*narrow*: $t(11) = 6.54$, $p < .001$, $d = 3.94$; *broad*: $t(11) = 2.44$, $p < .05$, $d = 1.47$). Thus, these preliminary data give evidence of referent selection in a non-ostensive context using a purely associative system.

At test, in the *narrow* condition, the robot selected the correct exemplar in response to newly fast-mapped novel names at levels approaching chance (0.33 for all known, novel and retention trials; $t(11) = 1.99$, $p = .071$, $d = 1.20$) but was unable to reliably generalise novel names to completely new exemplars (Wilcoxon $W = 33$, *ns*; note that a nonparametric test was used to analyse this binary variable). This pattern broadly reflects the empirical results and a computational dynamic neural field

replication (Twomey & Horst, Paper 1 & Paper 4, this thesis), however a full replication is required to establish whether the result for the retention trials is robust.

Finally, in the *broad* condition, the robot did not recall or generalise newly fast-mapped category labels (*narrow*: $t(11) = 1.34$, *ns*; *broad*: Wilcoxon $W = 63$, $p = .057$). Generalisation trials approached significance, however, and again, it is possible that given a full replication of the empirical task with three categories and 18 trials, the model would show robust word learning.

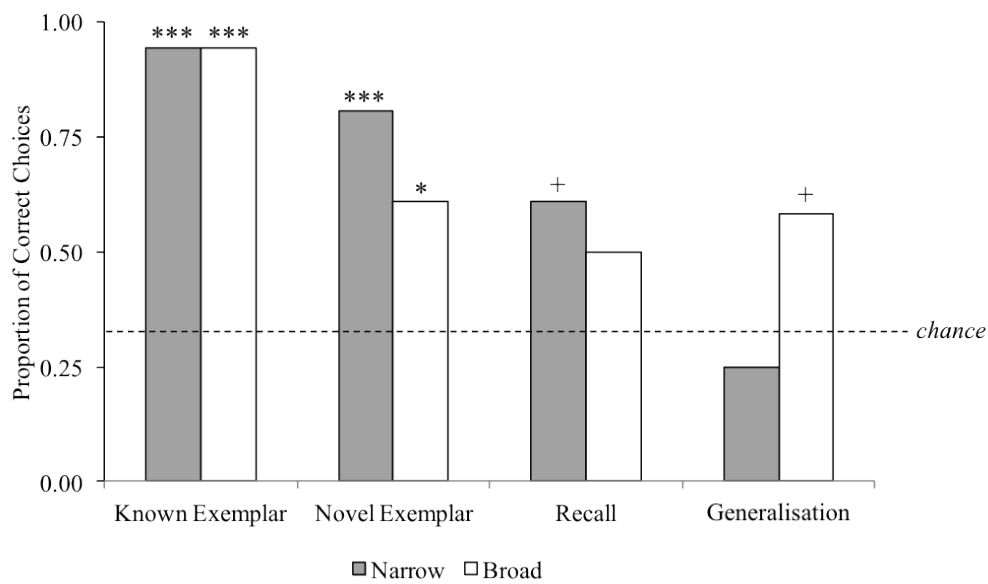


Figure 5. Results from the robot task. *** $p < .001$, * $p < .05$, + $p < .08$.

Discussion

The results from this preliminary study reflect the results reported by Twomey and Horst (2011). The robust demonstration of referent selection without supervision lays the foundation not only for a full replication of Twomey and Horst (2011), but also for further research into the mechanisms underlying fast mapping, word learning and categorisation.

Moreover, during the referent selection phase the model demonstrated in-the-moment categorisation. We go a step further than many existing models of categorisation, however, by embedding our simulation in an embodied system situated in a perceptual and temporal environment that closely reflects the environment of the child. Importantly, because of the differences in test trial design between the empirical and robot tasks any interpretation of these data must remain cautious. Nonetheless, the results point toward the exciting possibility of object category learning using an embodied, dynamic-associative system.

General Discussion

What does this robotic demonstration offer that existing empirical data or computational models do not? First, the project described here suggests a process-based account of the much-debated mechanisms underlying fast mapping: we have demonstrated that apparently inference-based referent selection (that is, using a strategy such as mutual exclusivity) can emerge for free from the dynamic interaction of the SOM-based architecture and the model's learning history (for a similar demonstration in a different dynamic connectionist model see Horst, McMurray & Samuelson, 2006; McMurray, et al., 2009).

Specifically, referent selection in this task context depended on inhibition. That is, in the central hub, strong excitatory connections between known labels and known exemplars inhibited the formation of new connections between novel labels and known exemplars, and between known labels and novel exemplars. When the model encountered a novel label, then, the only connection that was not the subject of inhibition was the potential mapping between novel label and novel object. This was the mapping that the model formed, reflecting children's ability to learn novel category labels in the absence of explicit teaching of labels via positive/negative feedback and/or

ostensive labelling. Thus, the robot exhibits a cognitive behaviour softly assembled from perceptual input (the stimuli presented to the robot) processed in the context of multiple timescales (pretrained vocabulary, in-task word learning), an account of fast mapping that would not be possible if only tested empirically.

Second, the iCub provides an appreciably more ecologically valid environment in which to situate what is fundamentally a model of neural dynamics. Specifically, the visual inputs to the system consist of image data, rather than the more abstract “neural activation” found in some purely computational models of cognitive development (e.g., Mareschal, French & Quinn, 2000; Rogers & McClelland, 2004; Twomey & Horst, Papers 4 & 5, this thesis; Samuelson, Perry & Spencer, 2011; amongst many others). Further, the stimuli presented to the iCub’s cameras were the same objects presented to the children by Twomey and Horst (2011) in the same physical context (on a white table, presented by the same experimenter), in the same testing timescale, following longer-term previous experience with known category exemplars and their labels. Thus, not only did we observe the emergence via simple associations of reliable referent selection and categorisation, but we also observed it in an environment much closer to that experienced by children than purely computational models can offer (Papers 4 and 5, this thesis). It is important to note, however, that “embodiment” as instantiated in this model is limited. Specifically, due to limitations imposed by time constraints, the robot did not receive proprioceptive feedback from the positioning of its limbs and torso, unlike children. This is problematic for the current paper, given that a similar system has shown an effect of body position on the robot’s (and children’s) ability to map words to objects via spatial locations. For a truly embodied replication, future work should focus on including proprioceptive feedback in the current system.

Third, in terms of categorisation, this model shows how categories can be scaffolded from the input and firmly grounded in environmental and temporal contexts. All the categories learned by the iCub were learned purely via real-time association of auditory labels with visual input: the robot had no in-built conceptual structure. So, for example, we had not programmed it to know that “spoon” was a utensil, or that “cow” was an animal; it simply learned to associate a cow-shaped object with the label “cow” (see also Horst et al., 2006; McMurray et al., 2009). Similarly, when the robot formed categories based on the novel exemplars it encountered during the familiarisation phase, it did so purely on the basis of visual similarity and a shared label. In line with existing research, these data suggest that children’s object categories can be perceptually-based (Gliga, Mareschal, & Johnson, 2008; Gogate & Hollich, 2010; Kovack-Lesh & Oakes, 2007; Rakison, 2000).

In order to firmly ground future projects in current work, initial research must focus on replicating the empirical data. However, the model as described here provides a foundation for future research, which will contribute to a number of research threads that have emerged in the field of cognitive development in recent years. Unlike most computational models, embodied robots allow us to investigate how cognitive abilities are shaped by and develop in the context of physical and social interaction with the environment. As such, the current project provides a methodological bridge between data from the word learning, categorisation and embodied cognition literatures. While the current work does not directly address social cueing, we believe the categorisation and fast mapping skills demonstrated here represent a crucial step toward such investigations.

The model presented here is perhaps the simplest example of the ERA architecture, yet the phenomena demonstrated are equally apparent in more complex

versions. Current and ongoing work with more complex examples of the model demonstrate the combination of cross-situational learning, bodily/spatial biases, fast-mapping, mutual exclusivity, and simple grammar learning to produce a learning system that is more than the sum of its parts (work in preparation). Further work is planned to explore developmental transitions (Morse, Belpaeme, Cangelosi, Floccia, 2011), social learning, and long-term learning.

In conclusion, increasing interest in computational models and cognitive processes has coincided with cognitive robotics' growing emphasis on the need to understand development. The converging interests of the two formerly remote fields have begun to produce exciting interdisciplinary collaborations such as the project presented here. Specifically, this chapter described an embodied robotic replication of a categorisation experiment conducted with young children which demonstrated that embodied computational approaches to understanding cognition can go far in resolving longstanding debates as to the processes that drive cognitive development. The preliminary data presented here are just one example of the benefits to be gained from the integration of research in cognitive development, computational modelling and developmental robotics (Simmering, et al., 2010). In embarking on interdisciplinary projects, each discipline stands to gain from the focus on the dynamic and temporally-contingent coupling of cognition, body and environment.

Acknowledgements

This research was supported by an EU Cognition II grant to K.E.T and A.F.M and ESRC +3 Quota award ES/H014942/ to K.E.T.

References

- Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, 72(2), 416-430. DOI: 10.1111/1467-8624.00287
- Alvy Ray, S. (1978). Colour gamut transform pairs. *SIGGRAPH Computer Graphics*, 12(3), 12-19. DOI: 10.1145/965139.807361
- Aslin, R., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *Journal of Biomedical Optics*, 10(1). DOI: 10.1117/1.1854672
- Aslin, R., & Salapatek, P. (1975). Saccadic localization of visual targets by the very young human infant. *Attention, Perception, & Psychophysics*, 17(3), 293-302. DOI: 10.3758/BF03203214
- Au, T. K. F., & Glusman, M. (1990). The principle of mutual exclusivity in word learning - to honor or not to honor. *Child Development*, 61(5), 1474-1490. DOI: 10.2307/1130757
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3. PMID:2229167
- Baillargeon, R. e., & Wang, S.H. (2002). Event categorization in infancy. *Trends in Cognitive Sciences*, 6(2), 85-93. DOI: 10.1016/S1364-6613(00)01836-2
- Berthouze, L., & Metta, G. (2005). Epigenetic robotics: modelling cognitive development in robotic systems. *Cognitive Systems Research*, 6, 189-192. DOI: 10.1016/j.cogsys.2004.11.002
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

- Burton, A. M., V. Bruce & Hancock, P. J. B. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science* 23(1), 1-31. DOI: 10.1207/s15516709cog2301_1
- Burton, A. M., V. Bruce & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology* 81(3), 361-380. DOI: 10.1111/j.2044-8295.1990.tb02367.x
- Burton, A. M., A. W. Young, Bruce, V., Johnston, R. A., & Ellis, A. W (1991). Understanding covert recognition. *Cognition* 39(2): 129-166. DOI: 10.1016/0010-0277(91)90041-2
- Cangelosi, A., Belpaeme, T., Sandini, G., Metta, G., Fadiga, L., Sagerer, G., et al. (2008, 12-15 March). *The ITALK Project: Integration and transfer of action and language knowledge*. Paper presented at the Third ACM/IEEE International Conference on Human Robot Interaction Amsterdam.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development*, 15, 17-29.
- Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfield & S. A. Gelman (Eds.), *Mapping the Mind: Domain Spicificity in Cognition and Culture*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511752902.008
- Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, 38(6), 967-978. DOI: 10.1037/0012-1649.38.6.967
- Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development*, 50(2), 419-424. DOI: 10.2307/1129417

- Colunga, E., & Smith, L. B. (2003). The emergence of abstract ideas: evidence from networks and babies. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358(1435), 1205-1214. DOI: 10.1098/rstb.2003.1306
- Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: the self-organization of skilled noun learning. *Developmental Science*, 11(2), 195-203. DOI: 10.1111/j.1467-7687.2007.00665.x
- Cowan, N. (2003). Comparisons of developmental modelling frameworks and levels of analysis in cognition: connectionist and dynamic systems theories deserve attention, but don't yet explain attention. *Developmental Science*, 6(4), 440-447. DOI: 10.1111/1467-7687.00299
- Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science*, 298(5600), 2013-2015. DOI: 10.1126/science.1077066
- Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development*, 74(1), 168-178. DOI: 10.1111/1467-8624.00528
- Diesendruck, G., & Graham, S. A. (2010). Kind matters: A reply to Samuelson & Perone. *Cognitive Development*, 25, 149-153. DOI: 10.1016/j.cogdev.2010.02.003
- Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology*, 58(3), 418-431. DOI: 10.1006/jecp.1994.1043
- Fantz, R. L., & Fagan, J. F. (1975). Visual attention to size and number of pattern details by term and preterm infants during first 6 months. *Child Development*, 46(1), 3-18. DOI: 10.2307/1128828

- Faubel, C., & Schöner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks*, 21(4), 562-576. DOI 10.1016/j.neunet.2008.03.007
- Fava, E., Hull, R., & Bortfeld, H. (2011). Linking behavioural and neurophysiological indicators of perceptual tuning to language. *Frontiers in Psychology*, 2. PMID:21738514
- Feldman, J., & Bailey, D. (2000). Layered hybrid connectionist models for cognitive science. *Hybrid neural systems*, 14-27.
- Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behaviour and Development*, 7(1), 19-25. DOI: 10.1016/S0163-6383(84)80019-3
- Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of colour changes with colour term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), 18221-18225. DOI: 10.1073/pnas.0809952105
- French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3-to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology-General*, 133(3), 382-397. DOI: 10.1037/0096-3445.133.3.382
- Gershkoff-Stowe, L., & Smith, L. B. (2004). Shape and the first hundred nouns. *Child Development*, 75(4), 1098-1114. DOI: 10.1111/j.1467-8624.2004.00728.x
- Gliga, T., Mareschal, D., & Johnson, M. H. (2008). Ten-month-olds' selective use of visual dimensions in category learning. *Infant Behaviour & Development*, 31(2), 287-293. DOI: 10.1016/j.infbeh.2007.12.001

- Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117(2), 496. DOI: 10.1037/a0019049
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28(1), 99-108. DOI: 10.1037/0012-1649.28.1.99
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M., & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23-45. doi: doi:10.1017/S030500090001271X
- Gredebäck, G., Johnson, S. P., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology*, 35(1), 1-19. DOI: 10.1080/87565640903325758
- Grieser, D. A., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, 25(4), 577. DOI: 10.1037/0012-1649.25.4.577
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23-B34. DOI: 10.1016/S0010-0277(02)00186-5
- Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, 53(4), 310-344. DOI: S0010-0285(06)00025-9
10.1016/j.cogpsych.2006.04.003
- Heibeck, T. H., & Markman, E. M. (1987). Word learning in children - an examination of fast mapping.. *Child Development*, 58(4), 1021-1034. DOI: 10.2307/1130543
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., et al. (2000). *Breaking the language barrier: An emergentist coalition model for*

the origins of word learning. Monographs of the Society for Research in Child Development, Vol. 65(3)

- Hollich, G., & Prince, C. G. (2009). Comparing infants' preference for correlated audiovisual speech with signal-level computational models. *Developmental Science*, 12(3), 379-387. DOI: 10.1111/j.1467-7687.2009.00823.x
- Horst, J. S., Ellis, A. E., Samuelson, L. K., Trejo, E., Worzalla, S. L., Peltan, J. R., et al. (2009). Toddlers can adaptively change how they categorize: same objects, same session, two different categorical distinctions. *Developmental Science*, 12(1), 96-105. DOI: 10.1111/j.1467-7687.2008.00737.x
- Horst, J. S., Oakes, L. M., & Madole, K. L. (2005). What does it look like and what can it do? Category structure influences how infants categorize. *Child Development*, 76(3), 614-631. DOI: 10.1111/j.1467-8624.2005.00867.x
- Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128-157. DOI:10.1080/15250000701795598
- Horst, J. S., Samuelson, L. K., Kucker, S. C., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, 118(2), 234-244. DOI: 10.1016/j.cognition.2010.10.015
- Horst, J. S., Scott, E. J., & Pollard, J. A. (2011). The role of competition in word learning via referent selection. *Developmental Science*, 13(5), 706-713. DOI: 10.1111/j.1467-7687.2009.00926.x
- Houston-Price, C., Caloghris, Z., & Raviglione, E. Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125-150. DOI: 10.1111/j.1532-7078.2009.00009.x
- Iida, F., Pfeifer, R., Steels, L., and Kuniyoshi, Y., eds. (2004). *Embodied artificial intelligence*. Berlin: Springer.

- James, D. K. (2010). Fetal learning: a critical review. *Infant and Child Development*, 19(1), 45-54. DOI: 10.1002/icd.653
- Jaswal, V. K., & Markman, E. M. (2001). Learning proper and common names in inferential versus ostensive contexts. *Child Development Vol 72(3) May-Jun 2001*, 768-786. DOI: 10.1111/1467-8624.00314
- Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research*, 1299, 17-32. DOI: 10.1016/j.brainres.2009.07.008
- Johnson, S. P. (2010a). Development of visual perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 515-528. DOI: 10.1002/wcs.128
- Johnson, S. P. (2010b). Perceptual completion in infancy *Neoconstructivism: the New Science of Cognitive Development*. New York: Oxford University Press.
- Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1-3), 1-6. DOI: 10.1016/S0925-2312(98)00030-7
- Kovack-Lesh, K. A., Horst, J. S., & Oakes, L. M. (2008). The cat is out of the bag: The joint influence of previous experience and looking behaviour on infant categorization. *Infancy*, 13(4), 285-307. DOI: 10.1080/15250000802189428
- Kovack-Lesh, K. A., & Oakes, L. M. (2007). Hold your horses: How exposure to different items influences infant categorization. *Journal of Experimental Child Psychology*, 98(2), 69-93. DOI: 10.1016/j.jecp.2007.05.001. DOI: 10.1016/j.jecp.2007.05.001
- Kucker, S. C., & Samuelson, L. K. (2011). The first slow step: Differential effects of object and word-form familiarization on retention of fast-mapped words. *Infancy*.

- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. DOI: 10.1016/0885-2014(88)90014-7
- Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology*, 36(5), 635. DOI: 10.1037/0012-1649.36.5.635
- Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3-to 4-month-olds' sequential category learning. *Cognitive Science*, 26(3), 377-389. DOI: 10.1207/s15516709cog2603_8
- Markman, E. M. (1994). Constraints on word meaning in early language acquisition. *Lingua*, 92(1-4), 199-227. DOI: 10.1016/0024-3841(94)90342-5
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology*, 20(2), 121-157. DOI: 10.1016/0010-0285(88)90017-5
- Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, 14(1), 60-76. DOI: 10.1080/15250000802569702
- Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, 117(1), 1-31. DOI: 10.1037/a0018130. DOI: 10.1037/a0018130
- McMurray, B., Horst, J. S., & Samuelson, L. K. (under review). Using your lexicon at two timescales: investigating the interplay of word learning and word recognition. *Psychological Review*. DOI: 10.1093/acprof:oso/9780195300598.003.0011

- McMurray, B., Horst, J. S., Toscano, J., & Samuelson, L. K. (2009). Connectionist Learning and Dynamic Processing: Symbiotic Developmental Mechanism. In J. P. Spencer, M. Thomas & J. McClelland (Eds.), *Towards a New Grand Theory of Development? Connectionism and Dynamic Systems Theory Reconsidered* (pp. 218-249). New York: Oxford University Press. DOI: 10.1037/0033-295X.85.3.207
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review; Psychological Review*, 85(3), 207. DOI: 10.1037/0033-295X.85.3.207
- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198(4312), 75-78. DOI: 10.1126/science.198.4312.75. 10.1126/science.198.4312.75
- 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. Vol. 54(3-4). DOI: 10.2307/1166130
- Metta, G., Natale, L., Nori, F., Sandini, G., Vernon, D., Fadiga, L., et al. (2010). The iCub humanoid robot: An open-systems platform for research in cognitive development. *Neural Networks*, 23(8-9), 1125-1134. DOI: 10.1016/j.neunet.2010.08.010
- Morse, A. F., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2010a). *Thinking with your body: Modelling spatial biases in categorization using a real humanoid robot*. Paper presented at the 32nd Annual Conference of the Cognitive Science Society, Austin, TX.

- Morse, A. F., de Greeff, J., Belpaeme, T., & Cangelosi, A. (2010b). Epigenetic Robotics Architecture (ERA). *Autonomous Mental Development, IEEE Transactions on*, 2(4), 325-339.
- Morse, A. F., Belpaeme, T., Cangelosi, A., & Floccia, C. (2011). In: L. Carlson, C. Hölscher, & T. Shipley (eds.). Modelling U shaped performance curves in ongoing development. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 3034-3039). Austin, TX: Cognitive Science Society.
- Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, 6(4), 413-429. DOI: 10.1111/1467-7687.00296
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, 3. DOI: 10.3389/fpsyg.2012.00041. DOI: 10.3389/fpsyg.2012.00041
- Murphy, G. L. (2003). Ecological validity and the study of concepts. *Psychology of Learning and Motivation* 43, 1-41. San Diego: Academic Press.
- Murphy, G. L. (2004). *The big book of concepts*. Cambridge, Mass: The MIT Press.
- Namy, L. L. (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy*, 2(1), 73-86. DOI: 10.1207/S15327078IN0201_5
- Nosofsky, R. M., & Stanton, R. D. (2006). Speeded old-new recognition of multidimensional perceptual stimuli: Modelling performance at the individual-participant and individual-item levels. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 314-334. DOI: 10.1037/0096-1523.32.2.314

- Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3), 255-268. DOI: 10.1080/15248371003699977
- Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants' object examining: Habituation and categorization. *Cognitive Development*, 6(4), 377-392. DOI: 10.1016/0885-2014(91)90045-F
- Peniak, M., Morse A.F., Larcombe, C., Ramirez-Contla, S., & Cangelosi, A., (2011). *Aquila: An open-source GPU-accelerated toolkit for cognitive robotics research..* International Joint Conference on Neural Networks (IJCNN) 2011. San Jose, California, July 31- August 5, 2011
- Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, 120(3), 302-321. DOI: 10.1016/j.cognition.2010.11.015
- Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, 21(12), 1894-1902. DOI: 10.1177/0956797610389189
- Pfeifer, R., & Bongard, J. (2007). *How the body shapes the way we think*. Cambridge, Mass.: MIT Press.
- Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2), 665-681. DOI: 10.1016/j.cognition.2007.04.003
- Prince, C. G., & Hollich, G. J. (2005). Synching models with infants: a perceptual-level model of infant audio-visual synchrony detection. *Cognitive Systems Research*, 6(3), 205-228. DOI: 10.1016/j.cogsys.2004.11.006

- Quine, W. V. O. (1960). *Word and object*. Cambridge: MA: MIT Press.
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition*, 27(2), 145-179. DOI: 10.1016/0010-0277(87)90017-5
- Quinn, P. C. (2005). Young infants' categorization of human versus nonhuman animals: Roles for knowledge access and perceptual process. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building Object Categories in Developmental Time* (pp. 107-130). Mahwah, NJ: Lawrence Erlbaum Associates.
- Quinn, P. C., Burke, S., & Rush, A. (1993). Part-whole perception in early infancy: Evidence for perceptual grouping produced by lightness similarity. *Infant Behavior and Development*, 16(1), 19-42. DOI: 10.1016/0163-6383(93)80026-5
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, 22(4), 463-475. DOI: 10.1068/p220463
- Rakison, D. H. (2000). When a rose is just a rose: The illusion of taxonomies in infant categorization. *Infancy*, 1(1), 77-90. DOI:10.1207/S15327078in0101_07
- Rakison, D. H., & Yermolayeva, Y. (2011). How to identify a domain-general learning mechanism when you see one. *Journal of Cognition and Development*, 12(2), 134-153. DOI: 10.1080/15248372.2010.535228
- Rao, R. P. N. (2004). Bayesian computation in recurrent neural circuits. *Neural computation*, 16(1), 1-38. DOI: 10.1162/08997660460733976
- Regier, T., Kay, P., & Khetarpal, N. (2007). Colour naming reflects optimal partitions of colour space. *Proceedings of the National Academy of Sciences of the United States of America*, 104(4), 1436-1441. DOI:10.1073/pnas.0610341104. DOI: 10.1073/pnas.0610341104

- Ribar, R. J., Oakes, L. M., & Spalding, T. L. (2004). Infants can rapidly form new categorical representations. *Psychonomic Bulletin & Review*, 11(3), 536-541. DOI: 10.3758/BF03196607
- Riches, N. G., Tomasello, M., & Conti-Ramsden, G. (2005). Verb learning in children with SLI: frequency and spacing Effects. *Journal of Speech, Language, and Hearing Research*, 48, 1397-1411. DOI: 10.1044/1092-4388(2005/097)
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rogers, T. T., Rakison, D. H., & McClelland, J. L. (2004). U-Shaped curves in development: A PDP Approach. *Journal of Cognition and Development*, 5(1), 137-145. DOI: 10.1207/s15327647jcd0501_14
- Samuel, S. (1978). Measurement of infant visual acuity from pattern reversal evoked potentials. *Vision Research*, 18(1), 33-39. DOI: 10.1016/0042-6989(78)90074-3
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, 38(6), 1016-1037. DOI: 10.1037//0012-1649.38.6.1016.
- Samuelson, L. K., & Horst, J. S. (2008). Confronting complexity: insights from the details of behaviour over multiple timescales. *Developmental Science*, 11(2), 209-215. DOI 10.1111/j.1467-7687.2007.00667.x.
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, 6(12). DOI: 10.1371/journal.pone.0028095
- Schutte, A. R., & Spencer, J. P. (2010). Filling the gap on developmental change: Tests of a dynamic field theory of spatial cognition. *Journal of Cognition and Development*, 11(3), 328-355. DOI: 10.1080/15248371003700007 Pii 925511148

- Shahidullah, S., & Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Human Development*, 36(1), 13-26. DOI: 10.1016/0378-3782(94)90029-9
- Simmering, V. R., Triesch, J., Deak, G. O., & Spencer, J. P. (2010). A dialogue on the role of computational modelling in developmental science. *Child Development Perspectives*, 4(2), 152-158. DOI: 10.1111/j.1750-8606.2010.00134.x
- Slater, A., Quinn, P. C., Kelly, D. J., Lee, K., Longmore, C. A., McDonald, P. R., et al. (2010). The shaping of the face space in early infancy: Becoming a native face processor. *Child Development Perspectives*, 4(3), 205-211. DOI: 10.1111/j.1750-8606.2010.00147.x
- Slater, A., Von der Schulenburg, C., Brown, E., Badenoch, M., Butterworth, G., Parsons, S., et al. (1998). Newborn infants prefer attractive faces. *Infant Behaviour and Development*, 21(2), 345-354. DOI: 10.1016/S0163-6383(98)90011-X
- Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, N. Akhtar, M. Tomasello & G. Hollich (Eds.), *Becoming a Word Learner: A Debate on Lexical Acquisition* (pp. 51-80). New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780195130324.003.003
- Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: contextually cued attention and early word learning. *Cognitive Science*, 34(7), 1287-1314. DOI 10.1111/j.1551-6709.2010.01130.x
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13-19. DOI: 10.1111/1467-9280.00403

- Smith, L. B., & Samuelson, L. K. (2010). Objects in space and mind: From reaching to words. In K. S. Mix, L. B. Smith & M. Gasser (Eds.), *The spatial foundations of language and cognition*. New York: Oxford University Press.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411. DOI: 10.1037/0278-7393.24.6.1411
- Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38(2), 179-211. DOI: 10.1016/0010-0277(91)90051-5
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89-96. DOI: 10.1111/j.1467-7687.2007.00569.x
- Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science*, 6(4), 392-412. DOI: 10.1111/1467-7687.00295
- Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass.: MIT Press.
- Travis, K. E., Leonard, M. K., Brown, T. T., Hagler, D. J., Curran, M., Dale, A. M., et al. (2011). Spatiotemporal neural dynamics of word understanding in 12- to 18-month-old-infants. *Cerebral Cortex*, 21(8), 1832-1839. DOI: 10.1093/cercor/bhq259
- Twomey, K. E., & Horst, J. S.(Paper 1, this thesis). That's more like it: Multiple exemplars facilitate word learning.
- Twomey, K. E., & Horst, J. S.(Paper 4, this thesis). All things considered: Dynamic Field Theory captures effect of categories on children's word learning.

- Twomey, K. E., & Horst, J. S. (Paper 5, this thesis). Testing a dynamic neural field model of young children's category labelling.
- Vernon, D., Metta, G., & Sandini, G. (2007). A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents. *IEEE Transactions on Evolutionary Computation*, 11(2), 151-180.
- Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109(1), 163-167. DOI: 10.1016/j.cognition.2008.07.013
- Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology*, 43(3), 217-242. DOI: 10.1006/cogp.2001.0764
- Westermann, G., & Mareschal, D. (2009). Modelling the transition from perceptual to conceptual organization. In J. Mayor, N. Ruh & K. Plunkett (Eds.), *Connectionist models of behaviour and cognition* (Vol. 18, pp. 153-164).
- Woodward, A. L., & Markman, E. M. (1991). The mutual exclusivity bias in young children's word learning. *Developmental Review*, 11(2), 137-163. DOI: 10.1016/0273-2297(91)90005-9
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, 30(4), 553-566. DOI: 10.1037/0012-1649.30.4.553
- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47(5), 1220-1229. DOI: 10.1037/a0024023

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference.

Psychological Review, 114(2), 245. DOI: 10.1037/0033-295X.114.2.245

Younger, B. A., & Cohen, L. B. (1986). Developmental change in infants' perception of

correlations among attributes. *Child Development*, 57(3), 803-815. DOI:

10.2307/1130356

Zeschel, A., & Tuci, E. (2011). From symbol grounding to socially shared embodied

language knowledge. *Frontiers in Computational Neuroscience*. DOI:

10.3389/conf.fncom.2011.52.00025. DOI: 10.3389/conf.fncom.2011.52.00025

Ziemke, T. (2003). What's that thing called embodiment? In: Alterman & Kirsh (eds.)

Proceedings of the 25th Annual Conference of the Cognitive Science Society (pp.

1134-1139). Mahwah, NJ: Lawrence Erlbaum.

Afterword

Mapping the journey from languageless infant to fluent adult has been the objects of centuries of philosophical and scientific investigation (Aristotle, 335BC/1984; Quine, 1960; St Augustine, 397/1986). The papers presented in this thesis address the beginnings of this fundamentally human adventure – children’s early word learning. Earlier studies have focused on children’s ability to fast-map labels to entities in the real world (Carey & Bartlett, 1978). More recently, the interplay between labels and children’s environment have been the focus of both empirical and computational investigations (Axelsson, Churchley, & Horst, 2012; Houston-Price, Caloghris, & Raviglione, 2010; McMurray, Horst, & Samuelson, 2012; L. K. Samuelson, Smith, Perry, & Spencer, 2011). These studies take up this thread, investigating the effect of variation in what children see and the labels they learn for categories of object and action.

The empirical data presented here are in line with a dynamic systems view of word learning, in which in-the-moment fast mapping, word learning, environmental context and learning history all constitute components of one constantly interactive system, operating over different timescales and supporting the emergence of stable behaviours (Colunga & Smith, 2008; Larissa K. Samuelson & Horst, 2008; Thelen & Smith, 1994). On this view, “knowledge is process” – what children appear to know at a given time is the emergent product of the interactions-in-time between these component. In the context of this thesis, dynamic systems theory makes several testable predictions.

First, the DNF model suggests that word learning is a slow and two-stage process (see also Munro, Baker, McGregor, Docking, & Arculi, 2012). Children must encode both label and object across encounters. Although children may not robustly

learn a label-object mapping after a single exposure, then, they learn something – perhaps explaining children’s ability to extend novel labels without retention in Paper 1. This subtle difference between word learning and generalization, as well as the extent of children’s learning on each exposure, remain unaddressed. Understanding the moment-by-moment development of a word-label mapping is important for our interpretation of the existing word learning literature – what this thesis does show, however, is that lack of retention need not indicate lack of learning,

Second, DST states that cognition is situated both in the environment and in the body. In terms of the environment, this thesis demonstrates that perceptual variability does affect category label learning – indeed, in line with noun generalization studies, Paper 1 demonstrates that moderate exemplar variability *helps* object category label learning. Paper 2 hints at a similar phenomenon in action category label learning, pointing to interesting future work in comparing the processes underlying noun and verb learning. What these papers do not address, however, is how other environmental variability might affect word learning. Specifically, according to DST, stable behaviour emerges after a period of instability, or chaotic behaviour. Evidence from the adult domain suggest that manipulating the amount of environmental variability precipitates faster emergence of stable behaviours such as short-cut solution to solving simple puzzles. If learning – and specifically, word learning – can be described as a dynamic system, a comparable increase in environmental variability should facilitate the emergence of stable word learning behaviour in adults. What, then, would the effect be in word learning in children? Further, would variability in different modalities have differential effects?

Paper 7 presented pilot data from an embodied simulation of word learning. Although limited by time constraints, the paper provides exciting evidence that an

embodied robotic system can learn labels for object categories using low-level associative processes. However, although a similar system has been used to simulate children's spatial binding of labels to single objects (Morse, Belpaeme, Cangelosi, & Smith, 2010), the current system did not receive haptic feedback, so did not constitute the most stringent test possible of embodied categorization. This paper therefore provides the groundwork for an exciting future challenge: can a fully-embodied humanoid robotic system learn to label object categories?

It appears this thesis raises as many questions as it answers – and rightly so. In light of recent rapid advances in means of studying cognition and development and new insights from the integration of computational and empirical research, new questions emerge every day. The answers to these questions, however, will not only bring us a step closer to understanding how and where language acquisition begins, but will also provide a rich and informative picture of the temporal, cognitive and environmental landscape through which the journey takes us.

References

- Aristotle (1984). *The complete works of Aristotle: The revised Oxford translation*.
(Barnes, J., & others, Trans.). Princeton, NJ: Princeton University Press.
(Original work 335BC).
- Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time:
Why ostensive naming facilitates word learning. *Frontiers in Psychology*, 3.
- Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on
Child Language Development*, 15, 17–29. DOI:10.1037/0012-1649.38.6.967
- Colunga, E., & Smith, L. B. (2008). Knowledge embedded in process: the self-
organization of skilled noun learning. *Developmental Science*, 11(2), 195–203.
DOI:10.1111/j.1467-7687.2007.00665.x
- Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience
shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125–
150.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from
the interaction of online referent selection and slow associative learning.
Psychological Review, 119(4), 83877. DOI: 10.1037/a0029872
- Morse, A. F., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2010). Thinking with your
body: Modelling spatial biases in categorization using a real humanoid robot.
In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual
Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science
Society.
- Munro, N., Baker, E., McGregor, K., Docking, K., & Arculi, J. (2012). Why word
learning is not fast. *Frontiers in Psychology*, 3.

- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge: MIT Press.
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PloS one*, 6(12), e28095.
- Samuelson, Larissa K., & Horst, J. S. (2008). Confronting complexity: insights from the details of behavior over multiple timescales. *Developmental Science*, 11(2), 209–215. doi:10.1111/j.1467-7687.2007.00667.x
- St Augustine (1986). *The confessions of St. Augustine* (R. Warner, Trans.). New Kensington, PA: Whitaker House. (Original work 397AD)
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, Mass.: MIT Press.