COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1125–1130

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01132.x

# Mechanisms of Cognitive Development: Domain-General Learning or Domain-Specific Constraints?

Vladimir M. Sloutsky

*Center for Cognitive Science, The Ohio State University*

The issue of how people acquire knowledge in the course of individual development has fascinated researchers for thousands of years. Perhaps the earliest recorded effort to put forth a theoretical account belongs to Plato, who famously advocated the idea that knowledge of many abstract categories (e.g., ''equivalence'') is innate. Although Plato argued with his contemporaries who advocated the empirical basis of knowledge, it was the British empiricists who most forcefully put forth the idea of the empirical basis of knowledge, with John Locke offering the famous ''tabula rasa'' argument.

The first comprehensive psychological treatment of the problem of knowledge acquisition was offered by Piaget (1954), who suggested that knowledge emerges as a result of interactions between individuals and their environments. This was a radical departure from both extreme nativism and extreme empiricism. However, these ideas, as well those of empiricist-minded behaviorists, fell short of providing a viable account of many human abilities, most notably, language acquisition.

This inability prompted Chomsky to propose an argument that language cannot be acquired from the available linguistic input because it does not contain enough information to enable the learner to recover a particular grammar, while ruling out alternatives (Chomsky, 1980). Therefore, some knowledge of language must be innate to enable fast, efficient, and invariable language learning under the conditions of the impoverished linguistic input. This argument (i.e., known as the Poverty of the Stimulus argument) has been subsequently generalized to perceptual, lexical, and conceptual development. The 1990 Special Issue of *Cognitive Science* is an example of such generalization.

The current Special Issue on the mechanisms of cognitive development has arrived exactly 20 years after the first Special Issue. In the introduction to the 1990 Special Issue of *Cognitive Science*, Rochel Gelman stated:

Correspondence should be sent to Vladimir M. Sloutsky, Center for Cognitive Science, 208C Ohio Stadium East, 1961 Tuttle Park Place, Ohio State University, Columbus, OH 43210. E-mail: sloutsky.1@osu.edu

> Experience is indeterminant or inadequate for the inductions that children draw from it in that, even under quite optimistic assumptions about the nature and extent of the experiences relevant to a given induction, the experience is not, in and of itself, sufficient to justify, let alone compel, the induction universally drawn from it in the course of development. For example, there is nothing in the environment that supports a child's conclusion that the integers never end. (R. Gelman, 1990, p. 4)

If input is too impoverished to constrain possible inductions and to license the concepts that we have, the constraints must come from somewhere. It has been proposed that these constraints are internal—they come from the organism in the form of knowledge of ''core'' domains, skeletal principles, biases, or conceptual assumptions. To be useful in solving the indeterminacy problem, these constraints have to be (a) top-down, with higher-levels of abstraction appearing prior to lower levels (i.e., elements of an abstract structure must guide processing of specific instances), (b) a priori (i.e., these constraints have to precede learning rather than being consequence of learning), and (c) domain-specific (because generalizations in the domain of number differ drastically from those in the domain of biology, the principles guiding these generalizations should differ as well).

Formally, the Poverty of the Stimulus argument has the following structure: If (a) correct generalizations require many constraints and (b) the environment provides few, then (c) the constraints enabling correct generalizations do not come from the environment. While this argument is formally valid, its premise (b) and its conclusion (c) are questionable. Most important, do we know that the environment truly provides *few* constraints? And *how* do we know that?

The research featured in this Special Issue proposes an alternative way of meeting the challenge of understanding cognitive development. Instead of assuming top-down, a priori, domain-specific constraints, this research tries to understand how domain-general learning mechanisms may enable acquisition of knowledge for an organism functioning in an information-rich environment.

Chater and Christiansen (2010) focus on language learning and evolution and propose two critical ideas: (a) the idea of language adapting to biological machinery existing prior to the emergence of language and (b) the idea of ''C-induction.'' First, they argue that there is no credible account of how a richly structured, domain-specific, innate Universal Grammar could have evolved. They suggest that the solution to the logical problem of language evolution requires abandoning the notion of a domain-specific and innate Universal Grammar. As part of their second argument, Chater and Christiansen (2010) offer a critical distinction between ''natural'' and ''cultural'' induction (i.e., N-induction and C-induction). N-induction involves the ability to understand the natural world, whereas C-induction involves the ability to coordinate with other people. They argue that the problem of language acquisition has been traditionally misconstrued as a solution to an extremely difficult N-induction problem (i.e., the discovery of abstract syntax); however, according to the authors, the problem should be construed as a much easier problem of C-induction. Instead of inducing an arbitrary set of constraints (i.e., the problem of N-induction), individuals simply have to make the same guesses as everyone else. Crucially, this process of C-induction is made easier by

the fact that the others have the same biases as the learner and because language has been shaped by cultural evolution to fit those exact biases. Chater and Christiansen (2010) further suggest that the same line of argumentation is likely to extend to other kinds of development for which the learning of a culturally mediated system of knowledge is important.

Johnson's (2010) paper focuses on perceptual development. Perception has been at the center of our attempts to understand sources and origins of knowledge: How do people parse cluttered and occluded visual experience into separable objects? Does this ability develop over time through experience and learning, or is it based on some form of a priori knowledge (e.g., such as knowledge of objects)? In contrast to those advocating innate knowledge of objects, Johnson (2010) argues that there is no need to posit such innate knowledge. In his view, although some components of object perception (e.g., edge detection) may emerge from prenatal development (or even prenatal learning), other major components of object perception (e.g., perception of objects over occlusion) develop postnatally. According to Johnson's (2010) developmental proposal, initially perception of occluded objects requires support from multiple features, including the size of the occluding gap, the alignment of edges, and common motion. In the course of development, infants learn to perceive occluded objects independently of these features.

Kemp, Goodman, and Tenenbaum (2010) discuss how people learn about causal systems and generalize this knowledge to new situations. In particular, having learned that drug D has side effect E in person P, the learner may eventually generalize this knowledge to conclude that drugs have side effects on people. How is this learning achieved? One possible way of solving this problem is for the learner to have a highly constrained hypothesis space, specific to each knowledge domain. This fact has been at the heart of the nativist proposals arguing for innate sets of constraints specific to certain domains of knowledge. Although Kemp et al. (2010) agree that constraints are important for learning, they propose that these constraints do not have to be a priori—children can learn inductive constraints in some domains—and that these constraints subsequently support rapid learning within these domains. They develop and test a computational model of causal learning, demonstrating that constraints can be acquired and later used to facilitate learning of new causal structures. The critical idea is that when learners first encounter a new inductive task, their hypothesis space with respect to this task could be relatively broad and unconstrained. However, after experiencing several induction problems from that family, they induce a schema, or a set of abstract principles describing the structure of tasks in the family. These abstract principles constrain the hypotheses that learners apply to subsequent problems from the same family, and allow them to solve these problems given just a handful of relevant observations.

Sloutsky's (2010) paper discusses the development of concepts. It has been widely acknowledged that concepts allow nontrivial generalizations (e.g., that plants and animals are alive) and that concepts support reasoning. How do people acquire concepts? And given that generalizations are constrained (people generalize the property of being alive to garden flowers, but not to plastic flowers), where do these constraints come from? Unlike proposals arguing for a priori constraints, Sloutsky's (2010) proposal attempts to link conceptual development to a more general ability to form perceptual categories, which exhibits early developmental onset and is present across a wide variety of species. Sloutsky (2010) argues

that conceptual development progresses from simple perceptual grouping to highly abstract scientific concepts. This proposal of conceptual development has four parts. First, it is argued that categories in the world have different structure. Second, there might be different learning systems (subserved by different brain mechanisms) that evolved to learn categories of differing structures. Third, these systems exhibit differential maturational course, which affects how categories of different structures are learned in the course of development. And finally, an interaction of these components may result in the developmental transition from perceptual groupings to more abstract concepts.

Smith, Colunga, and Yoshida (2010) consider the role of attention in acquiring knowledge. They note that in her introduction to the 1990 Special Issue of *Cognitive Science*, Rochel Gelman asked, ''How is it that our young attend to inputs that will support the development of concepts they share with their elders?'' Gelman's analysis suggested that the problem cannot be solved without some form of innate knowledge (e.g., ''skeletal principles'') that guides learning in particular domains. Smith et al. (2010) give a different answer to this question. They suggest that the so-called knowledge domains are marked by multiple cue-outcome correlations that in turn correlate with context cues (e.g., the context of word learning may differ from the context of spatial orientation). In the course of learning, children learn to allocate attention to bundles of predictive cues in a given context (this process is called attentional learning). The outcome of this process has the appearance of domain specificity—children learn to differentially allocate attention to different cues in different contexts. In short, Smith et al. (2010) present an account of how domain-general processes (e.g., attentional learning) may give rise to behaviors that have the appearance of domain-specific knowledge.

One set of competencies appearing as a ''knowledge domain,'' or even as a dedicated module, is spatial cognition. Young children as well as a variety of nonhuman species have been found to exhibit sensitivity to spatial information, thus prompting some researchers to propose the existence of a dedicated and encapsulated geometric module. Twyman and Newcombe (2010) consider reasons to doubt the existence of this geometric module and offer a different account of the development of spatial abilities. This account is based on the idea of adaptive cue combination originally proposed by Newcombe and Huttenlocher (2006). According to the proposal, although some biological predispositions for processing of spatial information may exist, fully fledged representation and processing of spatial information emerges through interactions with and feedback from the environment. As a result, multiple sources of spatial and nonspatial information are integrated into a nonmodular and unified representation. Information that is high in salience, reliability, familiarity, and certainty, and low in variability, is given priority over other sources of information. In contrast to modularity proposals, according to the adaptive combination view, experience affects which cues are used in the combination and, as a consequence, the resulting representation. In particular, cues that lead to adaptive behaviors are more likely to be used again in the future, whereas cues that lead to maladaptive behaviors are less likely to be used. This position offers a clear view of how spatial abilities emerge and change in the course of development.

In addition to discussing the papers appearing in this Special Issue of *Cognitive Science*, Goldstone and Landy (2010) offer their view on the problem. They start with a simple observation that the idea of ''skeletal principles'' does not obviate the need for developmental explanations because skeletal structures themselves are subject to growth and development. Goldstone and Landy (2010) exemplify this idea by many systems (e.g., neural networks is but one example) whose internal structure is shaped by the nature of input. They conclude that the field of cognitive development has witnessed a major shift since the 1990 publication of the Special Issue of *Cognitive Science*—the field has moved from delineating specific constraints in domains such as language, motion, quantitative reasoning, social perception, and navigation to explicating mechanisms of how some of these constraints may emerge. Goldstone and Landy (2010) conclude that a new challenge for the study of cognitive development is to understand how general learning processes can give rise to learned domains, dimensions, categories, and contexts.

This Special Issue is a result of multiple efforts by multiple individuals. First, the authors deserve thanks for their willingness to write the papers and subject their work to a standard rigorous peer-review process that required multiple revisions. A set of anonymous reviewers who read the papers (and then revisions) also deserve appreciation. And special thanks go to Art Markman (the Executive Editor of *Cognitive Science*) and Caroline Verdier (the Managing Editor of *Cognitive Science*) who encouraged, supported, and guided the authors through the challenging process of putting together this *Special Issue*.

The collection of papers featured in this Special Issue focuses on the same topic as the 1990 Special Issue. However, the current set of papers offers solutions that are different from those offered in 1990. While in the 1990 issue the main argument was for domain-specific constraints that were considered to be the starting point of development, the current set attempts to understand how constraints emerge in the course of learning and development. Although particular accounts of how this knowledge emerges from domain-general processes may (and most likely will) change over time, the approach itself represents a substantial paradigm shift. Time will tell how successful this approach will be in answering the challenging questions of cognitive development.

## Acknowledgments

Writing of this manuscript was supported by grants from the NSF (REC 0208103), from the Institute of Education Sciences, U.S. Department of Education (R305H050125), and from NIH (R01HD056105) to Vladimir M. Sloutsky.

## References

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131–1157.

Chomsky, N. (1980). *Rules and representations*. Oxford, England: Blackwell.

Gelman, R. (1990). Structural constraints on cognitive development: Introduction to a special issue of Cognitive Science. *Cognitive Science*, *14*, 3–10.

Goldstone, R. L., & Landy, D. (2010). Domain-creating constraints. *Cognitive Science*, *34*(7), 1357–1377.

Johnson, S. (2010). How infants learn about the visual world. *Cognitive Science*, *34*(7), 1158–1184.

Kemp, C., Goodman, N., & Tenenbaum, J. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243.

Newcombe, N. S., & Huttenlocher, J. (2006). Development of spatial cognition. In W. Damon & R. Lerner (Series Eds.) and D. Kuhn & R. Seigler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception and language* (6th ed., pp. 734–776). Hoboken, NJ: John Wiley & Sons.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*(7), 1244–1286.

Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as Process: Contextually Cued Attention and Early Word Learning. *Cognitive Science*, *34*(7), 1287–1314.

Twyman, A. D., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, *34*(7), 1315–1356.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1131–1157

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01049.x

# Language Acquisition Meets Language Evolution

Nick Chater,[a] Morten H. Christiansen[b,c]

[a]*University College London*
[b]*Cornell University*
[c]*Santa Fe Institute*

Received 21 July 2008; received in revised form 26 November 2008; accepted 4 March 2009

**Abstract**

Recent research suggests that language evolution is a process of cultural change, in which linguistic structures are shaped through repeated cycles of learning and use by domain-general mechanisms. This paper draws out the implications of this viewpoint for understanding the problem of language acquisition, which is cast in a new, and much more tractable, form. In essence, the child faces a problem of induction, where the objective is to *coordinate* with others (C-induction), rather than to model the structure of the natural world (N-induction). We argue that, of the two, C-induction is dramatically easier. More broadly, we argue that understanding the acquisition of any cultural form, whether linguistic or otherwise, during development, requires considering the corresponding question of how that cultural form arose through processes of cultural evolution. This perspective helps resolve the ''logical'' problem of language acquisition and has far-reaching implications for evolutionary psychology.

*Keywords:* Biological adaptation; Cognitive development; Cultural evolution; Evolutionary psychology; Induction; Language acquisition; Language evolution; Natural selection; Universal grammar

## 1. Introduction

In typical circumstances, language changes too slowly to have any substantial effect on language acquisition. Vocabulary and minor pronunciation shifts aside, the linguistic environment is typically fairly stable during the period of primary linguistic development. Thus, researchers have treated language as, in essence, fixed, for the purposes of understanding language acquisition. Our argument, instead, attempts to throw light on the problem of language acquisition, by taking an evolutionary perspective, both concerning the biological evolution

Correspondence should be sent to Morten H. Christiansen, Department of Psychology, Cornell University, Ithaca, NY 14853. E-mail: christiansen@cornell.edu

of putative innate domain-specific constraints, and more importantly, the cultural evolution of human linguistic communication. We argue that understanding how language changes over time provides important constraints on theories of language acquisition; and recasts, and substantially simplifies, the problem of induction relevant to language acquisition.

Our evolutionary perspective casts many apparently intractable problems of induction in a new light. When the child aims to learn an aspect of human culture (rather than an aspect of the natural world), the learning problem is dramatically simplified—because culture (including language) is the product of past learning from previous generations. Thus, in learning about the cultural world, we are learning to ''follow in each other's footsteps''—so that our wild guesses are likely to be right—because the right guess is the most popular guess by previous generations of learners. Hence, considerations from language *evolution* dramatically shift our understanding of the problem of language *acquisition*; and we suggest that an evolutionary perspective may also require rethinking theories of the acquisition of other aspects of culture. In particular, in the context of learning about culture, rather than constraints from the natural world, we suggest that a conventional nativist picture, stressing domain-specific, innately specified modules, cannot be sustained.

The structure of the paper is as follows. In the next section, *Language as shaped by the brain,* we describe the logical problem of language evolution that confronts traditional nativist approaches, which propose that the brain has been adapted to language. Instead, we argue that language evolution is better understood in terms of cultural evolution, in which language has been adapted to the brain. This perspective results in a radically different way of looking at induction in the context of cultural evolution. In *C-induction and N-induction*, we outline the fundamental difference between inductive problems in which we must learn to coordinate with one another (C-induction), and those in which we learn aspects of the noncultural, natural world (N-induction). Crucially, language acquisition is, on this account, a paradigm example of C-induction. *Implications for learning and adaptation* shows: (a) that C-learning is dramatically easier than N-induction; and (b) that while innate domain-specific modules may have arisen through biological adaption to deal with problems of N-induction, this is much less likely for C-induction. Thus, while Darwinian selection may have led to dedicated cognitive mechanisms for vision or motor control, it is highly implausible that narrowly domain-specific mechanisms could have evolved for language, music, mathematics, or morality. The next section, *The emergence of binding constraints*, provides a brief illustration of our arguments, using a key case study from language acquisition. Finally, in *Discussion and implications*, we draw parallels with related work in other aspects of human development and consider the implications of our arguments for evolutionary psychology.

## 2. Language as shaped by the brain

Before most children can count to 10 or stand on one leg with their eyes closed for more than 10 s, they are already quite competent users of their native language. It seems that whatever inductive guesses children make about how language works, they tend to get it

right—even when presented with noisy and incomplete input. It is therefore widely assumed that there must be a tight fit between the mechanisms that children employ when acquiring language and the way in which language is structured and used. One way of explaining this close relationship is to posit the existence of domain-specific brain mechanisms dedicated to language acquisition—a Universal Grammar (UG)—through which the linguistic input is funneled (e.g., Chomsky, 1965, 1980). Current conceptions of UG vary considerably in terms of what is genetically specified, ranging from a set of universal linguistic principles with associated parameters in Principles and Parameter Theory (e.g., Crain, Goro, & Thornton, 2006; Crain & Pietroski, 2006), to a language-specific ''toolkit'' that includes structural principles relating to phrase structure (X-bar theory), agreement, and case-marking in Simpler Syntax (Culicover & Jackendoff, 2005; Pinker & Jackendoff, 2009), to the intricate recursive machinery that implements Merge within the Minimalist Program (e.g., Boeckx, 2006; Chomsky, 1995). However, despite the important theoretical differences between current approaches to UG, they all share the central assumption that the core components of UG, whatever their form, are fundamentally arbitrary, from the standpoint of building a system for communication. Thus, the abstract properties of UG do not relate to communicative or pragmatic considerations, nor from limitations on the mechanisms involved in using or acquiring language, or any other functional sources. Indeed, it has been argued that many aspects of UG may even hinder communication (e.g., Chomsky, 2005; Lightfoot, 2000), further highlighting the nonfunctional nature of UG.

The UG framework has been challenged with regard to its ability to account for language acquisition (e.g., Bates & MacWhinney, 1987; Pullum & Scholz, 2002; Tomasello, 2003), the neural basis of language (e.g., Elman et al., 1996; Müller, 2009), and purely linguistic phenomena (e.g., Croft, 2001; Goldberg, 2006; O'Grady, 2005). Whatever the merits are of these challenges (c.f., e.g., Crain & Pietroski, 2001; Laurence & Margolis, 2001; Wexler, 2004; Yang, 2002), our focus here is on what may be an even more fundamental predicament for UG theories: *the logical problem of language evolution* (Botha, 1999; Christiansen & Chater, 2008; Roberts, Onnis, & Chater, 2005; Zuidema, 2003). We argue that there is no credible account of how a richly structured, domain-specific, innate UG could have evolved. Instead, we propose that the direction of causation needs to be reversed: the fit between the neural mechanisms supporting language and the structure of language itself is better explained in terms of how language has adapted to the human brain, rather than vice versa. This solution to the logical problem of language evolution, however, requires abandoning the notion of a domain-specific UG.

### *2.1. The logical problem of language evolution*

As for any other biological structure, an evolutionary story for a putative UG can take one of two routes. One route is to assume that brain mechanisms specific to language acquisition have evolved over long periods of natural selection by analogy with the intricate adaptations for vision (e.g., Pinker & Bloom, 1990). The other rejects the idea that UG has arisen through adaptation and proposes that UG has emerged by nonadaptationist means (e.g., Bickerton, 1995; Gould, 1993; Jenkins, 2000; Lightfoot, 2000).

The nonadaptationist account can rapidly be put aside as an explanation for a domain-specific, richly structured UG. The nonadaptationist account boils down to the idea that some process of *chance variation* leads to the creation of UG. Yet the probability of randomly building a fully functioning, and completely novel, biological system by chance is infinitesimally small (Christiansen & Chater, 2008). To be sure, so-called evo-devo research in biology has shown how a single mutation can lead, via a cascade of genetic ramifications, to dramatic phylogenetic consequences (e.g., additional pairs of legs instead of antennae; Carroll, 2001). But such mechanisms cannot explain how a new, intricate, and functional system can arise de novo.[1]

What of the adaptationist account? UG is intended to characterize a set of universal grammatical principles that hold across all languages; it is a central assumption that these principles are arbitrary. This implies that many combinations of arbitrary principles will be equally adaptive—as long as speakers adopt the *same* arbitrary principles. Pinker and Bloom (1990) draw an analogy between UG and protocols for communication between computers: It does not matter what specific settings are adopted, as long as every agent adopts the same settings. Yet the claim that a particular linguistic ''protocol'' can become genetically embedded through adaptation faces three fundamental difficulties (Christiansen & Chater, 2008).

The first problem stems from the dispersion of human populations. Each subpopulation would be expected to create highly divergent linguistic systems. But, if so, each population will develop a UG as an adaptation to a *different* linguistic environment—and hence, UGs should, like other adaptations, diverge to fit their local environment. Yet modern human populations do not seem to be selectively adapted to learn languages from their own language groups. Instead, every human appears, to a first approximation, equally ready to learn any of the world's languages.[2] The second problem is that natural selection produces adaptations designed to fit the *specific* environment in which selection occurs, that is, a language with a specific syntax and phonology. It is thus puzzling that an adaptation for UG would have resulted in the genetic encoding of highly abstract grammatical properties, rather than fixing the superficial properties of one specific language. The third, and perhaps most fundamental, problem is that linguistic conventions change much more rapidly than genes do, thus creating a ''moving target'' for natural selection. Computational simulations have shown that even under conditions of relatively slow linguistic change, arbitrary principles do not become genetically fixed—this also applies when the genetic make-up of the learners is affecting the direction of linguistic change (Chater, Reali, & Christiansen, 2009; Christiansen, Chater, & Reali, in press).

Together, these arguments against adaptationist and nonadaptationist explanations of UG combine to suggest that there is no viable account of how such an innate domain-specific system for language could have evolved (for details, see Christiansen & Chater, 2008). It remains possible, though, that the origin of language did have a substantial impact on human genetic evolution. The above arguments only preclude biological adaptations for *arbitrary* features of language. There might have been features that are universally stable across linguistic environments that led to biological adaptations, such as the means of producing speech (e.g., Lieberman, 1984; but see also Hauser & Fitch, 2003), the need for enhanced memory capacity (Wynne & Coolidge, 2008), or complex pragmatic inferences (de Ruiter

& Levinson, 2008). However, these language features are likely to be functional—to facilitate language *use*—and thus would typically not be considered part of UG.

### 2.2. Language as shaped by multiple constraints

To escape the logical problem of language evolution, we need to invert the pattern of explanation underpinning the postulation of UG. Instead of viewing the brain as having a genetically specified, domain-specific system for language, which must somehow have arisen over the course of biological evolution, we see the key to language evolution to be evolutionary processes over language itself. Specifically, we view language as an evolving system, and the features of languages as having been shaped by repeated processes of acquisition and transmission across successive generations of language users (e.g., Christiansen, 1994; Culicover & Nowak, 2003; Deacon, 1997; Kirby & Hurford, 2002; Tomasello, 2003; for reviews, see Brighton, Smith, & Kirby, 2005; Christiansen & Chater, 2008). Aspects of language that are easy to learn and process, or are communicatively effective, tend to be retained and amplified; aspects of language which are difficult to learn or process, or which hinder communication, will, if they arise at all, rapidly be stamped out. Thus, the fit between the structure of language and the brains of language users comes about not because the brain has somehow evolved a genetically specified UG capturing the universal properties of language, but instead because language itself is shaped by the brain.

A key assumption of this evolutionary perspective is that language has been shaped by constraints from neural mechanisms that are not dedicated to language. But to what extent can such nonlinguistic constraints be identified and employed to explain linguistic structure previously ascribed to an innate UG? Christiansen and Chater (2008) identify four classes of constraint which simultaneously act to shape language.

#### 2.2.1. Perceptuo-motor factors

The motor and perceptual machinery underpinning language seems inevitably to influence language structure. The seriality of vocal output, most obviously, forces a sequential construction of messages. A perceptual system with a limited capacity for storing sensory input forces a code that can be interpreted incrementally (rather than the many practical codes in communication engineering, in which information is stored in large blocks). The noisiness and variability (across contexts and speakers) of vocal or signed signals may, moreover, provide a pressure toward dividing the phonological space across dimensions related to the vocal apparatus and to ''natural'' perceptual boundaries (e.g., de Boer, 2000; Oller, 2000; Oudeyer, 2005)—though such subdivisions may differ considerably from language to language and thus do not form a finite universal phonological inventory (Evans & Levinson, 2008).

#### 2.2.2. Cognitive limitations on learning and processing

Another source of constraints derives from the nature of cognitive architecture, including learning, processing, and memory. In particular, language processing involves extracting regularities from highly complex sequential input, pointing to a connection between

sequential learning and language: Both involve the extraction and further processing of discrete elements occurring in complex temporal sequences. It is therefore not surprising that sequential learning tasks have become an important experimental paradigm for studying language acquisition and processing (sometimes under the guise of ''artificial grammar/language learning'' or ''statistical learning''; for reviews, see Gómez & Gerken, 2000; Saffran, 2003); and, indeed, some linguists have argued that some important cross-linguistic regularities arise from sequential processing constraints (e.g., Hawkins, 1994, 2004; Kirby, 1999).

*2.2.3. Constraints from thought*

The structure of mental representation and reasoning must, we suggest, have a fundamental impact on the nature of language. The structure of human concepts and categorization must strongly influence lexical semantics; the infinite range of possible thoughts presumably is likely to promote tendencies toward compositionality in natural language (Kirby, 2007); the mental representation of time is likely to have influenced linguistic systems of tense and aspect (Suddendorf & Corballis, 2007); and, more broadly, the properties of conceptual structure may profoundly and richly influence linguistic structure (Jackendoff, 2000). While the Whorfian hypothesis that language influences thought remains controversial, there can be little doubt that thought profoundly influences language.

*2.2.4. Pragmatic constraints*

Similarly, language is likely to be substantially shaped by the pragmatic constraints involved in linguistic communication. Pragmatic processes may, indeed, be crucial in understanding many aspects of linguistic structure, as well as the processes of language change. Levinson (2000) notes that ''discourse'' and syntactic anaphora have interesting parallels, which provide the starting point for a detailed theory of anaphora and binding. As we discuss further below, Levinson argues that initially pragmatic constraints may, over time, become ''fossilized'' in syntax, leading to some of the complex syntactic patterns described by binding theory. Thus, one of the paradigm cases for arbitrary UG constraints may derive, at least in part, from pragmatics.

Christiansen and Chater (2008) note that the four types of constraints interact with one another, such that specific linguistic patterns may arise from a combination of several different types of constraints. For example, the patterns of binding phenomena discussed below are likely to require explanations that cut across the four types of constraints, including constraints on cognitive processing (O'Grady, 2005) and pragmatics (Levinson, 1987; Reinhart, 1983). That is, the explanation of any given aspect of language is likely to require the inclusion of multiple overlapping constraints deriving from perceptuo-motor factors, from cognitive limitations on learning and processing, from the way our thought processes work, and from pragmatic sources.

The idea of explaining language structure and use through the interaction of multiple constraints has a long pedigree within functionalist approaches to the psychology of language (e.g., Bates & MacWhinney, 1979; Bever, 1970; Slobin, 1973). The integration of multiple

constraints—or ''cues''—has risen to prominence in contemporary theories of language acquisition (see e.g., contributions in Golinkoff et al., 2000; Morgan & Demuth, 1996; Weissenborn & Höhle, 2001; for a review, see Monaghan & Christiansen, 2008). For example, 2nd-graders' initial guesses about whether a novel word refers to an object or an action is affected by the sound properties of that word (Fitneva, Christiansen, & Monaghan, in press), 3-4-year-olds' comprehension of relative clause constructions are affected by prior experience (Roth, 1984), 7-year-olds use visual context to constrain on-line sentence interpretation (Trueswell, Sekerina, Hill, & Logrip, 1999), and preschoolers' language production and comprehension is constrained by perspective taking (Nadig & Sedivy, 2002). Similarly, many current theories of adult language processing also involve the satisfaction of multiple constraints (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995), perhaps as a product of processes of language development driven by the integration of multiple cues to linguistic structure (e.g., Farmer, Christiansen, & Monaghan, 2006; Seidenberg & MacDonald, 2001; Snedeker & Trueswell, 2004).

We have considered some of the ways in which language is shaped by the brain. We now turn to the implications of this perspective on the induction problem that the child must solve in language acquisition.

## 3. C-induction and N-induction

Human development involves solving with two, inter-related, challenges: acquiring the ability to understand and manipulate the natural world (N-induction); and acquiring the ability to coordinate with each other (C-induction). Pure cases of these two types of problem are very different. In N-induction, the world imposes an external standard, against which performance is assessed. In C-induction, the standard is not external, but social: The key is that we do the *same* thing, not that we all do an objectively ''right'' thing. In reality, most challenges facing the child involve an intricate mixture of N- and C-induction—and teasing apart the elements of the problem that involve understanding the world, versus coordinating with others, may be very difficult. Nonetheless, we suggest that the distinction is crucially important, both in understanding development in general, and in understanding the acquisition of language, in particular.

To see why the distinction between N- and C-induction is important, consider the difference between learning the physical properties of the everyday world, and learning how to indicate agreement or disagreement using head movements. In order to interact effectively with the everyday world, the child needs to develop an understanding of persistent objects, exhibiting constancies of color and size, which move coherently, which have weight and momentum, and which have specific patterns of causal influences on other objects. The child's perceptuo-motor interactions with the everyday world (e.g., catching a ball; Dienes & McLeod, 1993) depend crucially on such understanding; and do so individualistically—in the sense that success or failure is, to a first approximation, independent of how other children, or adults, understand the everyday world. The child is a lone scientist (Gopnik, Meltzoff, & Kuhl, 1999; Karmiloff-Smith & Inhelder, 1973).

By contrast, in C-learning, the aim is to do as others do. Thus, consider the problem of appropriately deploying a repertoire of head movements to indicate agreement. Whereas there are rich objective constraints, derived from physics, on catching a ball, the problem of communication via head movements is much less constrained—from an abstract point of view, several mappings between overt expressions and underlying mental states may be equivalent. For example, in Northern Europe nodding one's head indicates ''yes,'' but in Greece nodding signals ''no.'' Similarly, in many places across the world, shaking one's head is used for ''no,'' but in Sri Lanka it indicates general agreement (Wang & Li, 2007). What is crucial for the child is that it comes to adopt the *same* pattern of head movement to indicate agreement as those around it. The child is here not a lone scientist, but a musician whose objective is not to attain any absolute pitch, but to be ''in tune'' with the rest of the orchestra.

Before we turn to the question of why C-induction is dramatically easier than N-induction, note that the distinction between N- and C-induction is conceptually distinct from the debate between nativist and empiricist accounts of development (although it has striking implications for these accounts, as we shall see). Table 1 illustrates this point with a range of examples from animal behavior. Thus, in many species, innate constraints appear fundamental to solving N- and C-induction problems. Innate solutions concerning problems of N-induction include basic processes of flying, swimming, and catching prey, as well as highly elaborate and specific behaviors such as nest building. And such innate constraints are equally dominant in determining coordination between animals. Thus, for example, from a functional point of view, patterns of movement might translate into information about food sources in a range of ways; but genetic constraints specify that honey bees employ a *particular* dance (Dyer, 2002). This amounts to solving a problem of C-induction (although solving it over phylogenetic time, via natural selection, rather than solving it over ontogenetic time, via learning), because it is a problem of coordination: The bees must adopt the *same* dance with the same interpretation (and indeed dances do differ slightly between bee species). Courtship, rutting, and play behaviors may often have the same status—the ''rules'' of

Table 1
A tentative classification of a sample of problems of understanding and manipulating the world (N-induction) versus coordinating with others (C-induction) in nonhuman animals

| | Innate Constraints Dominant | Learning Dominant |
|---|---|---|
| N-induction | Locomotion and perceptual-motor control (Alexander, 2003); hunting, foraging, and feeding (Stephens et al., 2007); nest building (Healy et al., 2008) | Learning own environment (Healy & Hurly, 2004), identifying kin (Holmes & Sherman, 1982), learned food preferences and aversion (Garcia et al., 1955) |
| C-induction | Insect social behavior (Wilson, 1971), fixed animal communication systems (Searcy & Nowicki, 2001), including the bee waggle dance (Dyer, 2002), many aspects of play (Bekoff & Byers, 1998), and mate choice (Anderson, 1994) | Social learning (Galef & Laland, 2005), including imitative song-birds (Marler & Slabbekoorn, 2004) |

social interactions are genetically specified; but they are also somewhat arbitrary. The key is that these rules are coordinated across individuals—that a male courtship display is recognizable by relevant females, for example.

Equally, both N- and C-induction can be solved by learning. Animals learn about their immediate environment, where food is located, what is edible, and, in some cases, the identity of conspecifics—this is N-induction, concerning objective aspects of the world. Indeed, some learned behaviors (such as milk-bottle pecking in blue tits or food preparation techniques in chimpanzees or gorillas) may be learned from conspecifics, although whether by processes of emulation, imitation, or simpler mechanisms, is not clear (Hurley & Chater, 2005). To a modest degree, some nonhuman animals also learn to coordinate their behavior. For example, some song birds and whales learn their songs from others. Reproductive success depends on producing a ''good'' song defined in terms of the current dialect (Marler & Slabbekoorn, 2004), rather than achieving any ''objective'' standard of singing.

The distinction between problems of C- and N-induction is, then, conceptually separate from the question of whether an induction problem is solved over phylogenetic time, by natural selection (and specifically, by the adaptation of genetically encoded constraints), or over ontogenetic time, by learning. Nonetheless, the distinction has two striking implications for the theory of development, and, in particular, for language acquisition. First, as we shall argue, C-induction is dramatically easier than N-induction; and many aspects of language acquisition seem paradoxically difficult because a problem of C-induction is mischaracterized as a problem of N-induction. Second, the child's ability to solve C-induction problems, including language acquisition, must primarily be based on cognitive and neural mechanisms *that predate the emergence of the cultural form to be learned*. That is, natural selection cannot lead to the creation of dedicated, domain-specific learning mechanisms for solving C-induction problems (e.g., innate modules for language acquisition). By contrast, such mechanisms may be extremely important for solving N-induction problems. Table 2,

Table 2
A tentative classification of sample problems of understanding and manipulating the world (N-induction) versus coordinating with others (C-induction) in human development

| | Innate Constraints Dominant | Learning Dominant |
|---|---|---|
| N-induction | Low-level perception, motor control (Crowley & Katz, 1999), perhaps core naïve physics (Carey & Spelke, 1996) | Perceptual, motor, and spatial learning (Johnson, this issue, Newcombe, this issue; Shadmehr & Wise, 2005); science and technology (Cartwright, 1999) |
| C-induction | Understanding other minds (Tomasello et al., 2005), pragmatic interpretation (de Ruiter & Levinson, 2008), social aspects of the emotions (Frank, 1988), basic principles of cooperation, reciprocation, and punishment (Fehr & Gächter, 2002; Olson & Spelke, 2008) | Language, including syntax, phonology, word learning, and semantics (Smith, this issue), linguistic categorization (Sloutsky, this issue; Tenenbaum, this issue). Other aspects of culture (Geertz, 1973), including music, art, social conventions, ritual, religion, and moral codes |

somewhat speculatively, considers examples from human cognition, including some of the topics considered in this special issue. Rather than focusing in detail on each of these cases, we focus here on the general distinction between N-induction and C-induction, before turning to our brief illustrative example, binding constraints.

## 4. Implications for learning and adaptation

Suppose that some natural process yields the sequence 1, 2, 3… How does it continue? Of course, we have far too little data to know. It might oscillate (1, 2, 3, 2, 1, 2, 3, 2, 1…), become ''stuck'' (1, 2, 3, 3, 3, 3…), exhibit a Fibonacci structure (1, 2, 3, 5, 8…), and any of an infinity of more or less plausible alternatives. This indeterminacy makes the problem of N-induction of structure from the natural world difficult, although not necessarily hopelessly so, in the light of recent developments in statistics and machine learning (Chater & Vitányi, 2007; Harman & Kulkarni, 2007; Li & Vitányi, 1997; Tenenbaum, Kemp, & Shafto, 2007).

But consider the parallel problem of C-learning—we need not guess the ''true'' continuation of the sequence. We only have to *coordinate* our predictions with those of other people in the community. This problem is very much easier. From a psychological point of view, the overwhelmingly natural continuation of the sequence is ''…4, 5, 6….'' That is, most people are likely to predict this. Thus, coordination emerges easily and unambiguously on a specific infinite sequence, even given a tiny amount of data.

Rapid convergence of human judgments, from small samples of data, is observed across many areas of cognition. For example, Feldman (1997) and Tenenbaum (1999) show that people converge on the same categories incredibly rapidly, given a very small number of perceptual examples; and rapid convergence from extremely limited data is presupposed in intelligence testing, where the majority of problems are highly indeterminate, but responses nonetheless converge on a single answer (e.g., Barlow, 1983). Moreover, when people are allowed to interact, they rapidly align their choice of lexical items and frames of reference, even when dealing with novel and highly ambiguous perceptual input (e.g., Clark & Wilkes-Gibbs, 1986; Pickering & Garrod, 2004). Finally, consider a striking, and important class of examples from game theory in economics. In a typical coordination game, two players simultaneously choose a response; if it is the same, they both receive a reward; otherwise, they do not. Even when given very large sets of options, people often converge in ''one shot.'' Thus, if asked to select time and meeting place in New York, Schelling (1960) found that people generated several highly frequent responses (so-called focal points) such as ''twelve noon at Grand Central Station,'' so that players might potentially meet successfully, despite choosing from an almost infinite set of options. The corresponding problem of N-induction (i.e., of guessing the time and place of an arbitrarily chosen event in New York) is clearly hopelessly indeterminate; but as a problem of C-induction, where each player aims to coordinate with the other, it is nonetheless readily solved.

C-induction is, then, vastly easier than N-induction—essentially because, in C-induction, human cognitive biases inevitably work in the learner's favor as those biases are shared

with other people, with whom coordination is to be achieved. In N-induction, the aim is to predict Nature—and here, our cognitive biases will often be an unreliable guide.

Language acquisition is a paradigm example of C-induction. There is no human-independent ''true'' language, to which learners aspire. Rather, today's language is the product of yesterday's learners; and hence language acquisition requires *coordinating* with those learners. What is crucial is not *which* phonological, syntactic, or semantic regularities children prefer, when confronted with linguistic data; it is that they prefer the *same* linguistic regularities—each generation of learners needs only to follow in the footsteps of the last.

Note that the existence of very strong cognitive biases is evident across a wide range of learning problems—from categorization, to series completion, to coordinating a meeting. Thus, the mere existence of strong biases in no way provides evidence for a dedicated innate ''module'' embodying such biases. From this point of view, a key research question concerns the nature of the biases that influence language acquisition—these biases will help explain the structures that are, or are not, observed in the world's languages. Moreover, the *stronger* the biases (e.g., flowing from the interaction of perceptuo-motor factors, cognitive limitations on learning and processing, and constraints from thought and pragmatics, as described above), the *greater* the constraints on the space of possible languages, and hence the *easier* the problem of language acquisition.

Language, and other cultural phenomena, can therefore be viewed as evolving systems, and one of the most powerful determinants of which linguistic or cultural patterns are invented, propagated, or stamped out, is how readily those patterns are learned and processed. Hence, the learnability of language, or other cultural structures, is not a puzzle demanding the presence of innate information, but rather an inevitable consequence of the process of the incremental construction of language, and culture more generally, by successive generations (Deacon, 1997; Kirby & Hurford, 2002; Zuidema, 2003).

The first implication we have drawn from the distinction between C-induction and N-induction is that C-induction is dramatically easier than N-induction. But there is a second important implication, concerning the feasibility of the biological adaptation of specific inductive biases—that is, whether genetically encoded domain-specific modules could have arisen through Darwinian selection. This possibility looks much more plausible for problems of N-induction than for C-induction.

Many aspects of the natural world are fairly stable. Thus, across long periods of evolutionary time, there is little change in the low-level statistical regularities in visual images (Field, 1987), in the geometric properties of optic flow, stereo, or structure-from-motion (Ullman, 1979), or in the coherence of external visual and auditory ''objects'' (e.g., Bregman, 1990). These aspects of the environment therefore provide a stable selectional pressure over which natural selection can operate—often over times scales of tens or hundreds of millions of years. Just as the sensory and motor apparatus are exquisitely adapted to deal with the challenges of the natural environment, so it is entirely plausible that the neural and cognitive machinery required to operate this apparatus is equally under genetic control, at least to some substantial degree (e.g., Crowley & Katz, 1999). Indeed, in many organisms, including many mammals, much complex perceptual-motor behavior is functioning within hours of birth. Perceptuo-motor function appears to be considerably

delayed in human infancy, but it is nonetheless entirely plausible that some innate neural structures are conserved, or perhaps even elaborated, in humans. More broadly, it is at least prima facie plausible that biases regarding many problems of N-induction might be established by natural selection.

Consider, by contrast, the case of C-induction. While the natural world is stable, the behaviors on which people coordinate are typically highly *un*stable. Thus, the choice of meeting place in New York will, clearly, depend on contingent historical and cultural factors; but more importantly, cultural and linguistic conventions are in general highly labile—for example, the entire Indo-European language group, including Armenian, Finnish, Hindi, Ukrainian, and Welsh, which exhibit huge variations in case systems, word order, and phonology, have diverged in just 10,000 years (Gray & Atkinson, 2003). Moreover, ''focal points'' on which people can converge may emerge very rapidly during an experiment; for example, different pairs of participants rapidly develop one of a wide range of classifications in a task involving novel tangrams (Clark & Wilkes-Gibbs, 1986), and complex patterns of conventions can arise very rapidly in the emergence of languages. For example, Nicaraguan sign language has emerged within three decades, created by deaf children with little exposure to established languages (Senghas, Kita, & Özyürek, 2004). Thus, from this perspective, Pinker and Bloom's (1990) analogy between the evolution of vision and language breaks down because the former is primarily a problem of N-induction and the latter a problem of C-induction.

To summarize, C-induction involves learning what others will do; but what others will do is highly variable—and, crucially, changes far more rapidly than genetic change. Suppose that a particular set of cultural conventions is in play (a specific language, or religious or moral code). Learners with an inductive bias which, by chance, makes these conventions particularly easy to acquire will be favored. But there is no opportunity for those innate biases to spread through the population, because long before substantial natural selection can occur, those conventions will no longer apply, and a bias to adopt them will, if anything, be likely to be a disadvantage (Chater et al., 2009; Christiansen et al., in press). Hence, Darwinian selection will favor agents that are generalists—that is, can adapt to the changing cultural environment. It will, in particular, not involve the *coevolution* of genes and specific, though initially arbitrary, cultural conventions. Rapid cultural evolution (e.g., fast-changing linguistic, moral, or social systems) will automatically lead to a fit between culture and learners—because cultural patterns can only be created and propagated if they are easy to learn and use. But cultural evolution will work *against* biological (co)evolution in the case of malleable aspects of culture—rapid cultural change leads to a fast-changing cultural environment, which serves as a ''moving target'' to which biological adaptation cannot occur (c.f., Ancel, 1999).

There has, indeed, been extensive computational and mathematical analysis of the process of cultural evolution, including some models of language change (e.g., Batali, 1998; Hare & Elman, 1995; Kirby, Dowman, & Griffiths, 2007; Nettle, 1999; Niyogi, 2006; Nowak, Komarova, & Niyogi, 2001; Richerson & Boyd, 2005). Learning or processing constraints on learners provide one source of constraint on how such cultural evolution proceeds. Under some restricted conditions, learning biases specify a ''fixed'' probability

distribution of linguistic/cultural forms, which from cultural evolution can be viewed as sampling (Griffiths & Kalish, 2005). In the general case, though, historical factors can also be crucially important—once a culture/language has evolved in a particular direction, there may be no way to reverse the process. This observation seems reasonable in the light of numerous one-directional ''clines'' observed in empirical studies of language change (Comrie, 1989).

While arbitrary conventions, in language or other aspects of culture, typically change rapidly, and hence do not provide a stable target upon which biological evolution can operate, there may be important aspects of language and culture that are *not* arbitrary—that is, for which certain properties have functional advantages. For example, the functional pressure for communicative efficiency might explain why frequent words tend to be short (Zipf, 1949), and the functional pressure to successfully engage in repeated social interactions may explain the tendency to show reciprocal altruism (Trivers, 1971). Such aspects of culture could potentially provide a stable environment against which biological selection might take place. Moreover, ''generalist'' genes for dealing with a fast-changing cultural environment may also be selected for. Thus, it is in principle possible that the human vocal apparatus, memory capacity, and perhaps the human auditory system, might have developed specific adaptations in response to the challenges of producing and understanding speech, although the evidence that this actually occurred is controversial (e.g., Lieberman, 1984; but see also Hauser & Fitch, 2003). But genes encoding aspects of culture that were initially freely varying, and not held constant by functional pressure, could not have arisen through biological evolution (Chater et al., 2009).

While the issues discussed above apply across cognitive domains, we illustrate the payoff of this standpoint by considering a particularly central aspect of language—binding constraints—which has been viewed as especially problematic for nonnativist approaches to language acquisition, and to provide strong grounds for the postulation of innate language-specific knowledge.

## 5. The emergence of binding constraints

The problem of binding, especially between reflexive and nonreflexive pronouns and noun phrases, has for a long time been a theoretically central topic in generative linguistics (Chomsky, 1981); and the principles of binding appear both complex and arbitrary. Binding theory is thus a paradigm case of the type of information that has been proposed to be part of an innate UG (e.g., Crain & Lillo-Martin, 1999; Reuland, 2008), and it provides a challenge for theorists who do not assume UG. As we illustrate, however, there is a range of alternative approaches that provide a promising starting point for understanding binding as arising from domain-general factors. If such approaches can make substantial in-roads into the explanation of key binding principles, then the assumption that binding constraints are arbitrary language universals and must arise from an innate UG is undermined. Indeed, according to the latter explanation, apparent links between syntactic binding principles and pragmatic factors must presumably be viewed as mere coincidences—rather than as

originating from the ''fossilization'' of pragmatic principles into syntactic patterns by processes such as grammaticalization (Hopper & Traugott, 1993).

The principles of binding capture patterns of use of, among other things, reflexive pronouns (e.g., *himself*, *themselves*) and accusative pronouns (e.g., *him*, *them*). Consider the following examples, where subscripts indicate co-reference and asterisks indicate ungrammaticality:

(1) That $John_i$ enjoyed $himself_i$/*$him_i$ amazed $him_i$/*$himself_i$.
(2) $John_i$ saw $himself_i$/*$him_i$/*$John_i$.
(3) *$He_i$/$he_j$ said $John_i$ won.

Why is it possible for the first, but not the second, pronoun to be reflexive, in (1)? According to generative grammar, the key concept here is *binding*. Roughly, a noun phrase *binds* a pronoun if it c-commands that pronoun, and they are co-referring. In an analogy between linguistic and family trees, an element c-commands its siblings and all their descendents. A noun phrase, NP, *A-binds* a pronoun if it binds it; and, roughly, if the NP is in either subject or object position. Now we can state simplified versions of Chomsky's (1981) three binding principles:

*Principle A*. Reflexives must be A-bound by an NP.
*Principle B*. Pronouns must not be A-bound by an NP.
*Principle C*. Full NPs must not be A-bound.

Informally, Principle A says that a reflexive pronoun (e.g., *herself*) must be used, if co-referring to a ''structurally nearby'' item (defined by c-command), in subject or object position. Principle B says that a nonreflexive pronoun (e.g., *her*) must be used otherwise. These principles explain the pattern in (1) and (2). Principle C rules out co-reference such as (3). *John* cannot be bound to *he*. For the same reason, *John likes John*, or *the man likes John* do not allow co-reference between subject and object.

Need the apparently complex and arbitrary principles of binding theory be part of the child's innate UG? Or can these constraints be explained as a product of more basic perceptual, cognitive, or communicative constraints? One suggestion, due to O'Grady (2005), considers the possibility that binding constraints may in part emerge from processing constraints (see Section 2.2.2). Specifically, he suggests that the language processing system seeks to resolve linguistic dependencies (e.g., between verbs and their arguments) at the first opportunity—a tendency that might not be specific to syntax, but which might be an instance of a general cognitive tendency to resolve ambiguities rapidly in linguistic (Clark, 1975) and perceptual input (Pomerantz & Kubovy, 1986). The use of a reflexive is assumed to signal that the pronoun co-refers with an available NP, given a local dependency structure.

Thus, in parsing (1), the processor reaches *That John enjoyed himself…* and makes the first available dependency relationship between *enjoyed*, *John*, and *himself*. The use of the reflexive, *himself*, signals that co-reference with the available NP, *John*, is intended (c.f., Principle A). With the dependencies now resolved, the internal structure of the resulting clause is ''closed off'' and the parser moves on: *[That [John enjoyed himself]] surprised him/*himself.* The latter *himself* is not possible because there is no appropriate NP available

to connect with (the only NP is [*that John enjoyed himself*]) which is used as an argument of *surprised*, but which clearly cannot co-refer with the *himself*. But in *John enjoyed himself*, *John* is available as an NP when *himself* is encountered.

By contrast, plain pronouns, such as *him*, are used in roughly complementary distribution to reflexive pronouns (c.f., Principle B). It has been argued that this complementarity arises pragmatically (Levinson, 1987; Reinhart, 1983); that is, given that the use of reflexives is highly restrictive, they are, where appropriate, more informative. Hence, by not using them, the speaker signals that the co-reference is not appropriate.[3] Thus, we can draw on the additional influence of *pragmatic* constraints (Section 2.2.4).

Finally, simple cases of Principle C can be explained by similar pragmatic arguments. Using *John sees John* (see [2] above), where the object can, in principle, refer to any individual named John, would be pragmatically infelicitous if co-reference were intended—because the speaker should instead have chosen the more informative *himself* in object position. O'Grady (2005) and Reinhart (1983) consider more complex cases related to Principle C, in terms of a processing bias toward so-called upward feature-passing, though we do not consider this here.

The linguistic phenomena involved in binding are extremely complex and not fully captured by *any* theoretical account (indeed, the minimalist program [Chomsky, 1995]; has no direct account of binding but relies on the hope that the principles and parameters framework, in which binding phenomena have been described, can eventually be reconstructed from a minimalist point of view). We do not aim here to argue for any specific account of binding phenomena; but rather to indicate that many aspects of binding may arise from general processing or pragmatic constraints—such apparent relations to processing and pragmatics are, presumably, viewed as entirely coincidence according to a classical account in which binding constraints are communicatively arbitrary and expressions of an innate UG. Note, in particular, that it is quite possible that the complexity of the binding constraints arises from the interaction of *multiple* constraints. For example, Culicover and Jackendoff (2005) have recently argued that many aspects of binding may be semantic in origin. Thus, *John painted a portrait of himself* is presumed to be justified due to semantic principles concerning representation (the portrait is a representation of John), rather than any syntactic factors. Indeed, note too that, we can say: *Looking up, Tiger was delighted to see himself at the top of the leaderboard* where the reflexive refers to the name ''Tiger,'' not Tiger himself. And violations appear to go beyond mere representation—for example, *After a wild tee-shot, Ernie found himself in a deep bunker*, where the reflexive here refers to *his golfball.* More complex cases, involving pronouns and reflexives are also natural in this type of context, for example, *Despite* $Tiger_i$*'s mis-cued drive,* $Angel_j$ *still found* $himself_{(j's\ golfball)}$ *10 yards behind* $him_{(i's\ golfball)}$. There can, of course, be no purely syntactic rules connecting golfers and their golfballs; and presumably no general semantic rules either, unless such rules are presumed to be sensitive to the rules of golf (among other things, that each player has exactly one ball). Rather, the reference of reflexives appears to be determined by pragmatics and general knowledge—for example, we know from context that a golfball is being referred to; that golfballs and players stand in one-to-one correspondence; and hence that picking out an individual could be used to signal the corresponding golfball.

The very multiplicity of constraints involved in the shaping of language structure, which arises naturally from the present account, may be one reason why binding is so difficult to characterize in traditional linguistic theory. But these constraints do not pose any challenges for the child—because these constraints are the very constraints with which the child is equipped. If learning the binding constraints were a problem of N-induction (e.g., if the linguistic patterns were drawn from the language of intelligent aliens; or deliberately created as a challenging abstract puzzle), then learning would be extraordinarily hard. But it is not: it is a problem of C-induction. To the extent that binding can be understood as emerging from a complex of processing, pragmatic, or other constraints operating on past generations of learners, then binding will be readily learned by the new generations of learners, who will necessarily embody those very constraints.

It might be argued that if binding constraints arise from the interaction of a multiplicity of constraints, one might expect that binding principles across historically unrelated languages would show strong family resemblances (as they would, in essence, be products of cultural *co-evolution*), rather than being strictly identical, as is implicit in the claim that binding principles are universal across human languages. Yet it turns out that the binding constraints, like other putatively ''strict'' language universals, may not be universal at all, when a suitably broad range of languages is considered (e.g., Evans & Levinson, 2008). Thus, Levinson (2000) notes that, even in Old English, the equivalent of *He saw him* can optionally allow coreference (apparently violating Principle A). Putative counterexamples to binding constraints, including the semantic/pragmatic cases outlined above, can potentially be fended off, by introducing further theoretical distinctions—but such moves run the real risk of stripping the claim of universality of real empirical bite (Evans & Levinson, 2008). If we take cross-linguistic data at face value, the pattern of data seems, if anything, more compatible with the present account, according to which binding phenomena results from the operation of multiple constraints during the cultural evolution of language, than the classical assumption that binding constraints are a rigid part of a fixed UG, ultimately rooted in biology.

To sum up: Binding has been seen as paradigmatically arbitrary and specific to language; and the learnability of binding constraints has been viewed as requiring a language-specific UG. If the problem of language learning were a matter of N-induction—that is, if the binding constraints were merely a human-independent aspect of the natural world—then this viewpoint would potentially be persuasive. But language learning is a problem of C-induction—people have to learn the *same* linguistic system as each other. Hence, the patterns of linguistic structure will themselves have adapted, through processes of cultural evolution, to be easy to learn and process—or more broadly, to fit with the multiple perceptual, cognitive, and communicative constraints governing the adaptation of language. From this perspective, binding is, in part, determined by innate constraints—but those constraints predate the emergence of language (de Ruiter & Levinson, 2008).

In the domain of binding, as elsewhere in linguistics, this type of cultural evolutionary story is, of course, incomplete—though to no greater degree, arguably, than is typical in genetic evolutionary explanations in the biological sciences. We suggest that viewing language as a cultural adaptation provides, though, a powerful and fruitful framework within

which to explore the evolution of linguistic structure and its consequences for language acquisition.

## 6. Discussion and implications

The theme of this special issue concerns one of the fundamental questions in cognitive development: the degree to which development is driven by domain-general learning mechanisms or by innate domain-specific constraints. The papers herein illustrate a variety of key developments in approaches that stress the importance of domain-general mechanisms, in areas ranging from conceptual development, to spatial cognition, to language acquisition. Here, our narrow focus has been on language. But our argument involved stepping back from questions concerning the acquisition of language, to take an evolutionary perspective, both concerning the biological evolution of putative innate constraints and the cultural evolution of human linguistic communication. Based on an evolutionary analysis, we proposed reconsidering development in terms of two types of inductive problems: N-induction, where the problem involves learning some aspect of the natural world, and C-induction, where the key to solving the learning problem is to coordinate with others. In this light, we then briefly reevaluated a key puzzle for language acquisition—the emergence of binding constraints—which has traditionally been interpreted as providing strong support for the existence of an innate UG. In this final discussion, we point to some of the broader implications of our approach for language acquisition and human development.

### *6.1. The logical problem of language acquisition reconsidered*

We have argued that viewing the evolution of language as the outcome of cultural, rather than biological evolution (and hence as a problem of C-induction, rather than N-induction) leads to a dramatically different perspective on language acquisition. The ability to develop complex language from what appears to be such poor input has traditionally led many to speak of the ''logical'' problem of language acquisition (e.g., Baker & McCarthy, 1981; Hornstein & Lightfoot, 1981). One solution to the problem is to assume that learners have some sort of biological headstart in language acquisition—that their learning apparatus is precisely meshed with the structure of natural language. This viewpoint is, of course, consistent with theories according to which there is a genetically specified language module, language organ, or language instinct (e.g., Chomsky, 1986; Crain, 1991; Piattelli-Palmarini, 1989; Pinker, 1994; Pinker & Bloom, 1990). But if we view language acquisition as a problem of C-induction, then the learner's objective is merely to follow prior learners—and hence the patterns in language will inevitably be those that are most readily learnable. It is not that people have evolved to learn language; rather, language has evolved to fit the multiple constraints of human learning and processing abilities.

Whatever learning biases people have, so long as these biases are *shared* across individuals, learning should proceed successfully. Moreover, the viewpoint that children learn language using general-purpose cognitive mechanisms, rather than language-specific mechanisms, has also been advocated on independent grounds (e.g., Bates & MacWhinney, 1979, 1987; Deacon, 1997; Elman et al., 1996; Monaghan & Christiansen, 2008; Seidenberg & MacDonald, 2001; Tomasello, 2000, 2003).

This alternative characterization of language acquisition additionally offers a different perspective on linguistic phenomena that have typically been seen as requiring a UG account for their explanation, such as specific language impairment (SLI) and creolization. These phenomena are beyond the scope of this paper, so we can only sketch how they may be approached. For example, the acquisition problems in SLI may, on our account, be largely due to deficits in underlying sequential learning mechanisms that support language (see Ullman & Pierpont, 2005; for a similar perspective), rather than impaired language-specific modules (e.g., Gopnik & Crago, 1991; Pinker, 1994; Van der Lely & Battell, 2003). Consistent with this perspective, recent studies have shown that children and adults with SLI have impaired sequential learning abilities (e.g., Evans & Saffran, 2005; Hsu, Christiansen, Tomblin, Zhang, & Gómez, 2006; Tomblin, Mainela-Arnold, & Zhang, 2007). Although processes of creolization, in which children acquire consistent linguistic structure from noisy and inconsistent input, have been seen as evidence of UG (e.g., Bickerton, 1984), we suggest that creolization may be better construed as arising from cognitive constraints on learning and processing. The rapid emergence of a consistent subject-object-verb word order in the Al-Sayyid Bedouin Sign Language (Sandler, Meir, Padden, & Aronoff, 2005) is consistent with this suggestion. Additional research is required to flesh out these accounts in detail, but a growing bulk of work indicates that such accounts are indeed viable (e.g., Chater & Vitányi, 2007; Goldberg, 2006; Hudson Kam & Newport, 2005; O'Grady, 2005; Reali & Christiansen, 2005; Tomasello, 2003).

### *6.2. Cultural evolution meets evolutionary psychology*

How far do these arguments generalize from language acquisition to the development of the child's knowledge of culture more broadly? How far might this lead to a new perspective in evolutionary psychology, in which the fit between the brain and cultural forms is not explained in terms of domain-specific modules, but by the shaping of cultural forms to preexisting biological machinery?

Human development involves the transmission of an incredibly elaborate culture from one generation to the next. Children acquire language; lay theories and concepts about the natural, artificial, and psychological worlds; social and moral norms; a panoply of practical lore and skills; and modes of expression, including music, art, and dance. The absorption of this information is all the more remarkable given that so much of it appears to be acquired incidentally, rather than being a topic of direct instruction.

As with language, it is prima facie unclear how this astonishing feat of learning is accomplished. One natural line of explanation is to assume that there is a close fit between the cultural information to be transmitted and the prior assumptions of the child,

whether implicit or explicit. The strongest form of this position is that some, and perhaps the most central, elements of this information are actually innately ''built in'' to each learner—and hence that cultural transmission is built over a skeleton of genetically fixed constraints (e.g., Hauser, 2006). Generalizing from the case of UG, some evolutionary psychologists have likened the mind to a Swiss army knife, consisting of a variety of special-purpose tools (Barkow, Cosmides, & Tooby, 1992). The design of each of these special-purpose tools is presumed to have arisen through biological selection. More broadly, the key suggestion is that there is a close mesh between genes and culture—and that this mesh helps explain how cultural complexity can successfully be transmitted from generation to generation.

The processes by which any putative connection between genes and culture might arise are central to the study of human development; and understanding such processes is part of the wider project of elucidating the relationship between biological and cultural explanation in psychology, anthropology, and throughout the neural and social sciences. But here we wish to take a wider view of these familiar issues, from the point of view of historical *origins*: How did the mesh between genes and culture arise?

The origin of a close mutual relationship between any two systems raises the question: Which came first? A natural line, in considering this type of problem, is to consider the possibility of co-evolution—and hence that the claim that one, or the other, must come first is misleading. As we have argued, in the case of genes and language, the conditions under which such co-evolution can occur are surprisingly limited; but the same issues arise in relation to the putative co-evolution of genes and any cultural form. Let us now broaden the argument and consider the two clear-cut options: that culture comes first, and biological adaptation brings about the fit with cultural structure; or the biological structures come first, and cultural adaptation brings about the fit with these biological structures. As a short hand, let us call these the *biological evolution* and *cultural evolution* perspectives.

How might biological evolution work? If cultural conventions have a particular form, then people within that culture will, we may reasonably assume, have a selective advantage if they are able to acquire those conventions rapidly and easily. So, for example, suppose that human cultures typically (or even always) fit some specific moral, social, or communicative pattern. Hence, children who are able rapidly to learn these constraints will presumably have a selective advantage. Thus, it is possible that, after a sufficiently long period of biological adaptation to an environment containing such constraints, learners who are genetically biased in favor of those constraints might emerge, so that they learn these constraints from very little cultural input; and, at the extreme, learners might be so strongly biased that they require no cultural input at all.[4]

If, though, we assume that genetic (or more generally biological) structure is *developmentally* prior (i.e., that learners acquire their culture via domain-specific genetic constraints, adapted to cultural patterns), then it appears that culture must be *historically* prior. The cultural structure (e.g., the pattern of specific syntactic regularities) provides the key aspect of the environment to which genes have adapted. Thus, if a genetically specified and domain-specific system containing specific cultural knowledge has arisen through Darwinian

processes of selection, then such selection appears to require a preexisting cultural environment, to which biological adaptation occurs. However, this conclusion is in direct contradiction to the key assumption of the biological approach—because it presupposes that the cultural forms do *not* arise from biological constraints, but predate them. If culture could preexist biological constraints, then the reason to postulate such constraints almost entirely evaporates.[5]

But it is clear, in the light of the arguments above, that there is alternative cultural evolution perspective: that *biological* structure is prior, and that it is cultural forms that adapt, through processes of cultural transmission and variation (e.g., Boyd & Richerson, 2005; Mesoudi, Whiten, & Laland, 2006) to fit biological structure as well as possible. Specifically, the culture is viewed as shaped by endless variation and winnowing, in which forms and patterns which are readily learned and processed are adopted and propagated, whereas forms which are difficult to learn or process are eliminated. Not merely language, but culture in general, is shaped by the brain, rather than the reverse.

Cultural forms will, of course, also be shaped by functional considerations: Just as language has been shaped to support flexible and expressive communication, tool use may have been shaped by efficacy in hunting, flaying, and food preparation. But according to this viewpoint, the fit between learners and culture is underpinned by prior biological ''machinery'' *that predates that culture, and hence is not itself shaped to deal with cultural problems*. This biological machinery may very well be the product of Darwinian selection, but in relation to preexisting goals. Thus, for example, the perceptuo-motor and planning systems may be highly adapted for the processing of complex hierarchically structured sequences (e.g., Byrne & Byrne, 1993); and such abilities may then be co-opted as a partial basis for producing and understanding language (Conway & Christiansen, 2001). Similarly, the ability to ''read'' other minds may have developed to deal with elaborate social challenges in societies with relatively little cultural innovation (as in nonhuman primates); but such mind-reading might be an essential underpinning for language and the development of social and moral rules (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

### *6.3. Conclusion*

A key challenge for future research will be to identify the specific biological, cognitive, and social constraints that have shaped the structure of language through cultural transmission; to show how the selectional pressures imposed by these constraints lead to specific patterns in the world's languages; and to demonstrate how these constraints can explain particular patterns of language acquisition and processing. If we generalize our evolutionary approach to other aspects of cultural evolution and human development, then similar challenges will also lie ahead here in identifying specific constraints and explaining how these capture cross-cultural patterns in development. Importantly, this perspective on human evolution and development does not construe the mind as a blank slate; far from it: We need innate constraints to explain the various patterns observed across phylogenetic and ontogenetic time. Instead, we have argued that there are many innate constraints that shape language and other culturally based human skills but that these are unlikely to be

domain specific. Thus, as Liz Bates put it so elegantly (cited in Goldberg, 2008), ''It's not a question of Nature versus Nurture; the question is about the Nature of Nature.''

## Notes

1. It might be objected, in the light of the minimalist program in linguistics, that only a very modest biological adaptation specific to language—recursion—may be required (Hauser, Chomsky, & Fitch, 2002). This response appears to fall on the horns of a dilemma. On the one hand, if UG consists only of the operation of recursion, then traditional generativist arguments concerning the poverty of the stimulus, and the existence of language universals, have been greatly exaggerated—and indeed, an alternative, non-UG-based explanation of the possibility of language acquisition and the existence of putative language universals is required. This position, if adopted, seems to amount to a complete retraction of the traditional generativist position (Pinker & Jackendoff, 2005). On the other hand, if the argument from the poverty of the stimulus is still presumed to hold good, with its implication that highly specific regularities such as the binding constraints must be part of an innate UG, then the probability of such complex, arbitrary systems of constraints arising by chance is vanishingly small. To be sure, the minimalist explanation of many linguistic regularities is based on the recursive operation Merge—but, in reality, explanations of specific linguistic data require drawing on extensive and highly abstract linguistic machinery, which goes far beyond simple recursion (Adger, 2003; Boeckx, 2006).
2. Dediu and Ladd (2007) present statistical analyses of typological and genetic variation across Old World languages, suggesting that there may be differences in genetic biases for learning tonal versus sequential phonology. They argue that these biases are unlikely to be due to biological adaptations for language because the same mutations would have had to arise independently several times; instead, they propose that these genetic biases may have arisen for other reasons independent of language but once in place they would slowly have shaped individual languages over generations toward either incorporating tonal contrasts or not. This suggestion fits closely with our argument below that language has been shaped by the brain.
3. It is also possible, of course, that as with pragmatic patterns in general, this pattern may become increasingly conventionalized through use—a typical pattern in grammaticalization (Hopper & Traugott, 1993).
4. This style of explanation, by which traits that are initially acquired during environmental exposure during development may ultimately become innate—that is, independent of environmental input—is known as the Baldwin effect (Baldwin, 1896; see Weber & Depew, 2003, for discussion).
5. Of course, possible co-evolutionary processes between genes and culture complicates the argument but does not change the conclusion. For a more detailed discussion of these issues, in the context of language, see Christiansen and Chater (2008).

## Acknowledgments

Nick Chater was supported by a Major Research Fellowship from the Leverhulme Trust and by ESRC grant number RES-000-22-2768. Morten H. Christiansen was supported by a Charles A. Ryskamp Fellowship from the American Council of Learned Societies. We are grateful to Kenny Smith and two anonymous reviewers for their feedback on a previous version of this paper.

## References

Adger, D. (2003). *Core syntax*. Oxford, England: Oxford University Press.

Alexander, R. M. (2003). *Principles of animal locomotion*. Princeton, NJ: Princeton University Press.

Ancel, L. (1999). A quantitative model of the Simpson-Baldwin effect. *Journal of Theoretical Biology*, *196*, 197–209.

Anderson, M. B. (1994). *Sexual selection*. Princeton, NJ: Princeton University Press.

Baker, C. L., & McCarthy, J. J. (Eds.) (1981). *The logical problem of language acquisition*. Cambridge, MA: MIT Press.

Baldwin, J. M. (1896). A new factor in evolution. *American Naturalist*, *30*, 441–451.

Barkow, J., Cosmides, L. & Tooby, J. (Eds.) (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.

Barlow, H. B. (1983). Intelligence, guesswork, language. *Nature*, *304*, 207–209.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 405–426). Cambridge, England: Cambridge University Press.

Bates, E., & MacWhinney, B. (1979). A functionalist approach to the acquisition of grammar. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 167–209). New York: Academic Press.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–193). Hillsdale, NJ: Erlbaum.

Bekoff, M. & Byers, J. A. (Eds.) (1998). *Animal play: Evolutionary, comparative, and ecological perspectives*. Cambridge, England: Cambridge University Press.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In R. Hayes (Ed.), *Cognition and language development* (pp. 277–360). New York: Wiley & Sons.

Bickerton, D. (1984). The language bio-program hypothesis. *Behavioral and Brain Sciences*, *7*, 173–212.

Bickerton, D. (1995). *Language and human behavior*. Seattle, WA: University of Washington Press.

Boeckx, C. (2006). *Linguistic minimalism: Origins, concepts, methods, and aims*. New York: Oxford University Press.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*, 441–465.

Botha, R. P. (1999). On Chomsky's ''fable'' of instantaneous language evolution. *Language and Communication*, *19*, 243–257.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford, England: Oxford University Press.

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Brighton, H., Smith, K., & Kirby, S. (2005). Language as an evolutionary system. *Physics of Life Reviews*, *2*, 177–226.

Byrne, R. W., & Byrne, J. M. E. (1993). Complex leaf-gathering skills of mountain gorillas (*Gorilla g. berengei*): Variability and standardization. *American Journal of Primatology*, *31*, 241–261.

Carey, S., & Spelke, E. S. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

Carroll, S. B. (2001). Chance and necessity: The evolution of morphological complexity and diversity. *Nature*, *409*, 1102–1109.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge, England: Cambridge University Press.

Chater, N., Reali, F., & Christiansen, M. H. (2009). Restrictions on biological adaptation in language evolution. *Proceedings of the National Academy of Sciences*, *106*, 1015–1020.

Chater, N., & Vitányi, P. (2007). ''Ideal learning'' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, *51*, 135–163.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). *Rules and representations*. Oxford, England: Blackwell.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, *36*, 1–22.

Christiansen, M. H. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–558.

Christiansen, M. H., Chater, N., & Reali, F. (in press). The biological and cultural foundations of language. *Communicative and Integrative Biology*.

Clark, H. H. (1975). Bridging. In R. C. Schank & B. L. Nash-Webber (Eds.), *Theoretical issues in natural language processing* (pp. 169–174). New York: Association for Computing Machinery.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.

Comrie, B. (1989). *Language typology and language change*. Oxford, England: Blackwell.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, *5*, 539–546.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, *14*, 597–650.

Crain, S., Goro, T., & Thornton, R. (2006). Language acquisition is language change. *Journal of Psycholinguistic Research*, *35*, 31–49.

Crain, S., & Lillo-Martin, D. C. (1999). *An introduction to linguistic theory and language acquisition*. Oxford, England: Blackwell.

Crain, S., & Pietroski, P. (2001). Nature, nurture and universal grammar. *Linguistics and Philosophy*, *24*, 139–186.

Crain, S., & Pietroski, P. (2006). Is generative grammar deceptively simple or simply deceptive? *Lingua*, *116*, 64–68.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. New York: Oxford University Press.

Crowley, J., & Katz, L. (1999). Development of ocular dominance columns in the absence of retinal input. *Nature Neuroscience*, *2*, 1125–1130.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. New York: Oxford University Press.

Culicover, P. W., & Nowak, A. (2003). *Dynamical grammar*. Oxford, England: Oxford University Press.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: W. W. Norton.

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *Microcephalin* and *ASPM*. *Proceedings of the National Academy of Sciences*, *104*, 10944–10949.

Dienes, Z., & McLeod, P. (1993). How to catch a cricket ball. *Perception*, *22*, 1427–1439.

Dyer, F. C. (2002). The biology of the dance language. *Annual Review of Entomology*, *47*, 917–949.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Evans, N., & Levinson, S. (2008). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*.

Evans, J., & Saffran, J. R. (2005). *Statistical learning in children with specific language impairment*. Boston, MA: Paper presented at the Boston University Conference on Language Development.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*, 12203–12208.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology*, *41*, 145–170.

Field, D. J. (1987). Relations between the statistics of natural images and the response profiles of cortical cells. *Journal of the Optical Society of America*, *4*, 2379–2394.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (in press). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language*.

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: W.W. Norton.

Galef, B. G., & Laland, K. N. (2005). Social learning in animals: Empirical studies and theoretical models. *Bioscience*, *55*, 489–499.

Garcia, J., Kimeldorf, D. J., & Koelling, R. A.. (1955). Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science*, *122*, 157–158.

Geertz, C. (1973). *The interpretation of cultures: Selected essays*. New York: Basic Books.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. New York: Oxford University Press.

Goldberg, A. E. (2008). Universal Grammar? Or prerequisites for natural language? *Behavioral and Brain Sciences*, *31*, 522–523.

Golinkoff, R. M., Hirsh-Pasek, K., Bloom, L., Smith, L., Woodward, A., Akhtar, N., Tomasello, M., & Hollich, G. (Eds.) (2000). *Becoming a word learner: A debate on lexical acquisition*. New York: Oxford University Press.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, *4*, 178–186.

Gopnik, M., & Crago, M. B. (1991). Familial aggregation of a developmental language disorder. *Cognition*, *39*, 1–50.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: What early learning tells us about the mind*. New York: HarperCollins.

Gould, S. J. (1993). *Eight little piggies: Reflections in natural history*. New York: Norton.

Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, *426*, 435–439.

Griffiths, T. L., & Kalish, M. L. (2005). A Bayesian view of language evolution by iterated learning. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 827–832). Mahwah, NJ: Erlbaum.

Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, *56*, 61–98.

Harman, G., & Kulkarni, S. (2007). *Reliable reasoning: Induction and statistical learning theory*. Cambridge, MA: MIT Press.

Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Ecco/HarperCollins.

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*, 1569–1579.

Hauser, M. D., & Fitch, W. T. (2003). What are the uniquely human components of the language faculty? In M. H. Christiansen & S. Kirby (Eds.), *Language evolution* (pp. 158–181). Oxford, England: Oxford University Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.

Hawkins, J. A. (2004). *Complexity and efficiency in grammars*. Oxford, England: Oxford University Press.
Healy, S. D., & Hurly, T. A. (2004). Spatial learning and memory in birds. *Brain, Behavior and Evolution*, *63*, 211–220.
Healy, S. D., Walsh, P., & Hansell, M. (2008). Nest building in birds. *Current Biology*, *18*, R271–R273.
Holmes, W. G., & Sherman, P. W. (1982). The ontogeny of kin recognition in two species of ground squirrels. *American Zoologist*, *22*, 491–517.
Hopper, P., & Traugott, E. (1993). *Grammaticalization*. Cambridge, England: Cambridge University Press.
Hornstein, N., & Lightfoot, D. (Eds.) (1981). *Explanations in linguistics: The logical problem of language acquisition*. London: Longman.
Hsu, H.-J., Christiansen, M. H., Tomblin, J. B., Zhang, X., & Gómez, R. L. (2006). *Statistical learning of nonadjacent dependencies in adolescents with and without language impairment*. Madison, WI: Poster presented at the 2006 Symposium on Research in Child Language Disorders.
Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*, 151–195.
Hurley, S., & Chater, N. (Eds.) (2005). *Perspectives on imitation: From neuroscience to social science. Volume 1. Mechanisms of imitation and imitation in animals*. Cambridge, MA: MIT Press.
Jackendoff, R. (2000). *Foundations of language*. New York: Oxford University Press.
Jenkins, L. (2000). *Biolinguistics: Exploring the biology of language*. Cambridge, England: Cambridge University Press.
Karmiloff-Smith, A., & Inhelder, B. (1973). If you want to get ahead get a theory. *Cognition*, *3*, 195–212.
Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford, England: Oxford University Press.
Kirby, S. (2007). The evolution of meaning-space structure through iterated learning. In C. Lyon, C. Nehaniv, & A. Cangelosi (Eds.), *Emergence of communication and language* (pp. 253–268). Berlin: Springer Verlag.
Kirby, S., Dowman, M., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*, 5241–5245.
Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London: Springer Verlag.
Laurence, S., & Margolis, E. (2001). The poverty of the stimulus argument. *British Journal for the Philosophy of Science*, *52*, 217–276.
Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, *23*, 379–434.
Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.
Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity theory and its applications* (2nd ed.). Berlin: Springer.
Lieberman, P. (1984). *The biology and evolution of language*. Cambridge, MA: Harvard University Press.
Lightfoot, D. (2000). The spandrels of the linguistic genotype. In C. Knight, M. Studdert-Kennedy, & J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 231–247). Cambridge, England: Cambridge University Press.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.
Marler, P. & Slabbekoorn, H. (Eds.) (2004). *Nature's music: The science of birdsong*. San Diego, CA: Elsevier.
Mesoudi, A., Whiten, A., & Laland, K. (2006). Toward a unified science of cultural evolution. *Behavioral and Brain Sciences*, *29*, 329–383.
Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data (TILAR Series)* (pp. 139–163). Amsterdam: John Benjamins.

Morgan, J. L., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Müller, R.-A. (2009). Language universals in the brain: How linguistic are they? In M. H. Christiansen, C. Collins, & S. Edelman (Eds.), *Language universals* (pp. 224–252). New York: Oxford University Press.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*, 329–336.

Nettle, D. (1999). *Linguistic diversity*. Oxford, England: Oxford University Press.

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Nowak, M. A., Komarova, N. L., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*, 114–118.

O'Grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Mahwah, NJ: Erlbaum.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*, 222–231.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, *233*, 435–449.

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: From ''learning'' to parameter setting in biology and in the study of language. *Cognition*, *31*, 1–44.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–225.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow and Company.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Brain and Behavioral Sciences*, *13*, 707–727.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*, 201–236.

Pinker, S., & Jackendoff, R. (2009). The components of language: What's specific to language, and what's specific to humans? In M. H. Christiansen, C. Collins, & S. Edelman (Eds.), *Language universals* (pp. 126–151). New York: Oxford University Press.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance volume 2: Cognitive processes and performance* (pp. 36–1–36-46). New York: Wiley.

Pullum, G. K., & Scholz, B. (2002). Empirical assessment of stimulus poverty arguments. *Linguistic Review*, *19*, 9–50.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, *29*, 1007–1028.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. Chicago: Chicago University Press.

Reuland, E. (2008). Why neo-adaptationism fails. *Behavioral and Brain Sciences*, *31*, 531–532.

Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: Chicago University Press.

Roberts, M., Onnis, L., & Chater, N. (2005). Language Acquisition and Language Evolution: Two puzzles for the price of one. In M. Tallerman (Ed.), *Prerequisites for the evolution of language* (pp. 334–356). Oxford, England: Oxford University Press.

Roth, F. P. (1984). Accelerating language learning in young children. *Child Language*, *11*, 89–107.

de Ruiter, J. P., & Levinson, S. C. (2008). A biological infrastructure for communication underlies the cultural evolution of language. *Behavioral and Brain Sciences*, *31*, 518.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences*, *102*, 2661–2665.

Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Searcy, W. A., & Nowicki, S. (2001). *The evolution of animal communication: Realiability and deception in signaling systems*. Princeton, NJ: Princeton University Press.

Seidenberg, M. S., & MacDonald, M. (2001). Constraint-satisfaction in language acquisition. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 281–318). Westport, CT: Ablex.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*, 1779–1782.

Shadmehr, R., & Wise, S. P. (2005). *The computational neurobiology of reaching and pointing: A foundation for motor learning*. Cambridge, MA: MIT Press.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York: Holt, Rinehart & Winston.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*, 238–299.

Stephens, D. W., Brown, J. S., & Ydenberg, R. C. (2007). *Foraging: Behavior and ecology*. Chicago: University of Chicago Press.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*, 299–351.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of cognition and perception* (pp. 217–262). San Diego, CA: Academic Press.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In M. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems 11* (pp. 59–65). Cambridge, MA: MIT Press.

Tenenbaum, J. B., Kemp, C., & Shafto, P. (2007). Theory-based Bayesian models of inductive reasoning. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 114–136). Cambridge, England: Cambridge University Press.

Tomasello, M. (2000). Do you children have adult syntactic competence? *Cognition*, *74*, 209–253.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–691.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural learning in adolescents with and without specific language impairment. *Language Learning and Development*, *3*, 269–293.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergartenpath effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*, 399–433.

Van der Lely, H. K. J., & Battell, J. (2003). WH-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language*, *79*, 153–181.

Wang, D., & Li, H. (2007). Nonverbal language in cross-cultural communication. *Sino-US English Teaching*, *4*, 66–70.

Weber, B. H. & Depew, D. J. (Eds.) (2003). *Evolution and learning: The Baldwin effect reconsidered*. Cambridge, MA: MIT Press.

Weissenborn, J. & Höhle, B. (Eds.) (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Philadelphia, PA: John Benjamins.

Wexler, K. (2004). Lennenberg's dream. In L. Jenkins (Ed.), *Variation and universals in biolinguistics* (pp. 239–284). Amsterdam: Elsevier.

Wilson, E. O. (1971). *The insect societies*. Cambridge, MA: Harvard University Press.

Wynne, T., & Coolidge, F. L. (2008). A stone-age meeting of minds. *American Scientist*, *96*, 44–51.

Yang, C. (2002). *Knowledge and learning in natural language*. New York: Oxford University Press.

Zipf, G. K. (1949). *Human behavior and the principle of least-effort*. Cambridge, MA: Addison-Wesley.

Zuidema, W. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 51–58). Cambridge, MA: MIT Press.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1158–1184

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01127.x

# How Infants Learn About the Visual World

Scott P. Johnson

*Department of Psychology, University of California, Los Angeles*

Received 22 August 2008; received in revised form 30 April 2010; accepted 20 May 2010

**Abstract**

The visual world of adults consists of objects at various distances, partly occluding one another, substantial and stable across space and time. The visual world of young infants, in contrast, is often fragmented and unstable, consisting not of coherent objects but rather surfaces that move in unpredictable ways. Evidence from computational modeling and from experiments with human infants highlights three kinds of learning that contribute to infants' knowledge of the visual world: learning via association, learning via active assembly, and learning via visual-manual exploration. Infants acquire knowledge by observing objects move in and out of sight, forming associations of these different views. In addition, the infant's own self-produced behavior—oculomotor patterns and manual experience, in particular—is an important means by which infants discover and construct their visual world.

*Keywords:* Visual development; Cognitive development; Models of development; Object perception; Infants; Learning

## 1. Introduction

When we look around us, we encounter environments characterized by numerous objects and people at varying distances. The scene depicted in Fig. 1 is typical. It shows a garden in Southern California occupied by flora, people, and artifacts such as buildings and walkways. The scene is busy and cluttered. The objects have multiple parts and are located at various distances from the observer; nearer objects obscure farther objects—the trees hide part of the building's roof, and many of the people on the patio cannot be seen. In some instances, color, texture, size, and shape can serve as information for the unity of objects. The leaves on the trees, for example, are all roughly similar in appearance, and they are perceived as

Correspondence should be sent to Scott P. Johnson, Department of Psychology, Franz Hall, UCLA, Los Angeles, CA 90095. E-mail: scott.johnson@ucla.edu

Fig. 1. A visual scene.

grouping together. In other instances, however, we see objects as unified despite considerable discrepancies in these kinds of visual information: The girls on the stairs wear blue shorts and red shirts, yet we do not see these distinctions in color as denoting four ''parts'' of girls, but instead we see them as belonging to objects in common. In real-world scenes, motion of observers and of objects provides additional information to determine the contents of our surroundings. We can move through the environment, obtain new perspectives, and see parts of objects invisible from previous vantage points. And as objects move, we can track them across periods of temporary invisibility, often predicting their reappearance.

These facts about the visual world are at once ordinary and remarkable. They are ordinary because virtually every sighted observer experiences the environment as composed of separate, distinct objects of varying complexity and appearance. Yet they are nonetheless remarkable because of the intricate cortical machinery needed to produce them (Zeki, 1993): Several dozen areas of the brain, each responsible for processing a distinct aspect of the visual scene or coordinating the outputs of other areas, working in concert to yield a more-or-less seamless and coherent experience. Action systems are likewise elaborate (Gibson, 1950): eyes, head, and body, each with distinct control systems, working in concert to explore the visual environment.

In this article, I consider theory and research on the development of infants' perception of the visual environment, in particular object perception. The garden example illustrates some of the issues faced by researchers who wish to better understand these processes. Visual perception has a ''bottom-up'' foundation built upon coding and integrating distinctive kinds of visual information (variations in color, luminance, distance, texture, orientation, motion, and so forth). Visual perception also relies on ''top-down'' knowledge that observers bring to the scene, knowledge that aids in interpreting potentially ambiguous juxtapositions of visual attributes (such as the blue and red clothing on the girls). Both operate continuously in mature, sighted individuals who have had sufficient time and experience with which to learn about specific kinds of objects.

Young infants have not had as much time and experience with which to learn about objects, yet they inhabit the same world as adults and they encounter the same kinds of visual information. How do infants interpret visual scenes? Are they restricted to bottom-up processing, lacking the cognitive capacity to interpret visual information in a meaningful fashion, however that might be defined? Or might there be some capacity to perceive objects that is independent of visual experience? These questions have long been dominated by a tension between arguments for a learned or *constructed* versus an unlearned or *innate* system of object knowledge in humans. This article will examine the question of infants' object perception by considering these theoretical perspectives, and it attempts to answer the question with evidence from modeling of developmental processes and from empirical studies. I will restrict discussion to the developmental origins, in humans, of the ability to represent objects as coherent and complete across space and time—that is, despite partial or full occlusion—a definition of object perception akin to the *object concept* that was originally described by the eminent developmental psychologist Jean Piaget (more on this subsequently). The limited scope necessarily omits many interesting literatures on other topics related to object knowledge, such as object identity, numerosity, animacy, object-based attention, and so forth. Nevertheless, there has been a great deal of research effort directed at object concept development, and these investigations continue to bear on the question of infants' object perception by providing an increasingly rich base of evidence.

## 2. Theoretical considerations: Constructivism versus nativism

A perennial question of infant epistemology is the extent to which knowledge necessarily develops over time through learning and experience or whether some kinds of knowledge are available—''built in''—without any experience. This question has motivated innumerable experiments investigating infant knowledge of the physical and social environment (for recent reviews, see Carey, 2009; Johnson, 2010), and it has not yet been fully satisfied, although there have been suggestions to abandon it in light of mounting evidence from systems theory for multiple levels of developmental process (Spencer et al., 2009). I will illustrate these opposing views by outlining two theories of infant object perception: constructivist theory and nativist theory.

### *2.1. Constructivist theory*

Piaget was the first to provide a detailed theory of development of object knowledge in humans, and he amassed a great deal of evidence to support it (Piaget, 1954). The evidence was derived from observations of infants and young children as he engaged them in child-friendly games, such as hiding a desired toy under a cover and then observing the child's attempts to retrieve it. Piaget suggested that knowledge of objects and knowledge of space were inseparable—without knowledge of spatial relations, there could be no knowledge of objects *qua* distinct entities. Instead, objects, early in life, were disconnected forms that moved capriciously and randomly.

*Objectification* was Piaget's term for knowledge of the self and external objects as distinct entities, spatially segregated, persisting across time and space, and obeying causal constraints. Piaget suggested that objectification is rooted in the child's recognition of her own body as an independent object and her own movements as movements of objects through space, corresponding to movements of other objects she sees. This recognition, in turn, was thought to stem exclusively from exploration of objects via manual activity—that is, object knowledge is constructed by the child.

Because skilled reaching and grasping of objects is not available until some time after 4–6 months, Piaget proposed that object knowledge likewise is beyond the ken of young infants. Until this time, objects consist of little more than fleeting images, as noted previously. Active, intentional search behavior marks the inception of objectification, beginning with infants' visual accommodation to dropped objects (by looking at the floor), and culminating with infants' systematic, organized search for hidden objects in multiple possible locations (in one of Piaget's more complicated games) some time during the second year. As the action systems develop, therefore, so develops the capacity to interact with objects and discover their properties, most importantly the maintenance of object properties across disruptions in perceptual contact, as when objects go out of sight.

Piaget's theory enjoys strong support for many of the details of behavior that he so assiduously captured, and for raising awareness of the problems faced by infants as they navigate the visual world. The reasoning behind the developmental changes in behavior, however, has not seen the same level of enthusiasm. Numerous experiments have revealed that by 2–4 months, infants appear to maintain representations of partly and fully hidden objects across short delays, somewhat younger than allowed for on Piaget's account, and inconsistent with the emphasis on manual activity as the principal agent of developmental change. These experiments have led to views of infant object knowledge as relying on unlearned, or innate, foundations, and some of these views are described in the following section.

### *2.2. Nativist theory*

A central tenet of nativist theory is that some kinds of initial, unlearned knowledge form a central core around which more diverse, mature cognitive capacities are elaborated (Spelke, 1990; Spelke, Breinlinger, Macomber, & Jacobson, 1992). That is, some kinds of knowledge, including concepts of objects as coherent and continuous across occlusion, are *innate*. Philosophical discussions of innateness are ancient. Plato and Descartes, for example, proposed that some ideas, such as concepts of geometry or God or justice, were universal and available innately because they were unobservable or arose in the absence of any direct tutoring or instruction. With respect to infant knowledge, the focus of modern nativist theory is on *learnability*: According to this line of reasoning, in the absence of any viable theory of how humans come to understand object concepts so quickly after birth, in some cases well in advance of the manual search skills emphasized by Piaget, the assumption is that these concepts necessarily arise independent of experience.

Researchers of a nativist persuasion have offered three arguments for hypothesized innate object concepts. First, veridical object knowledge can be elicited in very young infants—as

young as 2 months of age, or perhaps even at birth—under a variety of circumstances, suggesting that early concepts emerge too quickly to have derived from postnatal learning. Second, infants' acquisition of object knowledge has been proposed to arise from *contrastive evidence*: opportunities to observe conditions under which an object behaves in a manner consistent or inconsistent with a particular concept (Baillargeon, 1994). On this view, a concept of persistence across occlusion, for example, must be innate, because there are no opportunities in the real world to observe objects going out of existence! Third, there is evidence from one nonhuman species—domestic chickens—for an unlearned capacity for *unity perception*, recognition of partly occluded shapes as similar to an unoccluded version of the same form (Regolin & Vallortigara, 1995). (In this experiment, newly hatched chicks were imprinted on a partly occluded cardboard triangle, and they subsequently chose to associate with a fully visible version of the triangle, rather than a version consisting of the previously seen triangle fragments. An experiment in which the imprinting/association objects were reversed showed the same result.)

As noted, there is compelling evidence from a variety of laboratories and experimental settings for representations of objects as solid entities that are spatiotemporally coherent and persistent by 2–4 months after birth. (Some of these experiments will be discussed in detail in subsequent sections of this article.) Nevertheless, the *developmental origins* of object concepts in human infants as a topic of investigation cannot be dismissed merely by noting competence in these experiments at a young age. Moreover, suggesting that infants learn about objects through only a single means, such as contrastive evidence, does not seem realistic. Unequivocal support for *innate* object concepts in humans would come from evidence for their emergence in the absence of visual experience—say, functionality at birth, or veridical object perception in blind individuals who have their sight restored. As we will see, experiments on infants' perception of partly occluded objects cast doubt on the viability of any of these varieties of innateness as the best descriptor of the development of object concepts. (Parenthetically, it is worth noting as well that the finding of unity perception in chicks remains an isolated phenomenon in the literature, inconsistent with experiments with another avian species—pigeons—which, as adults, apparently see partly occluded objects only in terms of their visible surfaces, not as having hidden parts; Sekuler, Lee, & Shettleworth, 1996.)

### *2.3. Summary*

Piaget set the stage for decades of fruitful research that has established the availability of functional object concepts in the first year after birth, and Piaget's own theory of how knowledge may be constructed by the child's own behavior has been tremendously influential. Numerous experiments in the past few decades, however, have suggested that Piaget underestimated young infants' capacity to perceive object unity and boundaries under occlusion, so much so that alternate theories stressing innate contributions to object knowledge have appeared.

In the next section of this article, I describe evidence for developmental change early in postnatal life in how infants respond to partly and fully hidden objects. Evidence comes

from experiments that assess three kinds of perceptual completion in infancy: *spatial completion* (perception of partly hidden surfaces as continuous), *spatiotemporal completion* (perception of objects as continuing to exist upon becoming occluded), and *3D object completion* (perception of objects as solid in 3D space, with a back and sides that cannot be viewed from any one vantage point). Because evidence for developmental change requires an explanation, in the sections after that I describe models of development and empirical investigations of developmental mechanisms in infants. These models and investigations posit a central role for learning, and they suggest specific means by which learning—in models and in human infants—can lead to the kinds of object knowledge I have discussed.

## 3. Developmental change in infants' object perception

Piaget's observations led him to conclude that newborn infants have no true concepts of objects or space (Piaget, 1952). Neonates can discriminate among visible objects and track their motions, but when objects move out of sight or the baby's gaze falls elsewhere, previously encountered objects cease to exist for the infant. The first inklings of object concepts come from recognition memory, say when infants smile upon mother's return, beginning a few months after birth. Knowledge of objects as complete and coherent across gaps in perceptual contact imposed by occlusion did not come until later, after infants can grasp and reach objects and thereby come to more fully appreciate properties such as solidity, volume, and existence independent of the infant. These concepts did not come all at once. Piaget described one important concept as ''reconstruction of an invisible whole from a visible fraction,'' and it was evinced by retrieval of an object from under a cover when only a part of the object was visible. An appreciation of continued existence despite complete occlusion was evinced by removal of a obstacle hiding a desired toy, or pulling away a cover from a parent's face during peekaboo. These behaviors were not seen consistently until 6–8 months or so, marking for Piaget the advent of a wholly new set of object and spatial concepts.

As noted previously, research over the past several decades has yielded a wealth of evidence from multiple measures (e.g., looking, reaching, and cortical activity) making it clear that infants represent object properties even when the objects are partly or fully hidden, and these findings have led, in turn, to nativist theoretical views that have sought to overturn Piagetian ideas about how these representations arise in infants. Yet two important facts remain, facts suggesting that Piagetian theory may not be so far off the mark as concerns developmental changes in infants' object knowledge. First, newborn infants do not seem to perceive partly occluded objects as having hidden parts. Instead, neonates construe such stimuli solely in terms of their visible parts, failing to achieve spatial completion (Slater, Johnson, Brown, & Badenoch, 1996; Slater et al., 1990; but see Valenza, Leo, Gava, & Simion, 2006). There is a clear developmental progression in perceptual completion (Johnson, 2004), and this calls for an explanation of underlying mechanisms of change. And second, the majority of research on infants' perceptual completion in reality is broadly consistent with Piaget's observations: Infants provide evidence of representing partly occluded objects a few months after birth, and fully occluded objects by about the middle of the first

year. Some have claimed object permanence in infants on the basis of evidence from looking time studies (e.g., Baillargeon, 2008), but the short-term representations of hidden objects demonstrated in such experiments fall short of Piaget's criteria for full object permanence: accurate search in multiple locations for a hidden object, demonstrating knowledge of object persistence and the spatial relations between the object, the hiding locations, and the infant (Haith, 1998; Kagan, 2008).

In the remainder of this section, I will describe investigations of perceptual completion in infants. These investigations provide clear evidence for developmental change in how infants perceive occlusion events.

### *3.1. Spatial completion*

Adults and 4-month-old infants construe the ''rod-and-box'' display depicted in Fig. 2, left, as consisting of two parts, a single elongated object moving back and forth behind an occluding box (Kellman & Spelke, 1983). Neonates, in contrast, construe this display as consisting of *three* parts: two distinct object parts and occluder (Slater et al., 1990, 1996). These conclusions stem from looking time experiments in which infants first view the rod-and-box display repeatedly until habituation of looking occurs, defined as a decline in looking times toward the display (judged by an observer) according to a predetermined criterion. Following habituation, infants see two new displays, and their posthabituation looking patterns are thought to reflect a novelty preference. The 4-month-olds and neonates showed opposite patterns of preference. Looking longer at the ''broken'' rod parts indicates that they were relatively novel compared to the rod-and-box display—the 4-month-olds' response, suggestive of unity perception. Looking longer at the ''complete'' rod indicates that infants likely construed the rod-and-box display as composed of disjoint objects—the newborns' response. These results led to the more general conclusion that neonates are unable to perceive occlusion, and that occlusion perception emerges over the first several postnatal months (Johnson, 2004). That is, ''piecemeal'' or fragmented perception of the visual environment extends from birth through the first several months afterwards.

Two-month-olds were found to show an ''intermediate'' pattern of performance—no reliable posthabituation preference—implying that spatial completion is developing at this

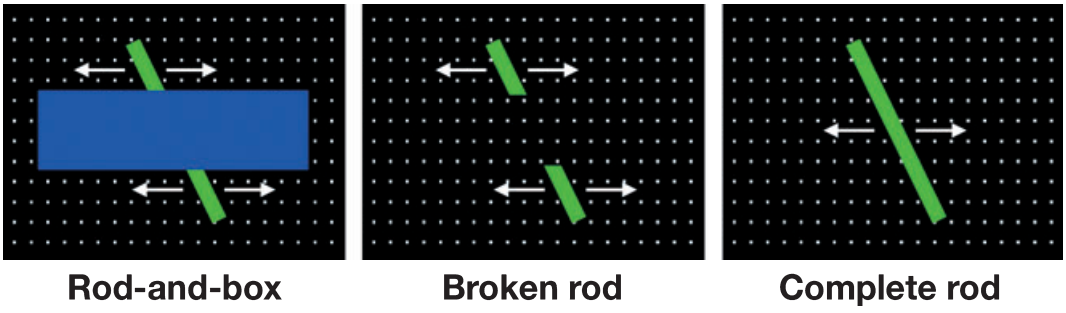


Fig. 2. Displays used in experiments that investigate spatial completion in young infants. Adapted from Johnson and Náñez (1995).

point but not yet complete (Johnson & Náñez, 1995). Additional studies examined the possibility that 2-month-olds will perceive unity if given additional perceptual support. The amount of visible rod surface revealed behind the occluder was enhanced by reducing box height and by adding gaps in it, and under these conditions 2-month-olds provided evidence of unity perception (Johnson & Aslin, 1995). (With newborns, however, this manipulation failed to reveal similar evidence—even with enhanced displays, newborns perceived the moving rod parts as disjoint objects; Slater et al., 1996; Slater, Johnson, Kellman, & Spelke, 1994.) These experiments served to pinpoint more precisely the time of emergence of spatial completion in infancy: the first several weeks or months after birth under typical circumstances.

Additional experiments explored the kinds of visual information infants use to perceive spatial completion. Kellman and Spelke (1983) reported that 4-month-olds perceived spatial completion only when the rod parts, with aligned outer edges, moved in tandem behind a stationary occluder. We replicated and extended this finding, showing in addition that 4-month-olds provided evidence of completion only when the rod parts were aligned (Johnson & Aslin, 1996). Later experiments revealed similar patterns of performance in 2-month-olds when tested using displays with different occluder sizes and edge arrangements, as seen in Fig. 3 (Johnson, 2004). Infants provided evidence of spatial completion obtained only when rod parts were aligned across a narrow occluder; in the other displays, infants provided evidence of disjoint surface perception.

One possible interpretation of these findings is that alignment, motion, and occluder width (i.e., the spatial gap) are interdependent contributions to spatial completion, such that common motion is detected most effectively when rod parts are aligned (Kellman & Arterberry, 1998). I examined this possibility by testing 2-month-olds' discrimination of different patterns of rod motion with varying orientations of rod parts and occluder widths. Under all tested conditions, infants discriminated the motion patterns, implying that motion discrimination was neither impaired nor facilitated by misalignment or occluder width (Johnson, 2004). It might be that motion contributes to infants' spatial completion in multiple ways, first serving to segment the scene into its constituent surfaces, and then serving to bind moving surfaces into a single object (Johnson, Davidow, Hall-Haro, & Frank, 2008).

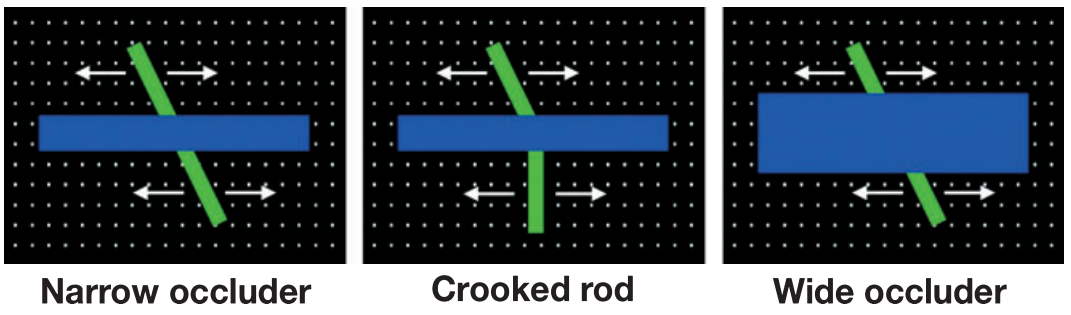


Fig. 3. Displays used to test the roles of occluder size and edge alignment in 2-month-olds' perception of spatial completion. Adapted from Johnson (2004).

In summary, experiments that explored development of spatial completion suggest that young infants analyze the motions and arrangements of visible surfaces. At birth, newborns perceive partly occluded surfaces as separate from one another and the background. Only later do infants integrate these surfaces into percepts of coherent, partly occluded objects. On this view, therefore, development of object knowledge begins with perception of visible object components, and it proceeds with increasing proficiency at representation of those object parts that cannot be discerned directly.

### *3.2. Spatiotemporal completion*

A number of studies using different methods have shown that young infants can maintain representations for hidden objects across brief delays (e.g., Aguiar & Baillargeon, 1999; Berger, Tzur, & Posner, 2006; Clifton, Rochat, Litovsky, & Perris, 1991). Yet newborn infants provide little evidence of spatial completion, raising the question of how perception of *complete* occlusion emerges during the first few months after birth. Apart from Piaget's observations, this question has received little serious attention until recently, in favor of accounts that stress innate object concepts (e.g., Baillargeon, 2008; Spelke, 1990).

To address this gap in our knowledge, my colleagues and I conducted experiments with object trajectory displays, asking whether infants perceive the trajectory as continuous across occlusion—spatiotemporal completion. We reasoned that manipulation of spatial and temporal characteristics of the stimuli, and observation of different age groups, might provide insights into development of spatiotemporal completion, as they did in the case of spatial completion.

These investigations revealed a fragmented-to-holistic developmental pattern and revealed spatial and temporal processing constraints as well, both sets of results in parallel with the investigations of spatial completion described in the previous section. Spatiotemporal completion was tested using similar methods: habituation to an occlusion display (Fig. 4), followed by broken and complete test displays, different versions of the partly hidden trajectory seen during habituation. At 4 months, infants treat the ball-and-box display depicted in Fig. 4 as consisting of two disconnected trajectories, rather than a single, partly hidden path (Johnson, Bremner, et al. 2003); evidence comes from a reliable preference for the continuous version of the test trajectory. By 6 months, infants perceived this trajectory as unitary, as revealed by a reliable preference for the discontinuous trajectory test stimulus. When occluder size was narrowed, however, reducing the spatiotemporal gap across which the trajectory had to be interpolated, 4-month-olds' posthabituation preferences (and thus, by inference, their percepts of spatiotemporal completion) were shifted toward the discontinuous, partway by an intermediate width, and fully by a narrow width, so narrow as to be only slightly larger than the ball itself. In 2-month-olds, this manipulation appeared to have no effect.

Reducing the spatiotemporal gap, therefore, facilitates spatiotemporal completion. Reducing the *temporal* gap during which an object is hidden, independently from the *spatial* gap, also supports spatiotemporal completion. Increasing the ball size (Fig. 5) can minimize the time out of sight as it passes behind the occluder, and this led 4-month-olds to perceive

its trajectory as complete. Accelerating the speed of a smaller ball as it passed behind the occluder (and appeared more quickly) had a similar effect (Bremner et al., 2005). On the other hand, altering the orientation of the trajectory impaired path completion (Fig. 5), unless the edges of the occluder were orthogonal to the path; these findings are similar to outcomes of experiments on edge misalignment described in the previous section (Bremner et al., 2007).

This work leads to three conclusions. First, spatiotemporal completion proceeds from processing parts of paths to complete trajectories. Second, there may a lower age limit for trajectory completion (between 2 months and 4 months), just as there appears to be for spatial completion (between birth and 2 months). Third, young infants' spatiotemporal completion is based on relatively simple parameters. Either a short time or short distance out of sight leads to perception of continuity, and this may occur because the processing load is reduced by these manipulations. The fragile nature of emerging spatiotemporal completion is underscored as well by results showing its breakdown when either occluder or path orientation is nonorthogonal.

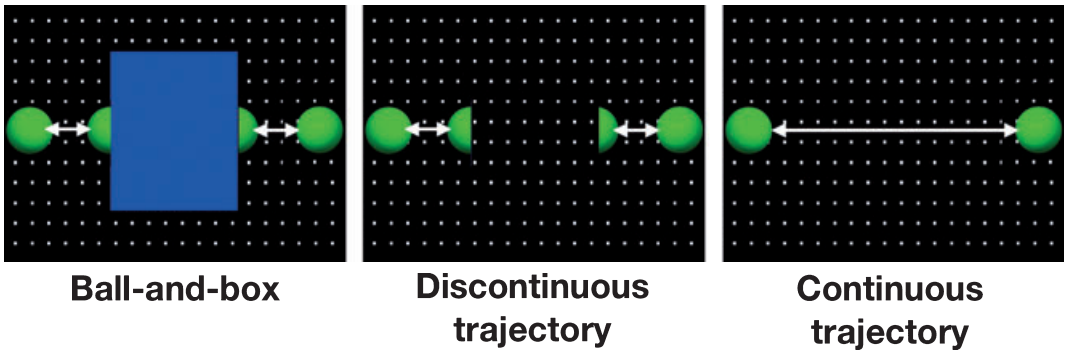


Fig. 4. Displays used in experiments that investigate spatiotemporal completion in young infants. Adapted from Johnson, Bremner, et al. (2003).

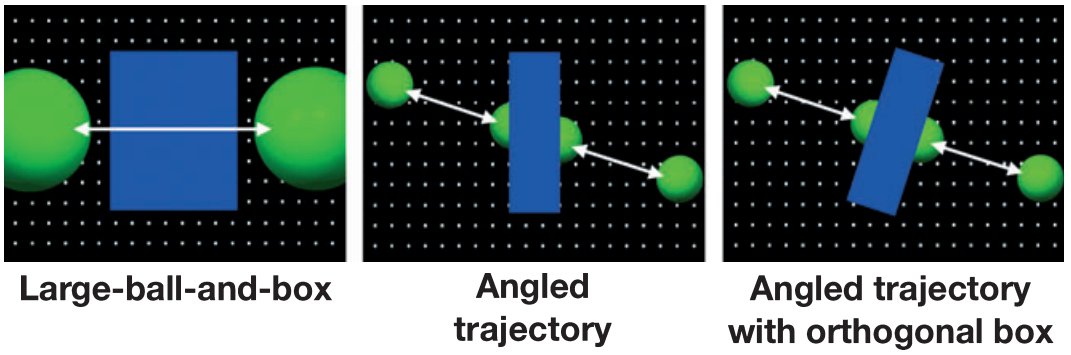


Fig. 5. Displays used to test the roles of occlusion duration and path orientation in 4-month-olds' perception of spatiotemporal completion. Adapted from Bremner et al. (2005, 2007).

### *3.3. 3D object completion*

Spatial and spatiotemporal completion consist of filling in the gaps in object surfaces that have been occluded by nearer ones. Solid objects also occlude parts of themselves such that we cannot see their hidden surfaces from our present vantage point, yet our experience of most objects is that of filled volumes rather than hollow shells. Perceiving objects as solid in three-dimensional space despite limited views constitutes 3D object completion. In contrast to spatial and spatiotemporal completion, little is known about development of 3D object completion. We recently addressed this question with a looking time paradigm similar to those described previously (Soska & Johnson, 2008). Four- and six-month-olds were habituated to a wedge rotating through 15 degrees around the vertical axis such that the far sides were never revealed (Fig. 6). Following habituation infants viewed two test displays in alternation, one an incomplete, hollow version of the wedge, and the other a complete, whole version, both undergoing a full 360 degree rotation revealing the entirety of the object shape. Four-month-olds showed no consistent posthabituation preference, but 6-month-olds looked longer at the hollow stimulus, indicating perception of the wedge during habituation as a solid, volumetric object in 3D space.

In a follow-up study (Soska, Adolph, & Johnson, 2010), we used these same methods with a more complex stimulus: a solid ''L''-shaped object with eight faces and vertices, as opposed to the five faces and six vertices in the wedge-shaped object described previously (Fig. 7). We tested 4-, 6-, and 9.5-month-olds. As in the Soska and Johnson (2008) study with the wedge stimulus, we found a developmental progression in 3D object completion: 4-month-olds' posthabituation looking times revealed no evidence for completion, whereas 9.5-month-olds consistently looked longer at the hollow test display, implying perception of the habituation object as volumetric in 3D space. At 6 months, interestingly, only the male infants showed this preference; females looked about equally at the two test displays. At 9.5 months, the male advantage had disappeared: Both males and females looked longer at the hollow shape.

One interpretation of the sex difference at 6 months is that infants who were successful at 3D object completion engaged in *mental rotation* in this task: manipulation of a mental image of the object and imagining it from a different perspective. Mental rotation is a cognitive skill for which men have an advantage relative to women (Shepard & Metzler, 1971),

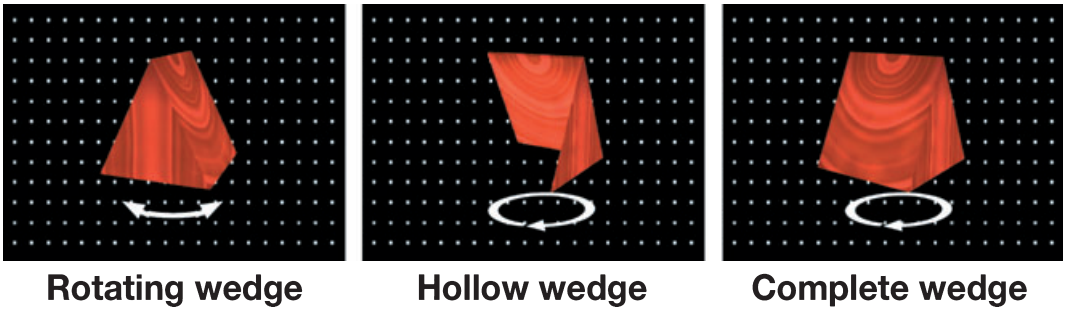


Fig. 6. Displays used in experiments that investigate 3D object completion in infants. Adapted from Soska and Johnson (2008).

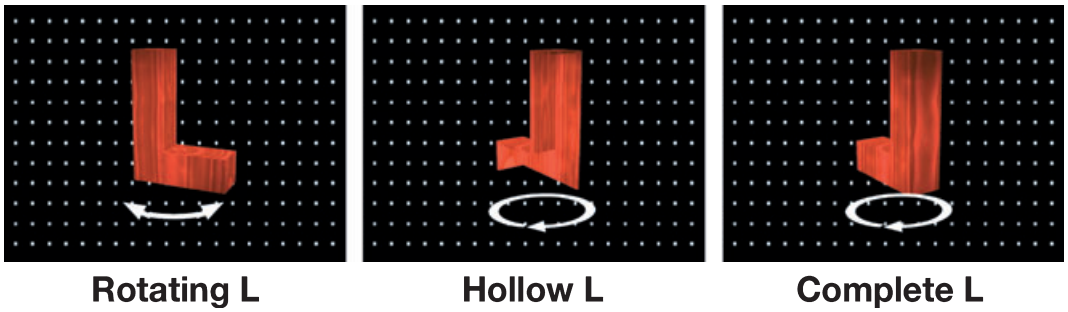


Fig. 7. Displays used in experiments that investigate 3D object completion in infants, with a more complex object relative to the Soska and Johnson (2008) study. Adapted from Soska and Johnson (unpublished data).

and two recent reports have provided evidence of a male advantage in young infants as well (Moore & Johnson, 2008; Quinn & Liben, 2008). It remains to be determined definitely whether mental rotation is involved in 3D object completion.

In summary, these data provide evidence for a developmental progression in infants' 3D object completion abilities, and for a role for stimulus complexity in infant performance. Both effects are consistent with the work on spatial and spatiotemporal completion described previously.

### *3.4. To be explained*

The research described in this section can be summarized as follows. Infants are born with a functional visual system sufficient to identify distinct regions of the visual scene and discriminate different regions from one another. Newborns can detect edges and motion, and there is even a rudimentary capacity for depth perception (e.g., size and shape constancy; Slater, 1995). Yet newborns do not perceive objects as do adults, and therefore they do not ''know'' the world to consist of overlapping objects at different distances that have hidden parts. These kinds of knowledge arise over the first several postnatal months. I described three kinds of perceptual completion—spatial, spatiotemporal, and 3D object completion— and described evidence for a developmental progression in each.

How does this happen? One way to deal with this question is to ignore or deny it, which in essence is the nativist position as it is commonly presented in the infant cognition literature (e.g., Spelke, 1990; Spelke & Kinzler, 2007; Spelke et al., 1992). Considering cognition more broadly, this is not necessarily an illegitimate approach. There are phenomena in the literature that would appear to be impossible to explain otherwise, such as the emergence of linguistic structures in the absence of any input (Goldin-Meadow & Mylander, 1998; Senghas, Kita, & Özyürek, 2004). Such instances are rare, however, and this fact leads many of us to a second way to deal with the question of origins of object knowledge: Confront it and determine what is needed to account for the developmental progression I have presented.

In the remainder of the article, I consider arguments and evidence in favor of a learning account of how infants come to perceive object occlusion. The hypothesis is that infants learn certain features of the world such as edges (and perhaps faces) from prenatal

developmental mechanisms tantamount to a kind of ''visual experience.'' Occlusion, in contrast, must be learned postnatally.

## 4. Evidence for prenatal ''visual experience'' and memory

Newborn infants have two prerequisite skills for the ability to learn occlusion: They are born seeing and they are born with the capacity for short-term recall. The oculomotor system is sufficiently functional to guide the point of gaze to desired targets in the environment. Three kinds of stimulus are particularly salient: high-contrast edges, motion, and faces (Kessen, Salapatek, & Haith, 1972; Slater, Morison, Town, & Rose, 1985; Valenza, Simion, Macchi Cassia, & Umiltà, 1996). Newborn infants look at such stimuli preferentially—that is, they tend to look longer at high-contrast contours and patterns of motion relative to homogenous regions, for example, and at face-like patterns relative to arrangements of similar features that do not match human faces (Slater, 1995; Slater et al., in press). The developmental mechanisms that yield these behaviors, consequently, must be in effect prior to birth.

An interesting fact about prenatal visual development prior to the onset of patterned visual input is that there is spontaneous yet organized activity in visual pathways from early on, activity that contributes to retinotopic ''mapping'' (Sperry, 1963). Mapping refers to the preservation of sensory structure, for example, the relative positions of neighboring points of visual space, from retina through the thalamus, primary visual cortex, and higher visual areas. One way in which mapping occurs is by ''waves'' of coordinated, spontaneous firing of receptors in the retina, prior to eye opening, observed in some nonhuman species such as chicks and ferrets (Wong, 1999). Waves of activity are propagated across the retinal surface at a point in development after connections to higher visual areas have formed; the wave patterns are then systematically propagated through to the higher areas. This might be one way by which correlated inputs remain coupled and dissimilar inputs become dissociated, and a likely means by which edges (which can be defined as simple, local interactions in the input) can be detected upon exposure to patterned visual scenes (Albert, Schnabel, & Field, 2008). Retinal waves also can serve as a foundation for development of representations of more complex visual patterns as infants gain exposure to the environment (Bednar & Miikkulainen, 2007).

Evidence for short-term memory at birth was presented previously: Neonates habituate to repeated presentation of stimuli, implying recognition of familiar patterns, and they recover interest to new stimuli, implying recognition of novel patterns. The rudiments of memory, therefore, undergo developments prenatally sufficient to support recognition across brief delays. These developments include the emergence and strengthening of neural connections within and between regions of cortical and subcortical regions known in adult primates to maintain activity to visual and spatial input across temporary delays in presentation. These regions include areas of prefrontal cortex (e.g., the principal sulcus, anterior arcuate, and inferior convexity) and their connections to areas that are involved in object and face processing, such as the temporal lobe (Goldman-Rakic, 1987, 1996).

The newborn baby, therefore, is equipped with perceptual and cognitive mechanisms sufficient to begin the process of learning about objects—to detect edges in the visual scene, to track motion, to recognize familiar items, and to discriminate different items presented simultaneously or in sequence. These facts have motivated a number of attempts to explore the developmental process of coming to perceive and act on objects using connectionist or other kinds of computational models. Some of these models are described next.

## 5. Modeling developmental processes

How can computational models help us understand the developmental process of occlusion perception? Models can address questions of object perception development and other developmental phenomena by constraining hypotheses about preexisting skill sets, the necessary inputs from the environment, and specific learning regimens, and how these considerations influence learning outcomes. Models have five features in common: (a) specification of a starting point, (b) a particular kind of training environment, (c) a particular means of gathering and retaining information, (d) a particular means of expressing acquired knowledge, and (e) learning from experience—that is, modification of responses based on some change in internal representations of the external environment that stem from feedback. These features can be manipulated by the modeler to shed light on the developmental process, analogous to variations in experimental designs in the laboratory. (For a recent review of models of infant object perception, see Mareschal & Bremner, 2009.)

Models of object perception development that have been presented in the literature learn by association in a simple two- or three-dimensional environment. They are trained with particular inputs and are ''queried'' periodically for the state of their learning about hidden regions or continued existence, when occluded, of the objects they ''see,'' and performance interpreted in light of the starting points, environment, and so forth, as mentioned previously.

### *5.1. Modeling association learning*

Mareschal and Johnson (2002) devised a connectionist model of unity perception based on three assumptions. First, infants can detect visual information relevant to object perception tasks prior to the effective utilization of this information. Second, experience viewing objects in motion is vital to perception of unity, because far objects move behind and emerge from near objects, providing support for the formation of associations between fully visible and partly occluded views. (In addition, young infants provide evidence of unity perception only when the visible parts undergo common motion; Kellman & Spelke, 1983.) Third, infants are equipped with short-term memory. None of these assumptions should reasonably be considered controversial or objectionable.

Given these assumptions, Mareschal and Johnson (2002) built a model using connectionist architecture: an input layer of inputs corresponding to a retina, hidden units whose computations formed representations of spatial completion, and an output layer that

provided a response in the form of a representation of a complete object, object parts, or an indeterminate response, after particular amounts of training in a specified environment. The architecture can be seen in Fig. 8, and an example of the training regimen can be seen in Fig. 9. Between the input and hidden units was a series of modules that processed the visual information in the training environment: the occluder, the motion patterns and orientation of rod parts, background texture, and points of intersection of the rod parts and occluder. The occluder remained stationary throughout each event, and the rod parts moved independently or in tandem. The model's task was to determine whether one or two objects (not including the occluder) were presented in the display. When the rod or rod parts were fully visible, the decision was accomplished directly, but it had to be inferred when there was an intersection of the rod and occluder. A memory trace of the previous portion of each event was stored and accumulated with increasing experience.

The model was set up to minimize the error between direct perception and the inference demanded by occlusion in response to training, and learned primarily by association: associations between views of partly hidden and fully visible ''objects,'' and how unity perception was best predicted by the visual cues present in each display. Because human infants are especially sensitive to motion patterns and orientation of the rod parts and rely on these cues to perceive unity (Johnson, 1997), we trained the models with events in which these cues were available or absent, in different combinations, to examine their contributions to unity perception alone or in tandem with other cues.

The models were able to learn unity readily, and their performance was strongly affected by the cues available, in particular a combination of cues to which infants, likewise, are sensitive and use to perceive unity in the lab setting: common motion, orientation, and relatability of the rod parts, and T-junctions (cues that specify the intersection of the rod parts and occluder). These models, therefore, demonstrate that object knowledge, at some level, can

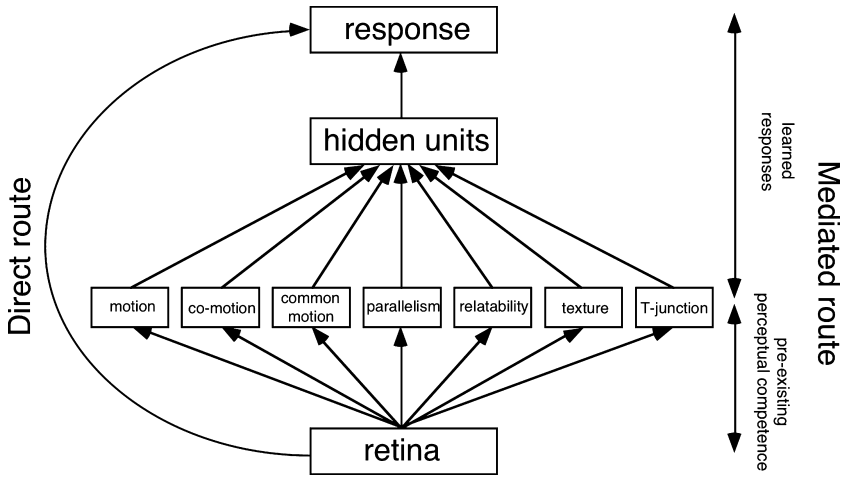


Fig. 8. Architecture of the Mareschal and Johnson (2002) model of perception of spatial completion. Adapted from Mareschal and Johnson (2002).

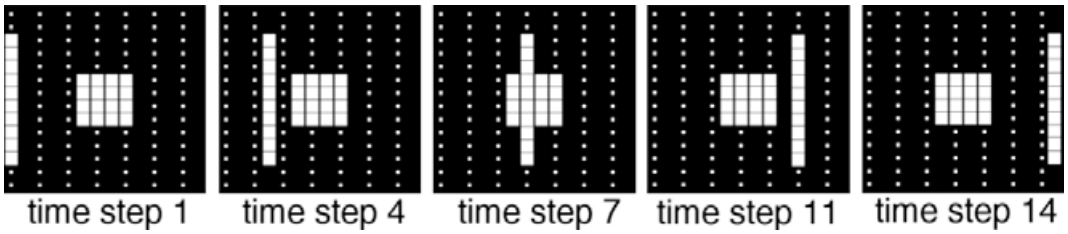


Fig. 9. Sample time steps illustrating the training regimen of the Mareschal and Johnson (2002) model.

be learned from proper experience in a structured environment, given appropriate lower-level perceptual skills—in this case, sensitivity to relevant visual information.

Models of fully hidden objects likewise have shown an important role for learning by association. Models described by Mareschal, Plunkett, and Harris (1999) were trained in a simple environment consisting of an occluder and a moving object that was small enough to be invisible for a number of time steps during an event in which it passed repeatedly back and forth along a horizontal trajectory. The models' task was to predict the location of this moving object in a future time step, and they learned to do this very quickly when it was fully visible. They were able to do so as well given repeated experience with a partly hidden trajectory—that is, a trajectory in which the object was briefly hidden. In other words, the models developed a representation of the moving object even in the absence of direct evidence for its existence—by virtue of a memory trace built from experience.

As noted previously, models are like experiments: A single modeling effort should not be taken to suggest that the specifics of the model's architecture or training provide any greater (or lesser) insights into human developmental processes than would a single set of experimental conditions tested in the lab. No doubt infants learn via association, but this is not all there is to how infants learn about the world at large. Other developmental phenomena are at work, and some of these phenomena have attracted the attention of modelers interested in exploring the origins of object knowledge. In the following section I will describe two recent models of visual development, each of which bears implications for infant object perception.

### *5.2. Modeling visual development*

The scope and contributions of the models I have described are limited, in part because the human visual system does not work or develop in the same way: Our retinas have a fovea and we move our eyes to points of interest in the scene, and visual development in human infants consists of formation and strengthening of neural circuits within, to, and from visual areas of the brain (Atkinson, 2000), as opposed to updating of weights within fixed connections between hard-wired, fully operational modules characteristic of many models (Rumelhart, McClelland, and the PDP Research Group, 1986), including ours. With these caveats in mind, Schlesinger, Amso, and Johnson (2007) created a computational model of infants' gaze patterns based on the idea of ''salience maps'' produced by visual modules

tuned to luminance, motion, color, and orientation in an input image (Itti & Koch, 2000). Salience was computed in part via a process of competition between visual features as the model received repeated exposures (or iterations) to the images, a strategy motivated by patterns of activity in the posterior parietal cortex that are suppressed in response to visual features that remain constant across exposure while increasing responses to features that change—thus highlighting their salience (Gottlieb, Kusunoki, & Goldberg, 1998). The model had a simulated fovea and the ability to direct ''gaze'' toward the most salient region in the image. We input an image of a moving rod-and-box stimulus to the model. After several iterations, the model quickly developed a salience map in which the rod segments were strongly activated, as activity for the edges of the occluder receded (Fig. 10). The model was intended to examine development of visual attention, not spatial completion per se, but given the success of this model and that of Mareschal and Johnson (2002), a model of ''learning from scanning'' seems feasible and likely to achieve important insights into human development.

The Schlesinger et al. (2007) model was designed to examine three kinds of cortical visual development in human infants, and their effects of these developments on scanning behavior and unity perception. The first was neural ''noise,'' uncorrelated activity among neurons within and across networks of cortical cells, which is characteristic of young infants' cortical function (Skoczenski & Norcia, 1998), and might make pattern detection initially difficult across disparate regions of the visual scene. The second was *horizontal* connections in visual area V1, which serve to strengthen responses of neighboring cells that code for a common edge in the visual scene (Burkhalter, Bernardo, & Charles, 1993), and whose development is likely to facilitate perception of edge connectedness across a gap. The third was *recurrent processing* of visual information, akin to repetition of the input within working memory and accomplished in primate parietal cortex (Gottlieb et al., 1998), which is analogous to modulating the time spent covertly ''comparing'' two or more targets.

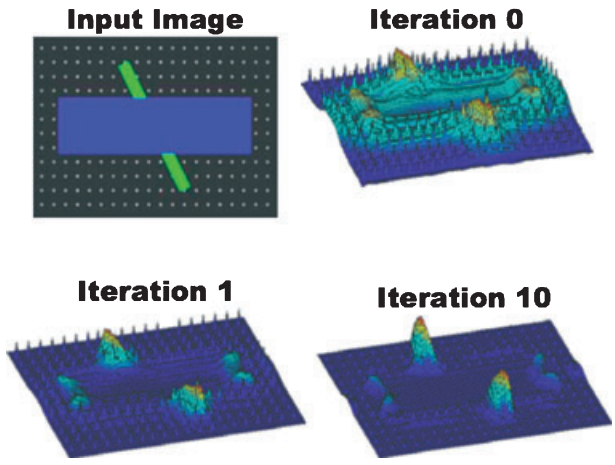


Fig. 10. Salience map yielded by the model of visual development produced by Schlesinger et al. (2007). Adapted from Schlesinger et al. (2007).

The parameters representing neural noise had little effect on performance, but the interaction of developments in horizontal connections and recurrent processing had a substantial effect on unity perception, with an ''ideal'' value for the horizontal connections set to a fairly low number, and the addition of recurrent loops beneficial to the model's ability to detect the unity of rod parts separated by the occluder. In other words, our model shows that growth of horizontal connections is neither necessary nor sufficient, whereas an increase in the duration of recurrent parietal activation is necessary and *almost* sufficient (i.e., it works for many values of horizontal connections).

### *5.3. Modeling vision and reaching*

In humans, eyes are situated in a head, which is situated on a body, which can move through space and to which are attached arms and hands that serve to act on objects. Part of the exploration process involves moving to and around objects, bringing them closer and rotating them to produce new points of view for further visual inspection. Models of motor development have demonstrated that neural networks can learn to coordinate vision and reaching during bouts of exploration of the environment. Kuperstein (1988) described a model in which representations of posture and arm position emerged from correlations between self-produced movement and sensory information about external target positions in space. The model was endowed with two retinas located in a head on a trunk, and a hand on an arm, all of which were free to move in 3D space. The model was allowed to grasp an object as information about its position was input from both the motor system and the visual system. The developmental process was similar to the Piagetian notion of *circular reactions* (Piaget, 1954), in which a developing system's behaviors are gradually honed after initial, sometimes lengthy, bouts of variability. The periods of variability were presumably used by the neural network to work out the coordination between sensorimotor feedback from different postures and positions of the limbs and the visual transformations they yielded. After correlations became more stable, the network learned to produce new, accurate patterns of reaching and grasping for objects that had not been encountered previously. In other words, the model had internalized the spatial mapping of limb position and visual coordinates—representations of the locations of the self and of external objects—from signals derived exclusively from sensory receptors and motor feedback, with no a priori knowledge of the objective ''features'' of objects.

Bullock, Grossberg, and Guenther (1993) introduced a model of eye-hand coordination with a similar goal: to examine emergence of correlations across sensory and motor systems without prespecified knowledge of how outputs from the two systems should be combined into a unitary representation. The phase of exploratory movements was termed *motor babbling* and constituted the principal learning period about the effects of movement on visual input. The goal of the model was to enact a reaching trajectory toward the object that was as direct as possible, on the basis of visual and sensory feedback. Following training, the model's reaches for objects of different sizes and shapes were geared toward transforming visual information about the target and the effector (the hand) so as to produce maximally efficient, goal-oriented movements.

These models demonstrate that the visual input from objects is determined by the action capabilities of the system and the affordances of objects in the context of those actions. They illustrate in addition that cognitive development is constrained by expanding motor control over actions that provide increasingly detailed information about objects and events in the world.

## 6. How infants learn about objects

In the previous two sections, I described the capacities of newborns to detect and remember key information about the visual world, information that is important for specifying objects: their segregation from one another and their relative distances, and the retention of this information for brief intervals sufficient to support recognition upon repeated encounters. I also described models of development object perception. These models demonstrated that a naïve system, given appropriate perceptual, cognitive, and motor skills and a suitable environment in which to learn, can perceive objects as complete and persistent despite occlusion, and can act on objects by detecting relevant information about their properties. Does the developmental process in human infants accord with these findings?

### *6.1. Infants learn about objects via association*

Consider first the possibility that infants learn about object occlusion via association. How might this work? The Mareschal and Johnson (2002) model learned to perceive partly hidden objects as complete in two ways: by associating objects with different visual cues in the input (i.e., texture, motion, junctions, and orientation), and by associating different views of objects with each other—a fully visible, complete rod that moved behind the occluder and then became partly occluded. The Mareschal et al. (1999) model was exposed to an object moving on a repetitive trajectory and quickly learned to predict its reappearance from behind an occluder. The model was set up to predict the location of the moving object based on its preceding position and trajectory, and the model maintained a memory trace of it when it was rendered invisible by the occluder. Is there evidence for similar processes in human infants?

To my knowledge, no one has tested the possibility that infants learn about objects by *associating* individual visual attributes with their coherence and persistence across occlusion, though the contributions of such visual attributes to perceptual completion have been investigated. Spatial completion has been observed in young infants (younger than 6 months) only when the rod parts are aligned, and moving in tandem behind the occluder (Johnson, 1997, 2004), in displays with the four visual cues examined in the Mareschal and Johnson (2002), but in the absence of one or more cues, spatial completion is disrupted (Kellman & Spelke, 1983). And spatiotemporal completion has been observed in young infants only when the trajectory is horizontal, not angled, and when the spatiotemporal gap imposed by the occluder is relatively short (Bremner et al., 2005, 2007; Johnson, Bremner, et al. 2003). But it is not clear that these studies provide evidence that association per se is an important mechanism of development in perceiving object occlusion.

Such evidence comes from experiments by Johnson, Amso, and Slemmer (2003), who examined 4- and 6-month-old infants' responses to object trajectory displays by recording predictive eye movements. We reasoned that a representation of a moving object would be revealed by a consistent pattern of fixations toward the far side of the occluder upon its occlusion. Infants were tested in one of four conditions. In the *baseline* condition, infants were shown the ball-box display depicted in Fig. 4 as eye movements were recorded with a corneal-reflection eye tracker. The display was presented for eight 30-s trials. In the *random* condition, infants viewed eight presentations of displays that were identical to the ball-box stimulus except the ball's point of reemergence after occlusion was randomized (left or right). In this case, anticipation offers no gain to the observer, who is just as likely to make perceptual contact with the ball if the point of gaze remains where the object moved out of view. (We hypothesized that anticipations in the random condition might be random eye movements themselves.) In the *training* condition, infants were first presented with four trials of the ball only, fully visible on its lateral trajectory (no occluder), followed by four trials with the ball-box display, as in the predictable condition. Finally, in the *generalization* condition, infants first viewed four trials with a vertical unoccluded trajectory, followed by four trials with a partly occluded horizontal trajectory.

In the baseline condition, 6-month-olds produced a significantly higher proportion of anticipatory eye movements than 4-month-olds, and a comparison of 4-month-olds' performance in the baseline versus random conditions revealed no reliable differences. This latter finding implies that any predictive eye movements we observed by 4-month-olds were actually not based on a mental representation of the occluded object and its motion, but instead were simply random eye movements scattered about the display that, by chance, happened to fit the criteria for categorization as predictive. Moreover, 4-month-olds' performance in the baseline condition did not improve across trials (as would be expected if the infants learned the repetitive sequence). In fact, there was a significant *decline* in anticipations across trials. These results indicate that eye movement patterns may have been driven more in the older age group by a veridical representation of the object on its path behind the occluder.

However, 4-month-olds in the training condition showed reliably more predictive eye movements relative to 4-month-olds in the baseline condition. Comparisons of the two 6-month-old groups, in contrast, revealed no significant differences. The boost in anticipation performance seen in the 4-month-old training group generalized from exposure to the vertical trajectory orientation, implying that infants in the training condition were not simply trained for facilitation of horizontal eye movements, but instead true representation-based anticipations.

How long does this effect of training last? Johnson and Shuwairi (2009) addressed this question with a replication of the Johnson, Amso, et al. (2003) experiment: Baseline and training conditions with 4-month-olds yielded similar results as the previous study. We extended these findings with three additional conditions: a *delay* condition, in which a 30-min wait was imposed between training (with an unoccluded trajectory) and test (with a partly occluded trajectory), and a *reminder* condition, identical to the delay condition except for the addition of a single additional training trial immediately before test. Performance in the delay condition was not significantly different from that of baseline, implying that the

gains produced by brief training did not survive the 30-min interruption prior to test. However, eye movement anticipations were facilitated by the reminder condition to the same extent as in the (immediate) training condition. (A fifth condition, *brief training*, consisted of a single training trial prior to immediate test, and this did not have any discernible effect on performance.)

Taken together, these findings suggest that there are consequential changes around 4 months after birth in representations of moving, occluded objects (Johnson, Bremner, et al. 2003). Such representations are sufficiently strong by 6 months to guide anticipatory looking behaviors consistently when viewing predictable moving object event sequences. Four-month-olds' anticipations under these conditions provided little evidence of veridical object representations. However, a short exposure to an unoccluded object trajectory induces markedly superior performance in our tracking task in this age group, and with a reminder, this training effect can last for a period of time outside the scope of short-term memory. These findings also help clarify the role of associative learning in object perception development. Infants did not seem to learn by viewing repetitive events that are perfectly predictable to adults (otherwise infants in the baseline conditions would have begun to show increased levels of anticipation after several trails viewing the occluded trajectory). Instead, infants learned by associating views of the fully visible object trajectory and the partly occluded object trajectory.

### *6.2. Infants learn about objects via ''active assembly''*

The Schlesinger et al. (2007) model discussed in the previous section highlighted two aspects of visual development that might have a key role in development of spatial completion: growth of horizontal connections among neurons and circuits in cortical visual area V1, and recurrent processing, which, we reasoned, served to compare aspects of the visual scene. How might these influence developing object perception skills?

Burkhalter and colleagues (Burkhalter, 1991; Burkhalter et al., 1993) have reported evidence for developments in horizontal connections in V1 from deceased fetuses and infants, across the period from 26 weeks postconception to 7 months after birth, but the precise role of these developments in object perception has not been documented. However, there are findings from experiments on spatial completion that bear on this question. Two-month-old infants have been found to perceive spatial completion in displays with a relatively narrow occluder, such that the rod parts are close together across the gap imposed by occlusion, but not when the occluder is wide (Johnson, 2004). (Similar findings were obtained in studies of spatiotemporal completion—reducing gap size facilitates perception of completion here as well.) Older infants are less susceptible to effects of widening this gap (Johnson, 1997). In addition, 4-month-old infants are more likely to look back and forth across a wide gap at the rod parts than are 2-month-olds (Johnson & Johnson, 2000). These results are to be expected if the visual system becomes better able to link aligned edges across a gap as connections between receptive fields become strengthened.

Other experiments examined the possibility that spatial completion develops from a constructive process—which I have termed *active assembly*—serving to integrate parts of

the visual scene into a coherent whole, in like fashion to recurrent processing discussed previously. Amso and Johnson (2006) and Johnson, Slemmer, and Amso (2004) observed 3-month-old infants in a spatial completion task and recorded infants' eye movements with a corneal reflection eye tracker during the habituation phase of the experiment. We found systematic differences in oculomotor scanning patterns between infants whose posthabituation test display preferences indicated unity perception and infants who provided evidence of perception of disjoint surfaces: ''Perceivers'' tended to scan more in the vicinity of the two visible rod segments, and to scan back and forth between them (Fig. 11). In a younger sample, Johnson et al. (2008) found a correlation between posthabituation preference—our index of spatial completion—and targeted visual exploration, operationalized as the proportion of eye movements directed toward the moving rod parts, which we reasoned was the most relevant aspect of the stimulus for perception of completion. (Precise localization of the point of gaze can be a challenge for these very young infants, attested by the fact that targeted scans almost always followed the rod as it moved, rarely anticipating its position.)

A relation between targeted visual exploration and spatial completion does not by itself pinpoint a causal role. Such evidence would come from experiments in which individual differences in oculomotor patterns were observed in both spatial completion and some other visual task, and this was recently reported by Amso and Johnson (2006). We found that both spatial completion and scanning patterns were strongly related to performance in an independent visual search task in which targets, defined by a unique feature (either motion or orientation) were placed among a large set of distracters. There were substantial individual differences in successful search, both in terms of detecting the target and the latency to do so, and these differences mapped clearly onto the likelihood of spatial completion. This finding is inconsistent with the possibility that scanning patterns were tailored specifically to perceptual completion, and instead suggests that a general facility with targeted visual behavior leads to improvements across multiple tasks.

Targeted visual exploration may make a vital contribution to the emergence of veridical object perception. As scanning patterns develop, they support binding of disparate visual features into unified percepts—active assembly of coherent objects from surface fragments, confirming the outcome of the Schlesinger et al. (2007) model of visual development. With the emergence of selective attention and other perception-action systems, infants become

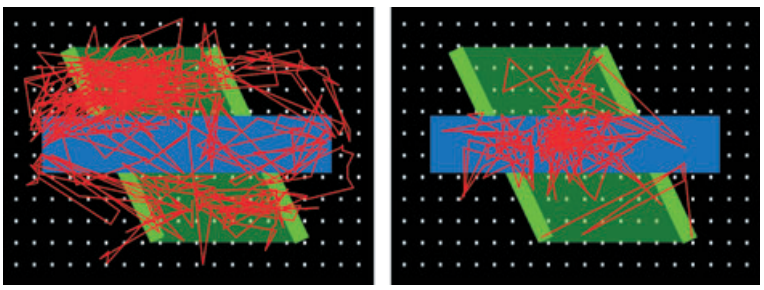

Fig. 11. Examples of individual differences in oculomotor scanning as 3-month-old infants view a rod-and-box display (Johnson et al., 2004).

increasingly active participants in their own perceptual development rather than passive recipients of information. Active engagement of the infant's visual attention is consistent with a key tenet of Piagetian theory—the central role of the child's own behavior in cognitive development—and with a constructivist view—the building of structure from constituent pieces. The following section describes another of these perception-action systems, visual-manual exploration, and its role in constructing volumetric objects.

### *6.3. Infants learn about objects via visual-manual exploration*

The Kuperstein (1988) and Bullock et al. (1993) models demonstrated that when perception and action develop in tandem their coordination can be an emergent property, with each influencing the other to the benefit of the exploratory capacity of the organism. Developments in perception and action have long been of interest to developmental psychologists, and there has been recent evidence to show that 3D object completion emerges as a consequence of improvements in infants' motor skills. Two types of motor skill, both of which develop rapidly at the same time that 3D object completion seems to emerge—4 to 6 months—may play a particularly important role: self-sitting and coordinated visual-manual exploration. Independent sitting frees the hands for play and promotes gaze stabilization during manual actions (Rochat & Goubet, 1995), and, therefore, self-sitting might encourage coordination of object manipulation with visual inspection as infants begin to play with objects, providing the infants with multiple views. In addition, manipulation of objects—touching, squeezing, mouthing—may promote learning about object form from tactile information.

To examine these possibilities, Soska, Adolph, and Johnson (2010) observed infants between 4.5 months and 7.5 months in a replication of the Soska and Johnson (2008) habituation experiment with the rotating wedge stimuli (Fig. 6). In the same testing session we assessed infants' manual exploration skills by observing spontaneous object manipulation in a controlled setting and obtained parental reports of the duration of infants' sitting experience. We reasoned that infants who had more self-sitting experience would in turn show a greater tendency to explore objects from multiple viewpoints and therefore have more opportunities to learn about objects' 3D forms outside the lab. Thus, within this age range, individual differences in self-sitting experience and coordinated visual-manual exploration were predicted to be related to individual differences in infants' looking preferences to the complete and incomplete object displays, our index of 3D object completion.

Our predictions were supported. We found strong and significant relations between both self-sitting and visual-manual coordination, from parents' reports and the motor skills assessment, and 3D object completion performance, assessed with the habituation paradigm. We recorded a number of other motor skills to explore how widespread the relations were within the perception-action systems under investigation, such as grasping, holding, and manipulation without visual inspection, and none were related to 3D object completion.

Self-sitting experience and coordinated visual-manual exploration were the strongest predictors of performance on the visual habituation task. The results of a regression analysis

yielded evidence that the role of self-sitting was indirect, influencing 3D completion chiefly in its support of infants' visual-manual exploration. Self-sitting infants performed more manual exploration while looking at objects than did nonsitters, and visual-manual object exploration is precisely the skill that provides active experience viewing objects from multiple viewpoints, thereby facilitating perceptual completion of 3D form. These results provide evidence for a cascade of developmental events following from the advent of visual-motor coordination, including learning from self-produced experiences.

## 7. Concluding remarks

In principle, perceptual completion and other object perception skills available early in postnatal life might develop solely from ''passive'' perceptual experience, because natural scenes are richly structured and characterized by a high degree of redundancy (Graham & Field, 2007) and infants gain a great deal of exposure to the visual world—on the order of several hundred hours—by 2 months after birth (Johnson, Amso, et al. 2003). Thus—in principle—infants might learn about objects by observing the world and acquiring associations between views of objects when fully visible and partly or fully occluded. But the findings yielded by the Amso and Johnson (2006) and Soska, Adolph, and Johnson (2010) experiments indicate that passive experience may be insufficient to learn about the full range of occlusion phenomena, and, together with modeling accounts of object perception development, broaden our conceptions of how infants learn about the visual world. Active assembly and visual-manual exploration provide information to the infant about her own control of an event while simultaneously generating multimodal information to inform developing object perception skills. For complex kinds of perceptual completion, such as 3D object completion, the coordination of posture, reaching, grasping, and visual inspection seems to be critical: Only the visual-manual skills involved in generating changes in object viewpoint—rotating, fingering, and transferring while looking—were related to 3D object completion.

Careful consideration of the evidence I have described reveals that no one account, such as Piagetian or nativist theories, encompasses the full range of changes that underlie the emergence of object concepts in infancy. Significant advances, nevertheless, have been achieved. The rudiments of object knowledge are evident in the first 6 months after birth, revealed by detailed observations of information-gathering processes available to infants, from which more complex representations of objects are constructed. But, notably, there is no pure case of development caused in the absence of either intrinsic or external influences (Elman et al., 1996; Quartz & Sejnowski, 1997). The question is what mechanisms are responsible for perceptual and cognitive development.

## Acknowledgments

Preparation of this article was supported by NIH grants R01-HD40432 and R01-HD48733.

## References


Aguiar, A., & Baillargeon, R. (1999). 2.5-month-old infants' reasoning about when objects should and should not be occluded. *Cognitive Psychology*, *39*, 116–157.

Albert, M. V., Schnabel, A., & Field, D. J. (2008). Innate visual learning through spontaneous activity patterns. *PLoS Computational Biology*, *4*, 1–8.

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *6*, 1236–1245.

Atkinson, J. (2000). *The developing visual brain*. New York: Oxford University Press.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science*, *3*, 133–140.

Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, *3*, 2–13.

Bednar, J. A., & Miikkulainen, R. (2007). Constructing visual function through prenatal and postnatal learning. In D. Mareschal, S. Sirois, G. Westermann, & M. H. Johnson (Eds.), *Neuroconstructivism: Perspectives and prospects*, Vol. 2. (pp. 13–37). New York: Oxford University Press.

Berger, A., Tzur, G., & Posner, M. L. (2006). Infant brains detect arithmetic errors. *Proceedings of the National Academy of Sciences (USA)*, *103*, 12649–12653.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Cheshire, A., & Spring, J. (2007). Conditions for young infants' failure to perceive trajectory continuity. *Developmental Science*, *10*, 613–624.

Bremner, J. G., Johnson, S. P., Slater, A. M., Mason, U., Foster, K., Cheshire, A., & Spring, J. (2005). Conditions for young infants' perception of object trajectories. *Child Development*, *74*, 1029–1043.

Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, *5*, 408–435.

Burkhalter, A. (1991). Developmental status of intrinsic connections in visual cortex of newborn human. In P. Bagnoli & W. Hodos (Eds.), *The changing visual system* (pp. 247–254). New York: Plenum Press.

Burkhalter, A., Bernardo, K. L., & Charles, V. (1993). Development of local circuits in human visual cortex. *Journal of Neuroscience*, *13*, 1916–1931.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Clifton, R. K., Rochat, P., Litovsky, R. Y., & Perris, E. E. (1991). Object representation guides infants' reaching in the dark. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 323–329.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton-Mifflin.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*, 279–281.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development*, *58*, 601–622.

Goldman-Rakic, P. S. (1996). The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London B*, *351*, 1445–1453.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*, 481–484.

Graham, D. J., & Field, D. J. (2007). Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, *21*, 149–164.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior & Development*, *21*, 167–179.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*, 1489–1506.

Johnson, S. P. (1997). Young infants' perception of object unity: Implications for development of attentional and cognitive skills. *Current Directions in Psychological Science*, *6*, 5–11.

Johnson, S. P. (2004). Development of perceptual completion in infancy. *Psychological Science*, *15*, 769–775.

Johnson, S. P. (Ed.) (2010). *Neoconstructivism: The new science of cognitive development*. New York: Oxford University Press.

Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. *Proceedings of the National Academy of Sciences (USA)*, *100*, 10568–10573.

Johnson, S. P., & Aslin, R. N. (1995). Perception of object unity in 2-month-old infants. *Developmental Psychology*, *31*, 739–745.

Johnson, S. P., & Aslin, R. N. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development*, *11*, 161–180.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' perception of object trajectories. *Child Development*, *74*, 94–108.

Johnson, S. P., Davidow, J., Hall-Haro, C., & Frank, M. C. (2008). Development of perceptual completion originates in information acquisition. *Developmental Psychology*, *44*, 1214–1224.

Johnson, S. P., & Johnson, K. L. (2000). Early perception-action coupling: Eye movements and the development of object perception. *Infant Behavior & Development*, *23*, 461–483.

Johnson, S. P., & Náñez, J. E. (1995). Young infants' perception of object unity in two-dimensional displays. *Infant Behavior & Development*, *18*, 133–143.

Johnson, S. P., & Shuwairi, S. M. (2009). Learning and memory facilitate predictive tracking in 4-month-olds. *Journal of Experimental Child Psychology*, *102*, 122–130.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy*, *6*, 185–201.

Kagan, J. (2008). In defense of qualitative changes in development. *Child Development*, *79*, 1606–1624.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.

Kessen, W., Salapatek, P., & Haith, M. (1972). The visual response of the human newborn to linear contour. *Journal of Experimental Child Psychology*, *13*, 9–20.

Kuperstein, M. (1988). Neural model of adaptive hand-eye coordination for single postures. *Science*, *239*, 1308–1311.

Mareschal, D., & Bremner, A. J. (2009). Modeling the origins of object knowledge. In B. Hood & L. Santos (Eds.), *The origins of object knowledge* (pp. 227–262). Oxford, England: Oxford University Press.

Mareschal, D., & Johnson, S. P. (2002). Learning to perceive object unity: A connectionist account. *Developmental Science*, *5*, 151–185.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviors in infancy. *Developmental Science*, *2*, 306–317.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science*, *19*, 1063–1066.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Quartz, S. R., & Sejnowski, T. J. (1997). The neural basis of cognitive development: An constructivist manifesto. *Behavioral and Brain Sciences*, *20*, 537–556.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*, 1067–1070.

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, *57*, 971–976.

Rochat, P., & Goubet, N. (1995). Development of sitting and reaching in 5- to 6-month-old infants. *Infant Behavior & Development*, *18*, 53–68.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Schlesinger, M., Amso, D., & Johnson, S. P. (2007). The neural basis for visual selective attention in young infants: A computational account. *Adaptive Behavior*, *15*, 135–148.

Sekuler, A. B., Lee, J. A. J., & Shettleworth, S. J. (1996). Pigeons do not complete partly occluded figures. *Perception*, *25*, 1109–1120.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*, 1779–1782.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701–703.

Skoczenski, A. M., & Norcia, A. M. (1998). Neural noise limitations on infant visual sensitivity. *Nature*, *391*, 697–700.

Slater, A. (1995). Visual perception and memory at birth. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research*, Vol. 9. (pp. 107–162). Norwood, NJ: Ablex.

Slater, A., Johnson, S. P., Brown, E., & Badenoch, M. (1996). Newborn infants' perception of partly occluded objects. *Infant Behavior & Development*, *19*, 145–148.

Slater, A., Johnson, S. P., Kellman, P. J., & Spelke, E. S. (1994). The role of three-dimensional depth cues in infants' perception of partly occluded objects. *Early Development and Parenting*, *3*, 187–191.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, *13*, 33–49.

Slater, A., Morison, V., Town, C., & Rose, D. (1985). Movement perception and identity constancy in the new-born baby. *British Journal of Developmental Psychology*, *3*, 211–220.

Slater, A., Quinn, P. C., Kelly, D. J., Lee, K., Longmore, C. A., McDonald, P. R., & Pascalis, O. (in press). The shaping of the face space in early infancy: Becoming a native face processor. *Child Development Perspectives*.

Soska, K. C., Adolph, K. A., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates 3D object completion. *Developmental Psychology*, *46*, 129–138.

Soska, K. C., & Johnson, S. P. (2008). Development of 3D object completion in infancy. *Child Development*, *79*, 1230–1236.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*, 605–632.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

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*, 79–87.

Sperry, R. W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and their connections. *Proceedings of the National Academy of Sciences (USA)*, *50*, 703–710.

Valenza, E., Leo, I., Gava, L., & Simion, F. (2006). Perceptual completion in newborn human infants. *Child Development*, *77*, 1810–1821.

Valenza, E., Simion, F., Macchi Cassia, V., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 892–903.

Wong, R. O. L. (1999). Retinal waves and visual system development. *Annual Review of Neuroscience*, *22*, 29–47.

Zeki, S. (1993). *A vision of the brain*. Cambridge, MA: Blackwell.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1185–1243

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01128.x

# Learning to Learn Causal Models

Charles Kemp,[a] Noah D. Goodman,[b] Joshua B. Tenenbaum[b]

[a]*Department of Psychology, Carnegie Mellon University*
[b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

Received 6 November 2008; received in revised form 11 June 2010; accepted 14 June 2010

**Abstract**

Learning to understand a single causal system can be an achievement, but humans must learn about multiple causal systems over the course of a lifetime. We present a hierarchical Bayesian framework that helps to explain how learning about several causal systems can accelerate learning about systems that are subsequently encountered. Given experience with a set of objects, our framework learns a causal model for each object and a *causal schema* that captures commonalities among these causal models. The schema organizes the objects into categories and specifies the causal powers and characteristic features of these categories and the characteristic causal interactions between categories. A schema of this kind allows causal models for subsequent objects to be rapidly learned, and we explore this accelerated learning in four experiments. Our results confirm that humans learn rapidly about the causal powers of novel objects, and we show that our framework accounts better for our data than alternative models of causal learning.

*Keywords:* Causal learning; Learning to learn; Learning inductive constraints; Transfer learning; Categorization; Hierarchical Bayesian models

## 1. Learning to learn causal models

Children face a seemingly endless stream of inductive learning tasks over the course of their cognitive development. By the age of 18, the average child will have learned the meanings of 60,000 words, the three-dimensional shapes of thousands of objects, the standards of behavior that are appropriate for a multitude of social settings, and the causal structures underlying numerous physical, biological, and psychological systems. Achievements like

Correspondence should be send to Charles Kemp, Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Baker Hall 340T, Pittsburgh, PA 15213. E-mail: ckemp@cmu.edu

these are made possible by the fact that inductive tasks fall naturally into families of related problems. Children who have faced several inference problems from the same family may discover not only the solution to each individual problem but also something more general that facilitates rapid inferences about subsequent problems from the same family. For example, a child may require extensive time and exposure to learn her first few names for objects, but learning a few dozen object names may allow her to learn subsequent names much more quickly (Bloom, 2000; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002).

Psychologists and machine learning researchers have both studied settings where learners face multiple inductive problems from the same family, and they have noted that learning can be accelerated by discovering and exploiting common elements across problems. We will refer to this ability as ''learning to learn'' (Harlow, 1949; Yerkes, 1943), although it is also addressed by studies that focus on ''transfer learning,'' ''multitask learning,'' ''lifelong learning,'' and ''learning sets'' (Caruana, 1997; Stevenson, 1972; Thorndike & Woodworth, 1901; Thrun, 1998; Thrun & Pratt, 1998). This paper provides a computational account of learning to learn that focuses on the acquisition and use of inductive constraints. After experiencing several learning problems from a given family, a learner may be able to induce a *schema*, or a set of constraints that captures the structure of all problems in the family. These constraints may then allow the learner to solve subsequent problems given just a handful of relevant observations.

The problem of learning to learn is relevant to many areas of cognition, including word learning, visual learning, and social learning, but we focus here on causal learning and explore how people learn and use inductive constraints that apply to multiple causal systems. A door, for example, is a simple causal system, and experience with several doors may allow a child to rapidly construct causal models for new doors that she encounters. A computer program is a more complicated causal system, and experience with several pieces of software may allow a user to quickly construct causal models for new programs that she encounters. Here we consider settings where a learner is exposed to a family of objects and learns causal models that capture the causal powers of these objects. For example, a learner may implicitly track the effects of eating different foods and may construct a causal model for each food that indicates whether it tends to produce indigestion, allergic reactions, or other kinds of problems. After experience with several foods, a learner may develop a schema (Kelley, 1972) that organizes these foods into categories (e.g., citrus fruits) and specifies the causal powers and characteristic features of each category (e.g., citrus fruits cause indigestion and have crescent-shaped segments). A schema of this kind should allow a learner to rapidly infer the causal powers of novel objects: for example, observing that a novel fruit has crescent-shaped segments might be enough to conclude that it causes indigestion.

There are three primary reasons why causal reasoning provides a natural setting for exploring how people learn and use inductive constraints. First, abstract inductive constraints play a crucial role in causal learning. Some approaches to causal learning focus on bottom-up statistical methods, including methods that track patterns of conditional independence or partial correlations (Glymour, 2001; Pearl, 2000). These approaches, however, offer at best a limited account of human learning. Settings where humans observe correlational

data without the benefit of strong background knowledge often lead to weak learning even when large amounts of training data are provided (Lagnado & Sloman, 2004; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). In contrast, both adults and children can infer causal connections from observing just one or a few events of the right type (Gopnik & Sobel, 2000; Schulz & Gopnik, 2004)—far fewer observations than would be required to compute reliable measures of correlation or independence. Top-down, knowledge-based accounts provide the most compelling accounts of this mode of causal learning (Griffiths & Tenenbaum, 2007).

Second, some causal constraints are almost certainly learned, and constraint learning probably plays a more prominent role in causal reasoning than in other areas of cognition, such as language and vision. Fundamental aspects of language and vision do not change much from one generation to another, let alone over the course of an individual's life. It is therefore possible that the core inductive constraints guiding learning in language and vision are part of the innate cognitive machinery rather than being themselves learned (Bloom, 2000; Spelke, 1994). In contrast, cultural innovation never ceases to present us with new families of causal systems, and the acquisition of abstract causal knowledge continues over the life span. Consider, for example, a 40-year-old who is learning to use a cellular phone for the first time. It may take him a while to master the first phone that he owns, but by the end of this process—and certainly after experience with several different cell phones—he is likely to have acquired abstract knowledge that will allow him to adapt to subsequent phones rapidly and with ease.

The third reason for our focus on causal learning is methodological, and it derives from the fact that learning to learn in a causal setting can be studied in adults and children alike. Even if we are ultimately interested in the origins of abstract knowledge in childhood, studying analogous learning phenomena in adults may provide the greatest leverage for developing computational models, at least at the start of the enterprise. Adult participants in behavioral experiments can provide rich quantitative judgments that can be compared with model predictions in ways that are not possible with standard developmental methods. The empirical section of this paper therefore focuses on adult experiments. We discuss the developmental implications of our approach in some detail, but a full evaluation of our approach as a developmental model is left for future work.

To explain how abstract causal knowledge can both constrain learning of specific causal relations and can itself be learned from data, we work within a hierarchical Bayesian framework (Kemp, 2008; Tenenbaum, Griffiths, & Kemp, 2006). Hierarchical Bayesian models include representations at several levels of abstraction, where the representation at each level captures knowledge that supports learning at the next level down (Griffiths & Tenenbaum, 2007; Kemp, Perfors, & Tenenbaum, 2007; Kemp & Tenenbaum, 2008). Statistical inference over these hierarchies helps to explain how the representations at each level are learned. Our model can be summarized as a three-level framework where the top level specifies a causal schema, the middle level specifies causal models for individual objects, and the bottom level specifies observable data. If the schema at the top level is securely established, then the framework helps to explain how abstract causal knowledge supports the construction of causal models for novel objects. If the schema at the upper level

is not yet established, then the framework helps to explain how causal models can be learned primarily from observable data. Note, however, that top-down learning and bottom-up learning are just two of the possibilities that emerge from our hierarchical approach. In the most general case, a learner will be uncertain about the information at all three levels, and will have to simultaneously learn a schema (inference at the top level) and a set of causal models (inference at the middle level) and make predictions about future observations (inference at the bottom level).

Several aspects of our approach draw on previous psychological research. Cognitive psychologists have discussed how abstract causal knowledge (Lien & Cheng, 2000; Shanks & Darby, 1998) might be acquired, and they have studied the bidirectional relationship between categorization and causal reasoning (Lien & Cheng, 2000; Waldmann & Hagmayer, 2006). Previous models of categorization have used Bayesian methods to explain how people organize objects into categories based on their features (Anderson, 1991) or their relationships with other objects (Kemp, Tenenbaum, Griffiths, Yamada, & Ueda, 2006), although not in a causal context. In parallel, Bayesian models of knowledge-based causal learning have often assumed a representation in terms of object categories, but they have not attempted to learn these categories (Griffiths & Tenenbaum, 2007). Here we bring together all of these ideas and explore how causal learning unfolds simultaneously across multiple levels of abstraction. In particular, we show how learners can simultaneously make inferences about causal categories, causal relationships, causal events, and perceptual features.

## 2. Learning causal schemata

Later sections will describe our framework in full detail, but this section provides an informal introduction to our general approach. As a running example we consider the problem of learning about drugs and their side-effects: for instance, learning whether blood-pressure medications cause headaches. This problem requires inferences about two *domains*—people and drugs—and can be formulated as a *domain-level problem*:

$$\texttt{ingests}(\texttt{person}, \texttt{drug}) \xrightarrow{?} \texttt{headache}(\texttt{person}) \tag{1}$$

The domain-level problem in Eqn. 1 sets up an *object-level* problem for each combination of a person and a drug. For example,

$$\texttt{ingests}(\texttt{Alice}, \texttt{Doxazosin}) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{2}$$

represents the problem of deciding whether there is a causal relationship between Alice taking Doxazosin and Alice developing a headache, and

$$\texttt{ingests}(\texttt{Bob}, \texttt{Prazosin}) \xrightarrow{?} \texttt{headache}(\texttt{Bob}) \tag{3}$$

represents a second problem concerning the effect of Prazosin on Bob. Our goal is to learn an *object-level causal model* for each object-level problem. In Fig. 1A there are six people

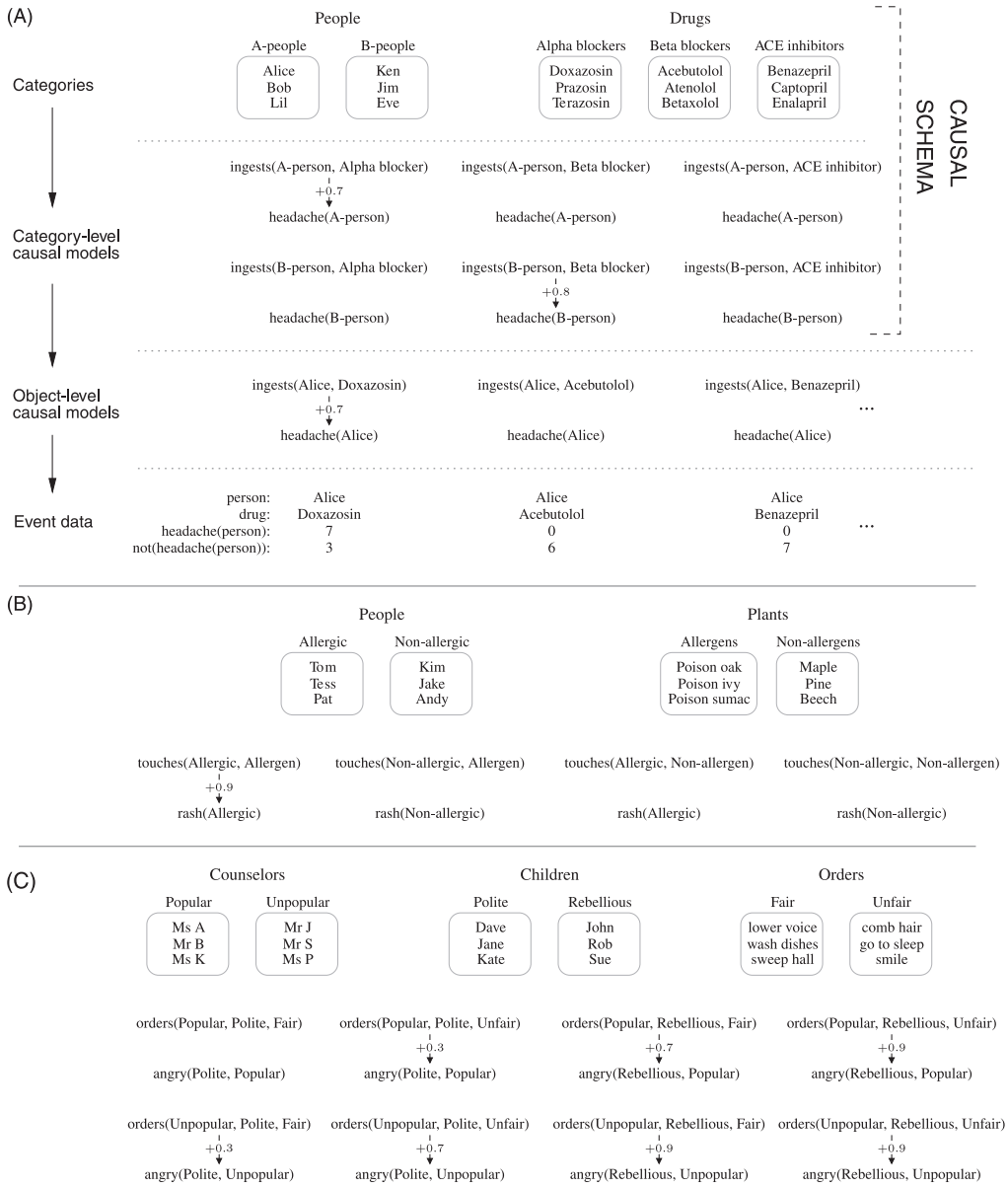


Fig. 1. Three settings where causal schemata can be learned. (A) The drugs and headaches example. The people are organized into two categories and the drugs are organized into three categories. The category-level causal models indicate that alpha blockers cause headaches in A-people and beta blockers cause headaches in B-people. There are 54 object-level causal models in total, one for each combination of a person and a drug, and three of these models are shown. The first indicates that Doxazosin often gives Alice headaches. The event data for learning these causal models are shown at the bottom level: Alice has taken Doxazosin 10 times and experienced a headache on seven of these occasions. (B) The allergy example. The schema organizes plants and people into two categories each, and the object-level models and event data are not shown. (C) The summer camp example. The schema organizes the counselors, the children, and the orders into two categories each.

and nine drugs, which leads to 54 object-level problems and 54 object-level models in total. Fig. 1A shows three of these object-level models, where the first example indicates that ingesting Doxazosin tends to cause Alice to develop headaches. The observations that allow these object-level models to be learned will be called event data or contingency data, and they are shown at the bottom level of Fig. 1A. The first column of event data indicates, for example, that Alice has taken Doxazosin 10 times and has experienced headaches on seven of these occasions.

The 54 object-level models form a natural family, and learning several models from this family should support inferences about subsequent members of the family. For example, learning how Doxazosin affects Alice may help us to rapidly learn how Doxazosin affects Bob, and learning how Alice responds to Doxazosin may help us to rapidly learn how Alice responds to Prazosin. This paper will explore how people learn to learn object-level causal models. In other words, we will explore how learning several of these models can allow subsequent models in the same family to be rapidly learned.

The need to capture relationships between object-level problems like Eqns. 2 and 3 motivates the notion of a causal schema. Each possible schema organizes the people and the drugs into categories and specifies causal relationships between these categories. For example, the schema in Fig. 1A organizes the six people into two categories (A-people and B-people) and the nine drugs into three categories (alpha blockers, beta blockers, and ACE inhibitors). The schema also includes *category-level* causal models that specify relationships between these categories. Because there are two categories of people and three categories of drugs, six category-level models must be specified in total, one for each combination of a person category and drug category. For example, the category-level models in Fig. 1A indicate that alpha blockers tend to produce headaches in A-people, beta blockers tend to produce headaches in B-people, and ACE inhibitors rarely produce headaches in either group. Note that the schema supports inferences about the object-level models in Fig. 1A. For example, because Alice is an A-person and Doxazosin is an alpha-blocker, the schema predicts that ingesting Doxazosin will cause Alice to experience headaches.

To explore how causal schemata are learned and used to guide inferences about object-level models, we work within a hierarchical Bayesian framework. The diagram in Fig. 1A can be transformed into a hierarchical Bayesian model by specifying how the information at each level is generated given the information at the level immediately above. We must therefore specify how the event data are generated given the object-level models, how the object-level models are generated given the category-level models, and how the category-level models are generated given a set of categories.

Although our framework is formalized as a top-down generative process, we will use Bayesian inference to invert this process and carry out bottom-up inference. In particular, we will focus on problems where event data are observed at the bottom level and the learner must simultaneously learn the object-level causal models, the category-level causal models and the categories that occupy the upper levels. After observing event data at the bottom level, our probabilistic model computes a posterior distribution over the representations at the upper levels, and our working assumption is that the categories and causal models learned by people are those assigned maximum posterior probability by our model. We do

not discuss psychological mechanisms that might allow humans to identify the representations with maximum posterior probability, but future work can explore how the computations required by our model can be implemented or approximated by psychologically plausible mechanisms.

Although it is natural to say that the categories and causal models are learned from the event data available at the bottom level of Fig. 1A, note that this achievement relies on several kinds of background knowledge. We assume that the learner already knows about the relevant domains (e.g., people and drugs) and events (e.g., ingestion and headache events) and is attempting to solve a problem that is well specified at the domain level (e.g., the problem of deciding whether ingesting drugs can cause headaches). We also assume that the existence of the hierarchy in Fig. 1A is known in advance. In other words, our framework knows from the start that it should search for *some* set of categories and *some* set of causal models at the category and object levels, and learning is a matter of finding the candidates that best account for the data. We return to the question of background knowledge in the General Discussion and consider the extent to which some of this knowledge might be the outcome of prior learning.

We have focused on the drugs and headaches scenario so far, but the same hierarchical approach should be relevant to many different settings. Suppose that we are interested in the relationship between touching a plant and subsequently developing a rash. In this case the domain-level problem can be formulated as

$$\texttt{touches}(\texttt{person}, \texttt{plant}) \xrightarrow{?} \texttt{rash}(\texttt{person})$$

We may notice that only certain plants produce rashes, and that only certain people are susceptible to rashes. A schema consistent with this idea is shown in Fig. 1B. There are two categories of plants (allergens and nonallergens), and two categories of people (allergic and nonallergic). Allergic people develop rashes after touching allergenic plants, including poison oak, poison ivy, and poison sumac. Allergic people, however, do not develop rashes after touching nonallergenic plants, and nonallergic people never develop rashes after touching plants.

As a third motivating example, suppose that we are interested in social relationships among the children and the counselors at a summer camp. In particular, we would like to predict whether a given child will become angry with a given counselor if that counselor gives her a certain kind of order. The domain-level problem for this setting is:

$$\texttt{orders}(\texttt{counselor}, \texttt{child}, \texttt{order}) \xrightarrow{?} \texttt{angry}(\texttt{child}, \texttt{counselor})$$

One possible schema for this setting is shown in Fig. 1C. There are two categories of counselors (popular and unpopular), two categories of children (polite and rebellious), and two categories of orders (fair and unfair). Rebellious children may become angry with a counselor if that counselor gives them any kind of order. Polite children accept fair orders

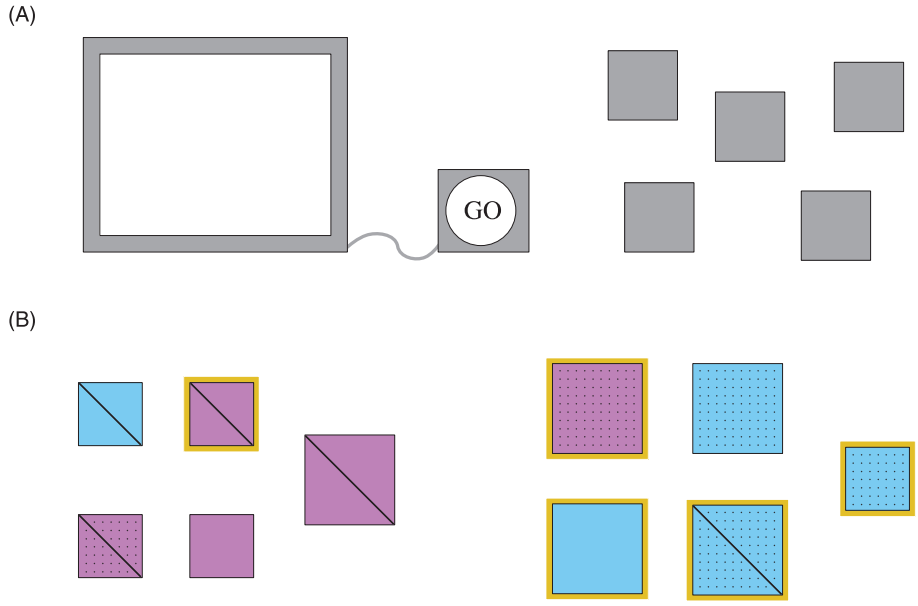


Fig. 2. Stimuli used in our experiments. (A) A machine and some blocks. The blocks can be placed inside the machine and the machine sometimes activates (flashes yellow) when the GO button is pressed. The blocks used for each condition of Experiments 1, 2, and 4 were perceptually indistinguishable. (B) Blocks used for Experiment 3. The blocks are grouped into two family resemblance categories: blocks on the right tend to be large, blue, and spotted, and tend to have a gold boundary but no diagonal stripe. These blocks are based on stimuli created by Sakamoto and Love (2004).

from popular counselors, but may become angry if a popular counselor gives them an unfair order or if an unpopular counselor gives them any kind of order.

Our experiments will make use of a fourth causal setting. Consider the blocks and the machine in Fig. 2A. The machine has a GO button, and it will sometimes activate and flash yellow when the button is pressed. Each block can be placed inside the machine, and whether the machine is likely to activate might depend on which block is inside. The domain-level problem for this setting is:

$$\texttt{inside}(\texttt{block},\texttt{machine})\ \texttt{\&}\ \texttt{button\_pressed}(\texttt{machine}) \xrightarrow{?} \texttt{activate}(\texttt{machine})$$

Note that the event on the left-hand side is a compound event which combines a state (a block is inside the machine) and an action (the button is pressed). In general, both the left- and right-hand sides of a domain-level problem may specify compound events that are expressed using multiple predicates.

One schema for this problem might organize the blocks into two categories: *active* blocks tend to activate the machine on most trials, and *inert* blocks seem to have no effect on the machine. Note that the blocks and machine example is somewhat similar to the drugs and headaches example: Blocks and drugs play corresponding roles, machines and people play

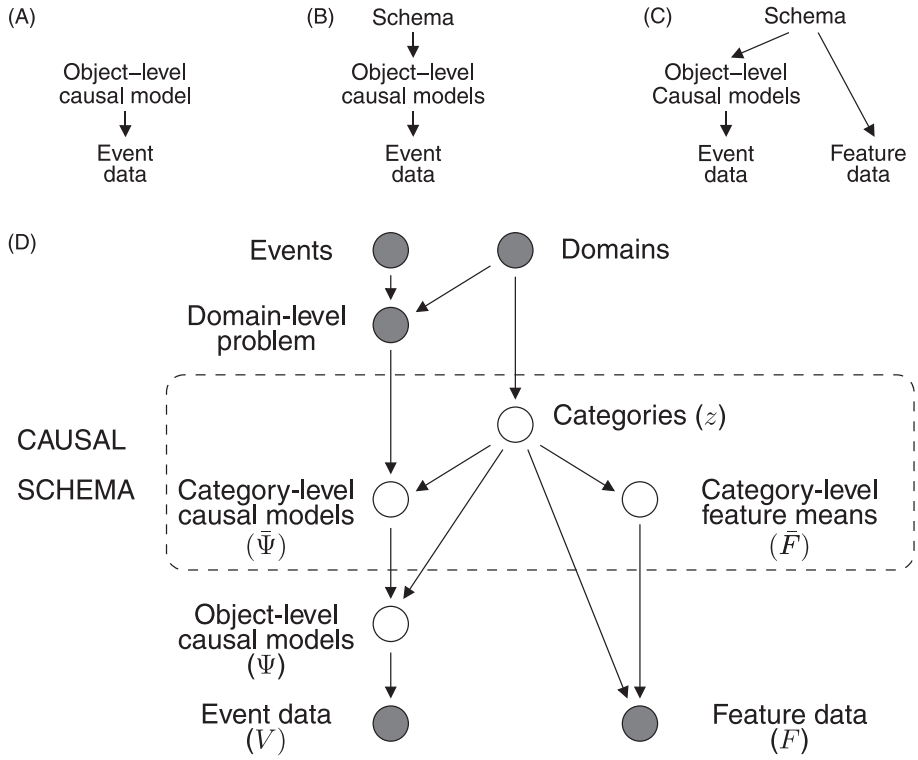


Fig. 3. A hierarchical Bayesian approach to causal learning. (A) Learning a single object-level causal model. (B) Learning causal models for multiple objects. The schema organizes the objects into categories and specifies the causal powers of each category. (C) A generative framework for learning a schema that includes information about the characteristic features of each category. (D) A generative framework that includes (A)–(C) as special cases. Nodes represent variables or bundles of variables and arrows indicate dependencies between variables. Shaded nodes indicate variables that are observed or known in advance, and unshaded nodes indicate variables that must be inferred. We will collectively refer to the categories, the category-level causal models, and the category-level feature means as a causal schema. Note that the hierarchy in Fig. 1A is a subset of the complete model shown here.

corresponding roles, and the event of a machine activating corresponds to the event of a person developing a headache.

The next sections introduce our approach more formally and we develop our framework in several steps. We begin with the problem of learning a single object-level model—for example, learning whether ingesting Doxazosin causes Alice to develop headaches (Fig. 3A). We then turn to the problem of simultaneously learning multiple object-level models (Fig. 3B) and show how causal schemata can help in this setting. We next extend our framework to handle problems where the objects of interest (e.g., people and drugs) have perceptual features that may be correlated with their categories (Fig. 3C). Our final analysis addresses problems where multiple members of the same domain may interact to produce an effect—for example, two drugs may produce a headache when paired although neither causes headaches in isolation.

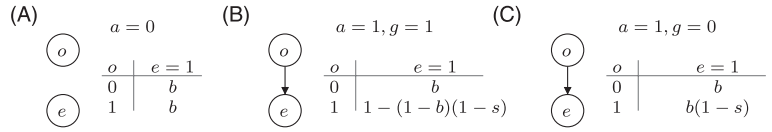


Fig. 4. Causal graphical models that capture three possible relationships between a cause $o$ and an effect $e$. Variable $a$ indicates whether there is a causal relationship between $o$ and $e$, variable $g$ indicates whether this relationship is generative or preventive, and variable $s$ indicates the strength of this relationship. A generative background cause of strength $b$ is always present.

Although we develop our framework in stages and consider several increasingly sophisticated models along the way, the result is a single probabilistic framework that addresses all of the problems we discuss. The framework is shown as a graphical model in Fig. 3D. Each node represents a variable or bundle of variables, and some of the nodes have been annotated with variable names that will be used in later sections of the paper. Arrows between nodes indicate dependencies—for example, the top section of the graphical model indicates that a domain-level problem such as

$$\texttt{ingests}(\texttt{person}, \texttt{drug}) \stackrel{?}{\rightarrow} \texttt{headache}(\texttt{person})$$

is formulated in terms of domains (people and drugs) and events ($\texttt{ingests}(\cdot,\cdot)$ and $\texttt{headache}(\cdot)$). Shaded nodes indicate variables that are observed (e.g., the event data) or specified in advance (e.g., the domain-level problem), and the unshaded nodes indicate variables that must be learned. Note that the three models in Fig. 3A–C correspond to fragments of the complete model in Fig. 3D, and we will build up the complete model by considering these fragments in sequence.

## 3. Learning a single object-level causal model

We begin with the problem of elemental causal induction (Griffiths & Tenenbaum, 2005) or the problem of learning a causal model for a single object-level problem. Our running example will be the problem

$$\texttt{ingests}(\texttt{Alice}, \texttt{Doxazosin}) \stackrel{?}{\rightarrow} \texttt{headache}(\texttt{Alice})$$

where the cause event indicates whether Alice takes Doxazosin and the effect event indicates whether she subsequently develops a headache. Let $o$ refer to the object Doxazosin, and we overload our notation so that $o$ can also refer to the cause event `ingests(Alice, Doxazosin)`. Let $e$ refer to the effect event `headache(Alice)`.

Suppose that we have observed a set of trials where each trial indicates whether or not cause event $o$ occurs, and whether or not the effect $e$ occurs. Data of this kind are often called contingency data, but we refer to them as event data $V$. We assume that the outcome of each trial is generated from an object-level causal model $M$ that captures the causal relationship between $o$ and $e$ (Fig. 5). Having observed the trials in $V$, our beliefs about the causal model can be summarized by the posterior distribution $P(M|V)$:

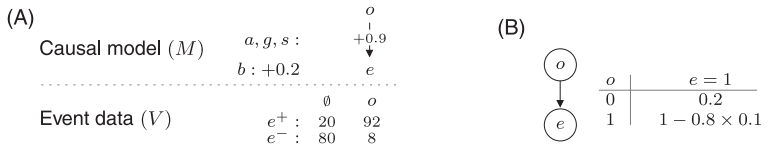


Fig. 5. (A) Learning an object-level causal model *M* from event data *V* (see Fig. 3A). The event data specify the number of times the effect was ($e^+$) and was not ($e^-$) observed when *o* was absent ($\emptyset$) and when *o* was present. The model *M* shown has $a = 1$, $g = 1$, $s = 0.9$, and $b = 0.2$, and it is a compact representation of the graphical model in (B).

$$P(M \mid V) \propto P(V \mid M)P(M). \tag{4}$$

The likelihood term $P(V \mid M)$ indicates how compatible the event data *V* are with model *M*, and the prior $P(M)$ captures prior beliefs about model *M*.

We parameterize the causal model *M* using four causal variables (Figs. 4 and 5). Let *a* indicate whether there is an arrow joining *o* and *e*, and let *g* indicate the polarity of this causal relationship ($g = 1$ if *o* is a generative cause and $g = 0$ if *o* is a preventive cause). Suppose that *s* is the strength of the relationship between *o* and *e*.[1] To capture the possibility that *e* will be present even though *o* is absent, we assume that a generative background cause of strength *b* is always present. We specify the distribution $P(e \mid o)$ by assuming that generative and preventive causes combine according to a network of noisy-OR and noisy-AND-NOT gates (Glymour, 2001).

Now that we have parameterized model *M* in terms of the triple (*a*,*g*,*s*) and the background strength *b*, we can rewrite Eq. 4 as

$$P(a, g, s, b \mid V) \propto P(V \mid a, g, s, b)P(a)P(g)P(s)P(b). \tag{5}$$

To complete the model we must place prior distributions on the four causal variables. We use uniform priors on the two binary variables (*a* and *g*), and we use priors $P(s)$ and $P(b)$ that capture the expectation that *b* will be small and *s* will be large. These priors on *s* and *b* are broadly consistent with the work of Lu, Yuille, Liljeholm, Cheng and Holyoak (2008), who suggest that learners typically expect causes to be necessary (*b* should be low) and sufficient (*s* should be high). Complete specifications of $P(s)$ and $P(b)$ are provided in Appendix A.

To discover the causal model *M* that best accounts for the events in *V*, we can search for the causal variables with maximum posterior probability according to Eq. 5. There are many empirical studies that explore human inferences about a single potential cause and a single effect, and previous researchers (Griffiths & Tenenbaum, 2005; Lu et al., 2008) have shown that a Bayesian approach similar to ours can account for many of these inferences. Here, however, we turn to the less-studied case where people must learn about many objects, each of which may be causally related to the effect of interest.

## 4. Learning multiple object-level models

Suppose now that we are interested in simultaneously learning multiple object-level causal models. For example, suppose that our patient Alice has prescriptions for many different drugs and we want to learn about the effect of each drug:

$$\begin{aligned}
\texttt{ingests(Alice, Doxazosin)} &\xrightarrow{?} \texttt{headache(Alice)}\\
\texttt{ingests(Alice, Prazosin)} &\xrightarrow{?} \texttt{headache(Alice)}\\
\texttt{ingests(Alice, Terazosin)} &\xrightarrow{?} \texttt{headache(Alice)}\\
&\vdots
\end{aligned}$$

For now we assume that Alice takes at most one drug per day, but later we relax this assumption and consider problems where patients take multiple drugs and these drugs may interact. We refer to the $i$th drug as object $o_i$, and as before we overload our notation so that $o_i$ can also refer to the cause event `ingests(Alice,` $o_i$).

Our goal is now to learn a set $\{M_i\}$ of causal models, one for each drug (Figs. 3b and 6). There is a triple $(a_i,g_i,s_i)$ describing the causal model for each drug $o_i$, and we organize these variables into three vectors, $\boldsymbol{a}$, $\boldsymbol{g}$, and $\boldsymbol{s}$. Let $\Psi$ be the tuple ($\boldsymbol{a}$, $\boldsymbol{g}$, $\boldsymbol{s}$, $b$) which includes all the parameters of the causal models. As before, we assume that a generative background cause of strength $b$ is always present.

One strategy for learning multiple object-level models is to learn each model separately using the methods described in the previous section. Although simple, this strategy will not succeed in learning to learn because it does not draw on experience with previous objects when learning a causal model for a novel object that is sparsely observed. We will allow information to be shared across causal models for different objects by introducing the notion of a causal schema. A schema specifies a grouping of the objects into categories and includes category-level causal models which specify the causal powers of each category. The schema in Fig. 6 indicates that there are two categories: objects belonging to category $c_A$ tend to prevent the effect and objects belonging to category $c_B$ tend to cause the effect. The strongest possible assumption is that all members of a category must play identical causal roles. For example, if Doxazosin and Prazosin belong to the same category, then the causal models for these two drugs should be identical. We relax this strong assumption and assume instead that members of the same category play similar causal roles. More precisely, we assume that the object-level models corresponding to a given category-level causal model are drawn from a common distribution.

Formally, let $z_i$ indicate the category of $o_i$, and let $\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$, $\bar{\boldsymbol{s}}$, and $\bar{b}$ be schema-level analogs of $\boldsymbol{a}$, $\boldsymbol{g}$, $\boldsymbol{s}$, and $b$. Variable $\bar{a}(c)$ is the probability that any given object belonging to category $c$ will be causally related to the effect, variables $\bar{g}(c)$ and $\bar{s}(c)$ specify the expected polarity and causal strength for objects in category $c$, and variable $\bar{b}$ specifies the expected strength of the generative background cause. Even though $\boldsymbol{a}$ and $\boldsymbol{g}$ are vectors of probabilities, Fig. 6 simplifies by showing each $\bar{a}(c)$ and $\bar{g}(c)$ as a binary variable. To generate a causal model for each object, we assume that each arrow variable $a_i$ is generated by tossing a coin with

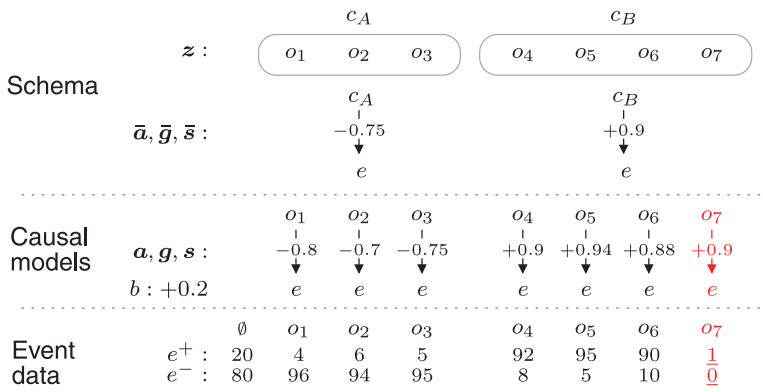


Fig. 6. Learning a schema and a set of object-level causal models (see Fig. 3B). $\boldsymbol{z}$ specifies a set of categories, where objects belonging to the same category have similar causal powers, and $\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$, and $\bar{\boldsymbol{s}}$ specify a set of category-level causal models. Note that the schema supports inferences about an object ($o_7$, counts underlined in red) that is very sparsely observed.

weight $\bar{a}(z_i)$, that each polarity $g_i$ is generated by tossing a coin with weight $\bar{g}(z_i)$, and that each strength $s_i$ is drawn from a distribution parameterized by $\bar{s}(z_i)$. Let $\bar{\Psi}$ be a tuple $(\bar{\boldsymbol{a}}, \bar{\boldsymbol{g}}, \bar{\boldsymbol{s}}, \bar{b})$ that includes all parameters of the causal schema. A complete description of each parameter is provided in Appendix A.

Now that the generative approach in Fig. 1A has been fully specified we can use it to learn the category assignments $\boldsymbol{z}$, the category-level models $\bar{\Psi}$, and the object-level models $\Psi$ that are most probable given the events $V$ that have been observed:

$$P(\boldsymbol{z}, \bar{\Psi}, \Psi \mid V) \propto P(V \mid \Psi) P(\Psi \mid \bar{\Psi}, \boldsymbol{z}) P(\bar{\Psi} \mid \boldsymbol{z}) P(\boldsymbol{z}). \tag{6}$$

The distribution $P(V \mid \Psi)$ is defined by assuming that the contingency data for each object-level model are generated in the standard way from that model. The distribution $P(\Psi \mid \bar{\Psi}, \boldsymbol{z})$ specifies how the model parameters $\Psi$ are generated given the category-level models $\bar{\Psi}$ and the category assignments $\boldsymbol{z}$. To complete the model, we need prior distributions on $\boldsymbol{z}$ and the category-level models $\bar{\Psi}$. Our prior $P(\boldsymbol{z})$ assigns some probability mass to all possible partitions but favors partitions that use a small number of categories. Our prior $P(\bar{\Psi} \mid \boldsymbol{z})$ captures the expectation that generative and preventive causes are equally likely a priori, that causal strengths are likely to be high, and that the strength of the background cause is likely to be low. Full details are provided in Appendix A.

Fig. 6 shows how a schema and a set of object-level causal models (top two levels) can be simultaneously learned from the event data $V$ in the bottom level. All of the variables in the figure have been set to values with high posterior probability according to Eq. 6: for instance, the partition $\boldsymbol{z}$ shown is the partition with maximum posterior probability. Note that learning a schema allows a causal model to be learned for object $o_7$, which is very sparsely observed (see the underlined entries in the bottom level of Fig. 6). On its own, a single trial might not be very informative about the causal powers of this object, but experience

with previous objects allows the model to predict that $o_7$ will produce the effect about as regularly as the other members of category $c_B$.

To compute the predictions of our model we used Markov chain Monte Carlo methods to sample from the posterior distribution in Eq. 6. A more detailed description of this inference algorithm is provided in Appendix A, but note that this algorithm is not intended as a model of psychological processing. The primary contribution of this section is the computational theory summarized by Eq. 6, and there will be many ways in which the computations required by this theory can be approximately implemented.

## 5. Experiments 1 and 2: Testing the basic schema-learning model

Our schema-learning model attempts to satisfy two criteria when learning about the causal powers of a novel object. When information about the new object is sparse, predictions about this object should be based primarily on experience with previous objects. Relying on past experience will allow the model to go beyond the sparse and noisy observations that are available for the novel object. Given many observations of the novel object, however, the model should rely heavily on these observations and should tend to ignore its observations of previous objects. Discounting past experience in this way will allow the model to be flexible if the new object turns out to be different from all previous objects.

Our first two experiments explore this tradeoff between conservatism and flexibility. Both experiments used blocks and machines like the examples in Fig. 2. As mentioned already, the domain-level problem for this setting is:

$$\texttt{inside}(\texttt{block},\texttt{machine})\ \texttt{\&}\ \texttt{button\_pressed}(\texttt{machine}) \xrightarrow{?} \texttt{activate}(\texttt{machine})$$

In terms of the notation we have been using, each block is an object $o_i$, each button press corresponds to a trial, and the effect $e$ indicates whether the machine activates on a given trial.

Experiment 1 studied people's ability to learn a range of different causal schemata from observed events and to use these schemata to rapidly learn about the causal powers of new, sparsely observed objects. To highlight the influence of causal schemata, inferences about each new object were made after observing at most one trial where the object was placed inside the machine. Across conditions, we varied the information available during training, and our model predicts that these different sets of observations should lead to the formation of qualitatively different schemata, and hence to qualitatively different patterns of inference about new, sparsely observed objects.

Experiment 2 explores in more detail how observations of a new object interact with a learned causal schema. Instead of observing a single trial for each new object, participants now observed seven trials and judged the causal power of the object after each one. Some sequences of trials were consistent with the schema learned during training, but others were inconsistent. Given these different sequences, our model predicts how and when learners should overrule their schema-based expectations when learning a causal model for a novel

object. These predicted learning curves—or ''unlearning curves''—were tested against the responses provided by participants.

Our first two experiments were also designed in part to evaluate our model relative to other computational accounts. Although we know of no previous model that attempts to capture causal learning at multiple levels of abstraction, we will consider some simple models inspired by standard models in the categorization literature. These models are discussed and analyzed following the presentation of Experiments 1 and 2.

### *5.1. Experiment 1: One-shot causal learning*

Experiment 1 explores whether participants can learn different kinds of schemata and use these schemata to rapidly learn about the causal powers of new objects. In each condition of the experiment, participants initially completed a training phase where they placed each of eight objects into the machine multiple times and observed whether the machine activated on each trial. In different conditions they observed different patterns of activation across these training trials. In each condition, the activations observed were consistent with the existence of one or two categories, and these categories had qualitatively different causal powers across the different conditions. After each training phase, participants completed a ''one-shot learning'' task, where they made predictions about test blocks after seeing only a single trial involving each block.

#### *5.1.1. Participants*

Twenty-four members of the MIT community were paid for participating in this experiment.

#### *5.1.2. Stimuli*

The experiment used a custom-built graphical interface that displayed a machine and some blocks (Fig. 2A). Participants could drag the blocks around, and they were able to place up to one block inside the machine at a time. Participants could also interact with the machine by pressing the GO button and observing whether the machine activated. The blocks used for Experiments 1 and 2 were perceptually indistinguishable, and their causal powers could therefore not be predicted on the basis of their physical appearance.

#### *5.1.3. Design*

The experiment includes four within-participant conditions and the training data for each condition are summarized in Fig. 7. The first condition ($p = \{0, 0.5\}$) includes two categories of blocks: blocks in the first category never activate the machine, and blocks in the second category activate the machine about half the time. The second condition ($p = \{0.1, 0.9\}$) also includes two categories: blocks in the first category rarely activate the machine, and blocks in the second category usually activate the machine. The remaining conditions each include only one category of blocks: blocks in the third condition ($p = 0$) never activate the machine, and blocks in the fourth condition ($p = 0.1$) activate the machine rarely.

| Condition | | Training data | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\emptyset$ | $o_1$ | $o_2$ | $o_3$ | $o_4$ | $o_5$ | $o_6$ | $o_7$ | $o_8$ |
| $p = \{0, 0.5\}$ | $e^+$ : | 0 | 0 | 0 | 0 | 0 | 5 | 4 | 6 | 1 |
| | $e^-$ : | 10 | 10 | 10 | 10 | 1 | 5 | 6 | 4 | 0 |
| | | $\emptyset$ | $o_1$ | $o_2$ | $o_3$ | $o_4$ | $o_5$ | $o_6$ | $o_7$ | $o_8$ |
| $p = \{0.1, 0.9\}$ | $e^+$ : | 0 | 1 | 2 | 1 | 0 | 9 | 8 | 9 | 1 |
| | $e^-$ : | 10 | 9 | 8 | 9 | 1 | 1 | 2 | 1 | 0 |
| | | $\emptyset$ | $o_1$ | $o_2$ | $o_3$ | $o_4$ | $o_5$ | $o_6$ | | |
| $p = 0$ | $e^+$ : | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| | $e^-$ : | 10 | 10 | 10 | 10 | 10 | 10 | 10 | | |
| | | $\emptyset$ | $o_1$ | $o_2$ | $o_3$ | $o_4$ | $o_5$ | $o_6$ | | |
| $p = 0.1$ | $e^+$ : | 0 | 1 | 2 | 1 | 2 | 1 | 2 | | |
| | $e^-$ : | 10 | 9 | 8 | 9 | 8 | 9 | 8 | | |

Fig. 7. Training data for the four conditions of Experiment 1. In each condition, the first column of each table shows that the empty machine fails to activate on each of the 10 trials. Each remaining column shows the outcome of one or more trials when a single block is placed inside the machine. For example, in the $p = \{0, 0.5\}$ condition block $o_1$ is placed in the machine 10 times and fails to activate the machine on each trial.

### *5.1.4. Procedure*

At the start of each condition, participants are shown an empty machine and asked to press the GO button 10 times. The machine fails to activate on each occasion. One by one the training blocks are introduced, and participants place each block in the machine and press the GO button one or more times. The outcomes of these trials are summarized in Fig. 7. For example, the $p = \{0, 0.5\}$ condition includes eight training blocks in total, and the block shown as $o_1$ in the table fails to activate the machine on each of 10 trials. After the final trial for each block, participants are asked to imagine pressing the GO button 100 times when this block is inside the machine. They then provide a rating which indicates how likely it is that the total number of activations will fall between 0 and 20. All ratings are provided on a seven-point scale where one is labeled as ''very unlikely,'' seven is labeled as ''very likely,'' and the other values are left unlabeled. Ratings are also provided for four other intervals: between 20 and 40, between 40 and 60, between 60 and 80, and between 80 and 100. Each block remains on screen after it is introduced, and by the end of the training phase six or eight blocks are therefore visible onscreen. After the training phase two test blocks are introduced, again one at a time. Participants provide ratings for each block before it has been placed in the machine, and after a single trial. One of the test blocks ($o^+$) activates the machine on this trial, and the other ($o^-$) does not.

The set of four conditions is designed to test the idea that inductive constraints and inductive flexibility are both important. The first two conditions test whether experience with the training blocks allows people to extract constraints that are useful when learning about the causal powers of the test blocks. Conditions three and four explore cases where these

constraints need to be overruled. Note that test block $o^+$ is surprising in these conditions because the training blocks activate the machine rarely, if at all.

To encourage participants to think about the conditions separately, machines and blocks of different colors were used for each condition. Note, however, that the blocks within each condition were always perceptually identical. The order in which the conditions were presented was counterbalanced according to a Latin square design. The order of the training blocks and the test blocks within each condition was also randomized subject to several constraints. First, the test blocks were always presented after the training blocks. Second, in conditions $p = \{0, 0.5\}$ and $p = \{0.1, 0.9\}$ the first two training blocks in the sequence always belonged to different categories, and the two sparsely observed training blocks ($o_4$ and $o_8$) were always the third and fourth blocks in the sequence. Finally, in the $p = 0$ condition test block $o^+$ was always presented second, because this block is unlike any of the training blocks and may have had a large influence on predictions about any block which followed it.

### *5.1.5. Model predictions*

Fig. 8 shows predictions when the schema-learning model is applied to the data in Fig. 7. Each plot shows the posterior distribution on the activation strength of a test block: the probability $P(e \mid o)$ that the block will activate the machine on a given trial. Because the background rate is zero, this distribution is equivalent to a distribution on the causal power

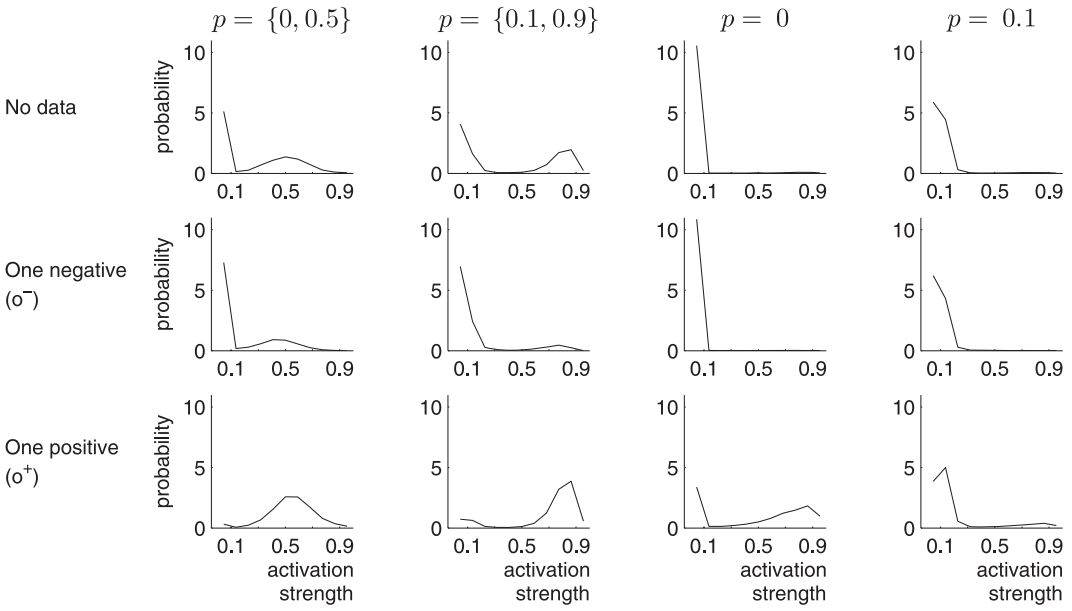


Fig. 8. Predictions of the schema-learning model for Experiment 1. Each subplot shows the posterior distribution on the activation strength of a test block. There are three predictions for each condition: The first row shows inferences about a test block before this block has been placed in the machine, and the remaining rows show inferences after a single negative ($o^-$) or positive ($o^+$) trial is observed. Note that the curves represent probability density functions and can therefore attain values greater than 1.

(Cheng, 1997) of the test block. Recall that participants were asked to make predictions about the number of activations expected across 100 trials. If we ask our model to make the same predictions, the distributions on the total number of activations will be discrete distributions with shapes similar to the distributions in Fig. 8.

The plots in the first row show predictions about a test block before it is placed in the machine. The first plot indicates that the model has discovered two causal categories, and it expects that the test block will activate the machine either very rarely or around half of the time. The two peaks in the second plot again indicate that the model has discovered two causal categories, this time with strengths around 0.1 and 0.9. The remaining two plots are unimodal, suggesting that only one causal category is needed to explain the data in each of the $p = 0$ and $p = 0.1$ conditions.

The plots in the second row show predictions about a test block ($o^-$) that fails to activate the machine on one occasion. All of the plots have peaks near 0 or 0.1. Because each condition includes blocks that activate the machine rarely or not at all, the most likely hypothesis is always that $o^-$ is one of these blocks. Note, however, that the first plot has a small bump near 0.5, indicating that there is some chance that test block $o^-$ will activate the machine about half of the time. The second plot has a small bump near 0.9 for similar reasons.

The plots in the third row show predictions about a test block ($o^+$) that activates the machine on one occasion. The plot for the first condition peaks near 0.5, which is consistent with the hypothesis that blocks which activate the machine at all tend to activate it around half the time. The plot for the second condition peaks near 0.9, which is consistent with the observation that some training blocks activated the machine nearly always. The plot for the third condition has peaks near 0 and near 0.9. The first peak captures the idea that the test block might be similar to the training blocks, which activated the machine very rarely. Given that none of the training blocks activated the machine, one positive trial is enough to suggest that the test block might be qualitatively different from all previous blocks, and the second peak captures this hypothesis. The curve for the final condition peaks near 0.1, which is the frequency with which the training blocks activated the machine.

*5.1.6. Results*

The four columns of Fig. 9 show the results for each condition. Each participant provided ratings for five intervals in response to each question, and these ratings can be plotted as a curve. Fig. 9 shows the mean curve for each question. The first row shows predictions before a test block has been placed in the machine (responses for test blocks $o^-$ and $o^+$ have been combined). The second and third rows show predictions after a single trial for test blocks $o^-$ and $o^+$.

The first row provides a direct measure of what participants have learned during the training for each condition. Note first that the plots for the four rows are rather different, suggesting that the training observations have shaped people's expectations about novel blocks. A two-factor ANOVA with repeated measures supports this conclusion, and it indicates that there are significant main effects of interval [$F(4,92) = 31.8$, $p < .001$] and condition [$F(3,69) = 15.7$, $p < .001$] but no significant interaction between interval and condition [$F(12,276) = 0.74$, $p > .5$].

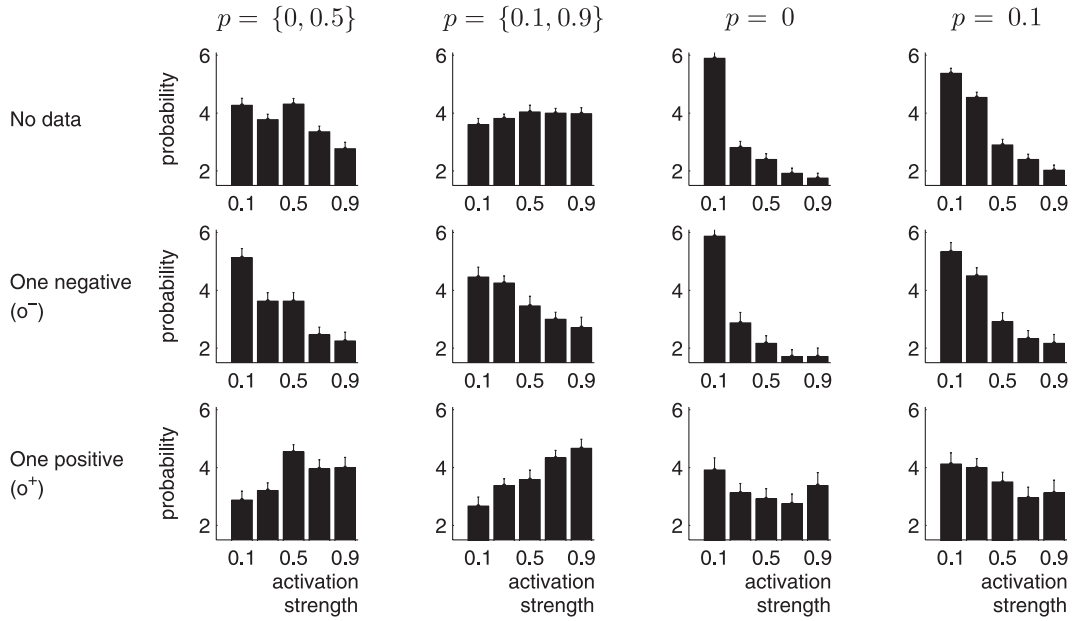


Fig. 9. Results for the four conditions in Experiment 1. Each subplot shows predictions about a new object that will undergo 100 trials, and each bar indicates the probability that the total number of activations will fall within a certain interval. The *x*-axis shows the activation strengths that correspond to each interval and the *y*-axis shows probability ratings on a scale from one (very unlikely) to seven (very likely). All plots show mean responses across 24 participants. Error bars for this plot and all remaining plots show the standard error of the mean.

In three of the four conditions, the human responses in the top row of Fig. 9 are consistent with the model predictions in Fig. 8. As expected, the curves for the $p = 0$ and $p = 0.1$ conditions indicate an expectation that the test blocks will probably fail to activate the machine. The curve for the $p = \{0, 0.5\}$ condition peaks in the same places as the model prediction, suggesting that participants expect that each test block will either activate the machine very rarely or about half of the time. The first (0.1) and third (0.5) bars in the plot are both greater than the second (0.3) bar, and paired sample *t* tests indicate that both differences are statistically significant ($p < .05$, one-tailed). The $p = \{0, 0.5\}$ curve is therefore consistent with the idea that participants have discovered two categories.

The responses for the $p = \{0.1, 0.9\}$ condition provide no evidence that participants have discovered two causal categories. The curve for this condition is flat or unimodal and does not match the bimodal curve predicted by the model. One possible interpretation is that learners cannot discover categories based on probabilistic causal information. As suggested by the $p = \{0, 0.5\}$ condition, learners might distinguish between blocks that never produce the effect and those that sometimes produce the effect, but not between blocks that produce the effects with different strengths. A second possible interpretation is that learners can form categories based on probabilistic information but require more statistical evidence than we provided in Experiment 1. Our third experiment supports this second interpretation and demonstrates that learners can form causal categories on the basis of probabilistic evidence.

Consider now the third row of Fig. 9, which shows predictions about a test block ($o^+$) that has activated the machine exactly once. As before, the differences between these plots suggest that experience with previous blocks shapes people's inferences about a sparsely observed novel block. A two-factor ANOVA with repeated measures supports this conclusion, and indicates that there is no significant main effect of interval [$F(4,92) = .46$, $p > .5$], but that there is a significant main effect of condition [$F(3,69) = 4.20$, $p < .01$] and a significant interaction between interval and condition [$F(12,276) = 6.90$, $p < .001$]. Note also that all of the plots in the third row peak in the same places as the curves predicted by the model (Fig. 8A). For example, the middle (0.5) bar in the $p = \{0, 0.5\}$ condition is greater than the bars on either side, and paired sample $t$ tests indicate that both differences are statistically significant ($p < .05$, one-tailed). The plot for the $p = 0$ condition provides some support for a second peak near 0.9, although a paired-sample $t$ test indicates that the difference between the fifth (0.9) and fourth (0.7) bars is only marginally significant ($p < .1$, one-tailed). Our second experiment explores this condition in more detail, and it establishes more conclusively that a single positive observation can be enough for a learner to decide that a block is different from all previously observed blocks.

Consider now the second row of Fig. 9, which shows predictions about a test block ($o^-$) that has failed to activate the machine exactly once. The plots in this row are all decaying curves, because each condition includes blocks that activate the machine rarely or not at all. Again, though, the differences between the curves are interpretable and match the predictions of the model. For instance, the $p = 0$ curve decays more steeply than the others, which makes sense because the training blocks for this condition never activate the machine. In particular, note that the difference between the first (0.1) and second (0.3) bars is greater in the $p = 0$ condition than the $p = 0.1$ condition ($p < .001$, one-tailed).

Although our primary goal in this paper is to account for the mean responses to each question, the responses of individual participants are also worth considering. Kemp (2008) presents a detailed analysis of individual responses and shows that in all cases except one the shape of the mean curve is consistent with the responses of some individuals. The one exception is the $o^+$ question in the $p = 0$ condition, where no participant generated a U-shaped curve, although some indicated that $o^+$ is unlikely to activate the machine and others indicated that $o^+$ is very likely to activate the machine on subsequent trials. This disagreement suggests that the $p = 0$ condition deserves further attention, and our second experiment explores this condition in more detail.

### *5.2. Experiment 2: Discovering new causal categories*

Causal schemata support inferences about new objects that are sparsely observed, but sometimes these inferences are wrong and will have to be overruled when a new object turns out to be qualitatively different from all previous objects. Experiment 1 provided some suggestive evidence that human learners will overrule a schema when necessary. In the $p = 0$ condition, participants observed six blocks that never activated the machine, then saw a single trial where a new block ($o^+$) activated the machine. The results in Fig. 9 suggest that some participants inferred that the new block might be qualitatively different from the

previous blocks. This finding suggests that a single observation of a new object is sometimes enough to overrule expectations based on many previous objects, but several trials may be required before learners are confident that a new object is unlike any of the previous objects. To explore this idea, Experiment 2 considers two cases where participants receive increasing evidence that a new object is different from all previously encountered objects.

### 5.2.1. Participants

Sixteen members of the MIT community were paid for participating in this experiment.

### 5.2.2. Design and procedure

The experiment includes two within-participant conditions ($p = 0$ and $p = 0.1$) that correspond to conditions 3 and 4 of Experiment 1. Each condition is very similar to the corresponding condition from Experiment 1 except for two changes. Seven observations are now provided for the two test blocks: for test block $o^-$, the machine fails to activate on each trial, and for test block $o^+$ the machine activates on all test trials except the second. Participants rate the causal strength of each test block after each trial and also provide an initial rating before any trials have been observed. As before, participants are asked to imagine placing the test block in the machine 100 times, but instead of providing ratings for five intervals they now simply predict the total number of activations out of 100 that they expect to see.

### 5.2.3. Model predictions

Fig. 10 shows the results when the schema-learning model is applied to the tasks in Experiment 2. In both conditions, predictions about the test blocks track the observations provided, and the curves rise after each positive trial and fall after each negative trial.

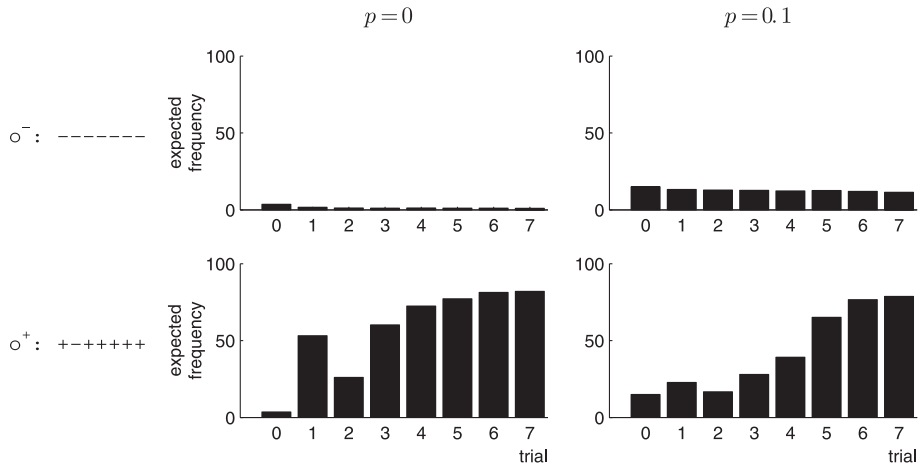


Fig. 10. Predictions of the schema-learning model for Experiment 2. A new block is introduced that is either similar ($o^-$) or different ($o^+$) from all previous blocks, and the trials for each block are shown on the left of the figure. Each plot shows how inferences about the causal power of the block change with each successive trial.

The most interesting predictions involve test block $o^+$, which is qualitatively different from all of the training blocks. The $o^+$ curves for both conditions attain similar values by the final prediction, but the curve for the $p = 0$ condition rises more steeply than the curve for the $p = 0.1$ condition. Because the training blocks in the $p = 0.1$ condition activate the machine on some occasions, the model needs more evidence in this condition before concluding that block $o^+$ is different from all of the training blocks.

The predictions about test block $o^-$ also depend on the condition. In the $p = 0$ condition, none of the training blocks activates the machine, and the model predicts that $o^-$ will also fail to activate the machine. In the $p = 0.1$ condition, each training block can be expected to activate the machine about 15 times out of 100. The curve for this condition begins at around 15, then gently decays as $o^-$ repeatedly fails to activate the machine.

### *5.2.4. Results*

Fig. 11 shows average learning curves across 16 participants. The curves are qualitatively similar to the model predictions, and as predicted the $o^+$ curve for the $p = 0$ condition rises more steeply than the corresponding curve for the $p = 0.1$ condition. Note that a simple associative account might predict the opposite result, because the machine in condition $p = 0.1$ activates more times overall than the machine in condition $p = 0$. To support our qualitative comparison between the $o^+$ curves in the two conditions, we ran a two-factor ANOVA with repeated measures. Because we expect that the $p = 0$ curve should be higher than the $p = 0.1$ curve from the second judgment onwards, we excluded the first judgment from each condition. There are significant main effects of condition [$F(1,15) = 6.11$, $p < .05$] and judgment number [$F(6,90) = 43.21$, $p < .01$], and a significant interaction between condition and judgment number [$F(6,90) = 2.67$, $p < .05$]. Follow-up paired-sample $t$ tests indicate that judgments two through six are reliably greater in the $p = 0$ condition (in all

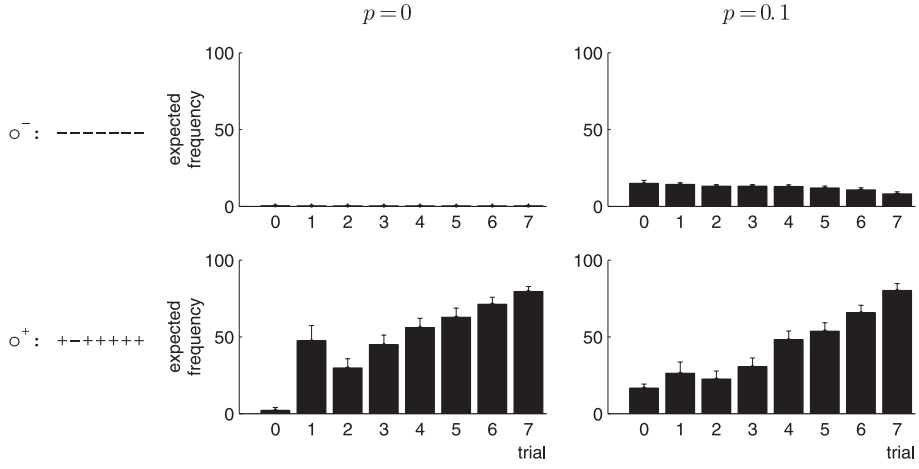


Fig. 11. Mean responses to Experiment 1. The average learning curves closely match the model predictions in Fig. 10.

cases $p < .05$, one-tailed), supporting the prediction that participants are quicker in the $p = 0$ condition to decide that block $o^+$ is qualitatively different from all previous blocks.

### 5.3. Alternative models

As mentioned already, our experiments explore the tradeoff between conservatism and flexibility. When a new object is sparsely observed, the schema-learning model assumes that this object is similar to previously encountered objects (Experiment 1). Once more observations become available, the model may decide that the new object is different from all previous objects and should therefore be assigned to its own category (Experiment 2). We can compare the schema-learning model to two alternatives: an *exemplar* model that is overly conservative, and a *bottom-up* model that is overly flexible. The exemplar model assumes that each new object is just like one of the previous objects, and the bottom-up model ignores all of its previous experience when making predictions about a new object.

We implemented the bottom-up model by assuming that the causal power of a test block is identical to its empirical power—the proportion of trials on which it has activated the machine. Predictions of this model are shown in Fig. 12. When applied to Experiment 1, the most obvious failing of the bottom-up model is that it makes identical predictions about all four conditions. Note that the model does not make predictions about the first row of Fig. 8A, because at least one test trial is needed to estimate the empirical power of a new block. When applied to Experiment 2, the model is unable to make predictions before any trials have been observed for a given object, and after a single positive trial the model leaps to the conclusion that test object $o^+$ will always activate the machine. Neither prediction matches the human data, and the model also fails to predict any difference between the $p = 0$ and $p = 0.1$ conditions.

We implemented the exemplar model by assuming that the causal power of each training block is identical to its empirical power, and that each test block is identical to one of the training blocks. The model, however, does not know which training block the test block will match, and it makes a prediction that considers the empirical powers of all training blocks, weighting each one by its proximity to the empirical power of the test block. Formally, the distribution $d_n$ on the strength of a novel block is defined to be

$$d_n = \frac{\sum_i w_i d_i}{\sum_i w_i} \tag{7}$$

where $d_i$ is the distribution for training block $i$, and is created by dividing the interval [0,1] into eleven equal intervals, setting $d_i(x) = 11$ for all values $x$ that belong to the same interval as the empirical power of block $i$, and setting $d_i(x) = 0$ for all remaining values. Each weight $w_i$ is set to $1 - \mid p_n - p_i \mid$, where $p_n$ is the empirical power of the novel block and $p_i$ is the empirical power of training block $i$. As Eq. 7 suggests, the exemplar model is closely related to exemplar models of categorization (Medin & Schaffer, 1978; Nosofsky, 1986).

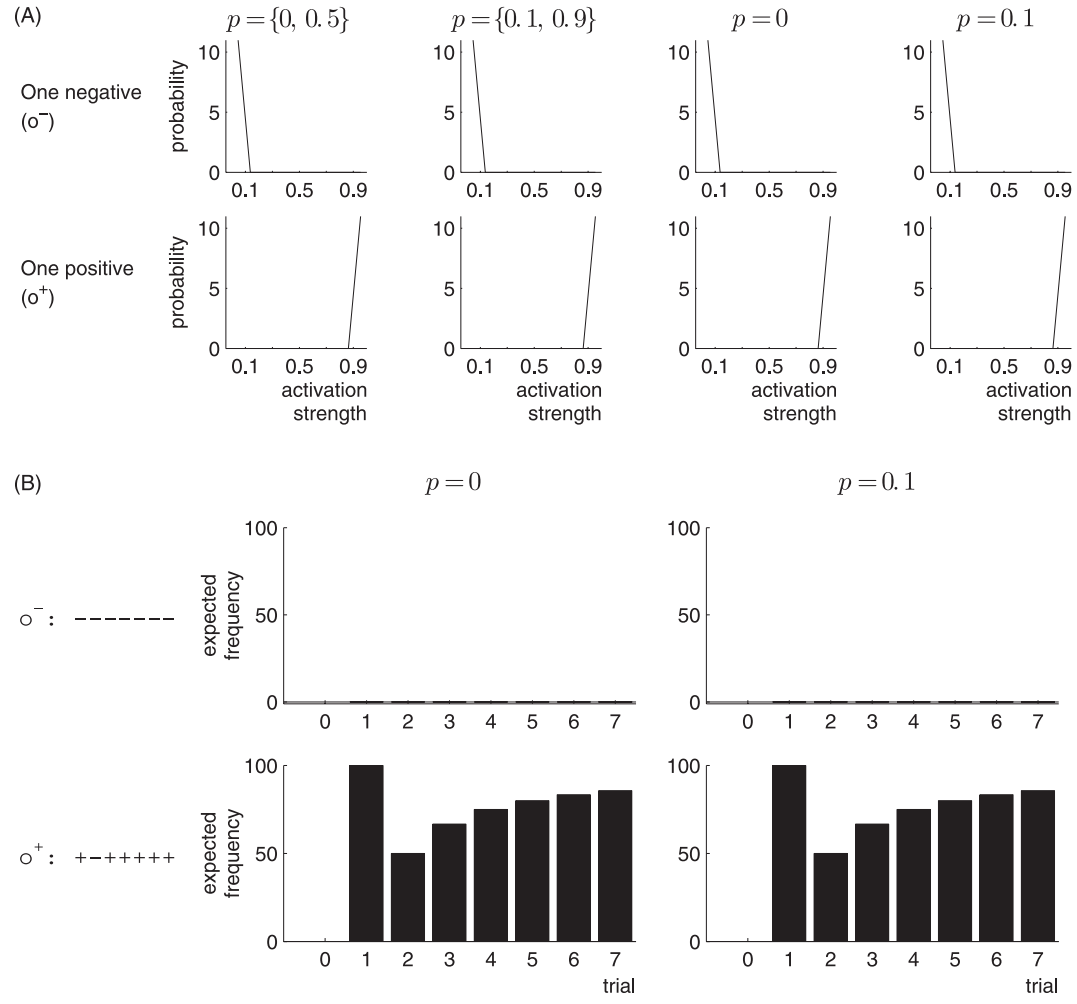


Fig. 12. Predictions of the bottom-up model for (A) Experiment 1 and (B) Experiment 2. In both cases the model fails to account for the differences between conditions.

Predictions of the exemplar model are shown in Fig. 13. The model accounts fairly well for the results of Experiment 1 but is unable to account for Experiment 2. Because the model assumes that test object $o^+$ is just like one of the training objects, it is unable to adjust when $o^+$ activates the machine more frequently than any previous object.

Overall, neither baseline model can account for our results. The bottom-up model is too quick to throw away observations of previous objects, and the exemplar model is unable to handle new objects that are qualitatively different from all previous objects. Other baseline models might be considered, but we are aware of no simple alternative that will account for all of our data.

Our first two experiments deliberately focused on a very simple setting where causal schemata are learned and used, but real-world causal learning is often more complex. The

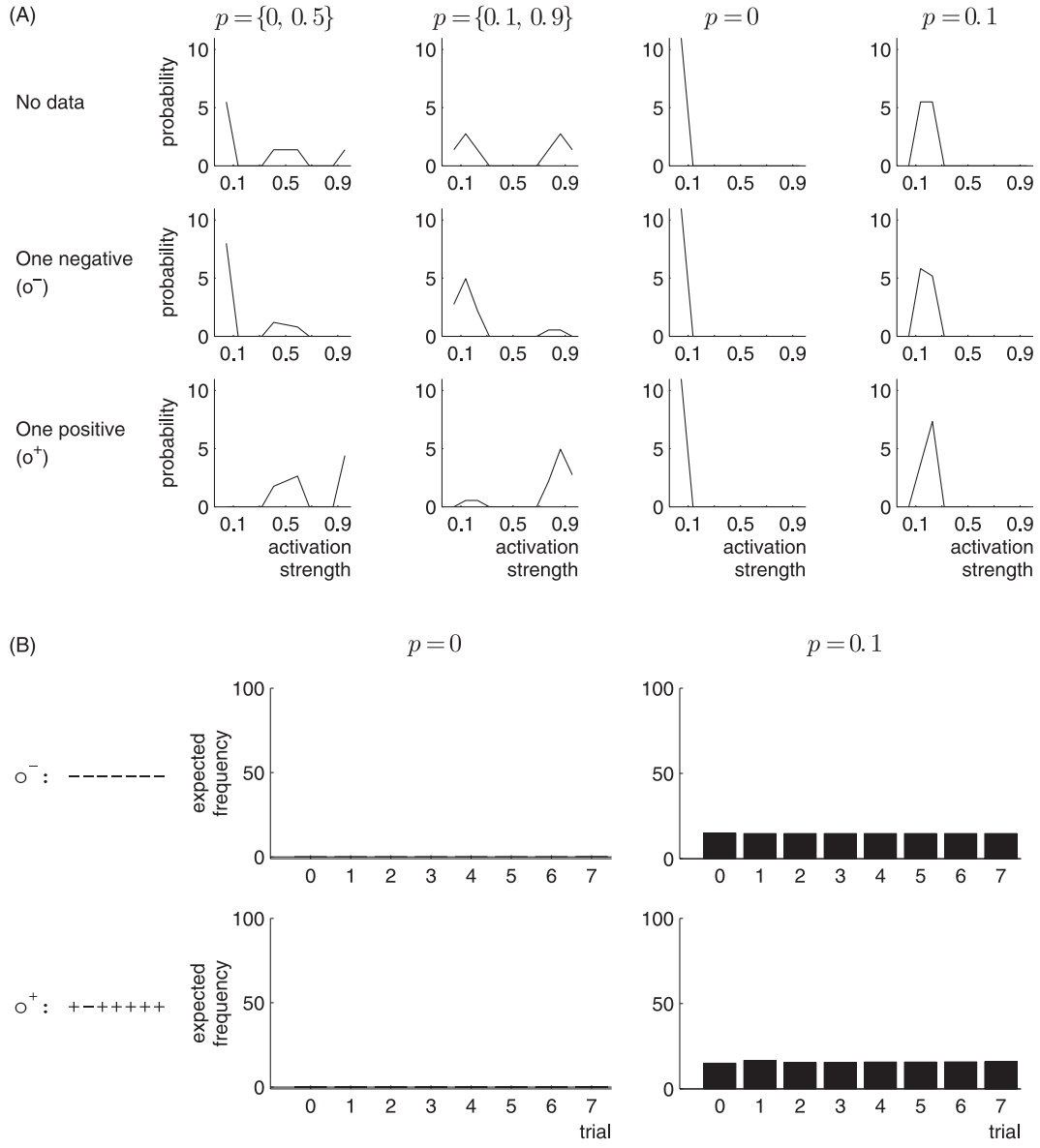


Fig. 13. Predictions of the exemplar model for (A) Experiment 1 and (B) Experiment 2. The model accounts fairly well for Experiment 1 but fails to realize that test block $o^+$ in Experiment 2 is qualitatively different from all previous blocks.

rest of the paper will address some of these complexities: in particular, we show how our framework can incorporate perceptual features and can handle contexts where causes interact to produce an effect.

## 6. Learning causal categories given feature data

Imagine that you are allergic to nuts, and that one day you discover a small white sphere in your breakfast cereal—a macadamia nut, although you do not know it. To discover the causal powers of this novel object you could collect some causal data—you could eat it and wait to see what happens. Probably, however, you will observe the features of the object, including its color, shape, and texture, and decide to avoid it because it is similar to other allergy-producing foods that you have encountered.

Our hierarchical Bayesian approach can readily handle the idea that members of a given category tend to have similar features in addition to similar causal powers (Figs. 3C and 14). Suppose that we have a matrix $F$ which captures many features of the objects under consideration, including their sizes, shapes, and colors. We assume that objects belonging to the same category have similar features. For instance, the schema in Fig. 14 specifies that objects of category $c_B$ tend to have features $f_1$ through $f_4$, but objects of category $c_A$ tend not to have these features. Formally, let the schema parameters include a matrix $\bar{F}$, where $\bar{f}_j(c)$ specifies the expected value of feature $f_j$ within category $c$ (Fig. 3D). Building on previous models of categorization (Anderson, 1991), we assume that the value of $f_j$ for object $o_i$ is generated by tossing a coin with bias $\bar{f}_j(z_i)$. Our goal is now to use the features $F$ along with the events $V$ to learn a schema and a set of object-level causal models:

$$P(\mathbf{z}, \bar{F}, \bar{\Psi}, \Psi \mid F, V) \propto P(F \mid \bar{F}, \mathbf{z}) P(\bar{F} \mid \mathbf{z}) P(V \mid \Psi) P(\Psi \mid \bar{\Psi}, \mathbf{z}) P(\bar{\Psi} \mid \mathbf{z}) P(\mathbf{z}). \tag{8}$$

There are many previous models for discovering categories of objects with similar features (Anderson, 1991; Love, Medin, & Gureckis, 2004), and feature-based categorization is

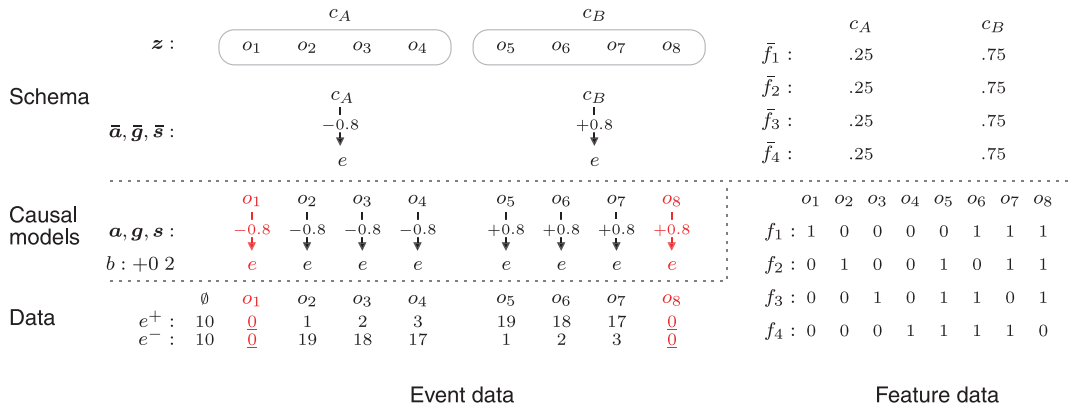


Fig. 14. Learning a schema and a set of object-level causal models given event and feature data (see Fig. 3C). Objects belonging to the same category have similar causal powers and similar features, and $\bar{f}_i$ specifies the expected value of feature $f_i$ within each category. Note that the schema supports inferences about the causal powers of two objects ($o_1$ and $o_8$, counts underlined in red) that are very sparsely observed. The event and feature data shown are similar to the data used for Experiment 3.

sometimes pitted against causal categorization (Gopnik & Sobel, 2000). Our schema-learning model is based on the idea that real-world categories are often distinguished both by their characteristic features and their characteristic causal interactions. More often than not, one kind of information will support the categories indicated by the other, but there will also be cases where the causal data and the feature data conflict. In a later section we show how our framework can learn whether causal data or feature data provide the more reliable guide to category membership.

## 7. Experiment 3: Combining causal and feature data

Our first two experiments suggest that causal schemata allow causal models for novel objects to be rapidly learned, sometimes on the basis of a single causal event. Our third experiment explores whether learners can acquire a causal model for an object on the basis of its perceptual features alone. The objects in this experiment can be organized into two family resemblance categories on the basis of their perceptual features, and these two categories are associated with different causal powers. Observing the features of a novel object should allow a learner to assign it to one of these categories and to make inferences about its causal powers.

### *7.1. Participants*

Twenty-four members of the MIT community were paid for participating in this experiment.

### *7.2. Procedure*

Participants are initially shown an empty machine that activates on 10 of the 20 trials. Ten blocks then appear on screen, and the features of these blocks support two family resemblance categories (see Figs. 2 and 15). Before any of the blocks is placed in the machine, participants are informed that the blocks are laid out randomly, and they are encouraged to drag them around and organize them in a way that will help them predict what effect they will have on the machine. Participants then observe 20 trials for blocks $o_1$ through $o_8$, and see that blocks $o_1$ through $o_4$ activate the machine rarely, but blocks $o_5$ through $o_8$ activate the machine most of the time. After 20 trials for each block, participants respond to the same question used in Experiment 1: They imagine 100 trials involving the block and rate how likely it is that the total number of activations will fall into each of five intervals. After this training phase, participants answer the same question for test blocks $o^-$ and $o^+$ without seeing *any* trials involving these blocks. Experiment 1 explored one-shot learning, and this new task might be described as zero-shot learning. After making predictions for the two test blocks, participants are asked to sort the blocks into two categories ‘‘according to their effect on the machine’’ and to explain the categories they chose.

| | $\emptyset$ | $o^-$ | $o_1$ | $o_2$ | $o_3$ | $o_4$ | $o_5$ | $o_6$ | $o_7$ | $o_8$ | $o^+$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{e}^+$ : | 10 | 0 | 3 | 2 | 1 | 2 | 18 | 18 | 17 | 19 | 0 |
| $\mathbf{e}^-$ : | 10 | 0 | 17 | 18 | 19 | 18 | 2 | 2 | 3 | 1 | 0 |
| | | | | | | | | | | | |
| $f_1$ : | | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| $f_2$ : | | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |
| $f_3$ : | | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| $f_4$ : | | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 |
| $f_5$ : | | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |

Fig. 15. Training data for Experiment 3. The event data are consistent with two categories: The first includes objects that prevent the machine from activating, and the second includes objects that activate the machine. Features $f_1$ through $f_5$ are ''family resemblance'' features that provide noisy information about the underlying categories.

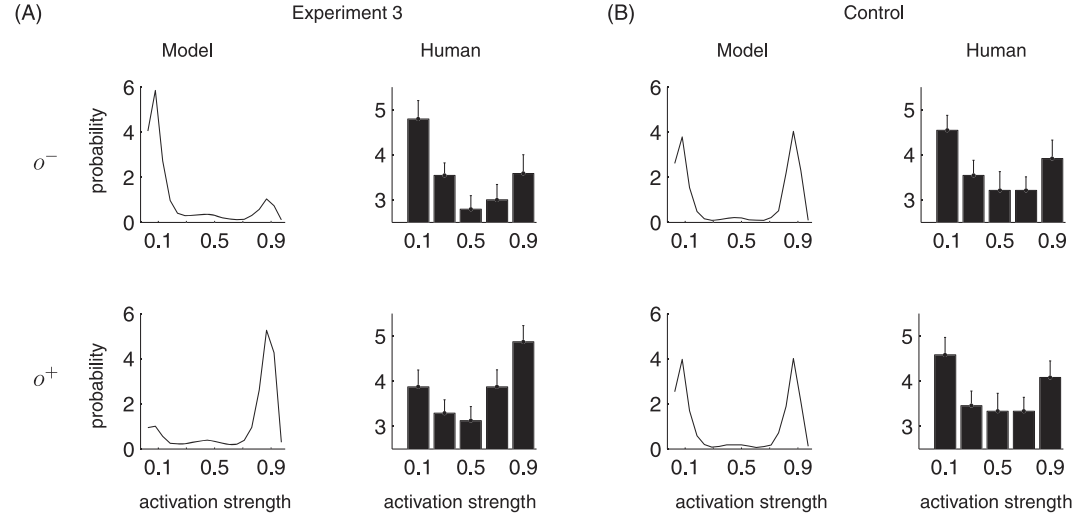


Fig. 16. Results for Experiment 3. (A) Model predictions and mean responses across 24 participants. Even though no trials are ever observed for objects $o^-$ and $o^+$, participants use the features of these objects to make predictions about their causal powers. (B) Model predictions and mean responses for a control task where all objects are perceptually identical.

### *7.3. Model predictions*

Predictions of the schema-learning model are shown in the left column of Fig. 16A. Each plot shows the probability that a test block will activate the machine on any given trial.[2] Both plots have two peaks, indicating that the model has discovered two categories but is not certain about the category assignments of the test blocks. The plots are skewed in opposite directions: based on the features of the test blocks, the model predicts that $o^-$ will activate the machine rarely, and that $o^+$ will activate the machine often. The left column of Fig. 16B shows predictions about a control task that is identical to Experiment 3 except that all blocks are perceptually identical. The curves are now symmetric, indicating that the model has no basis for assigning the test blocks to one category or another.

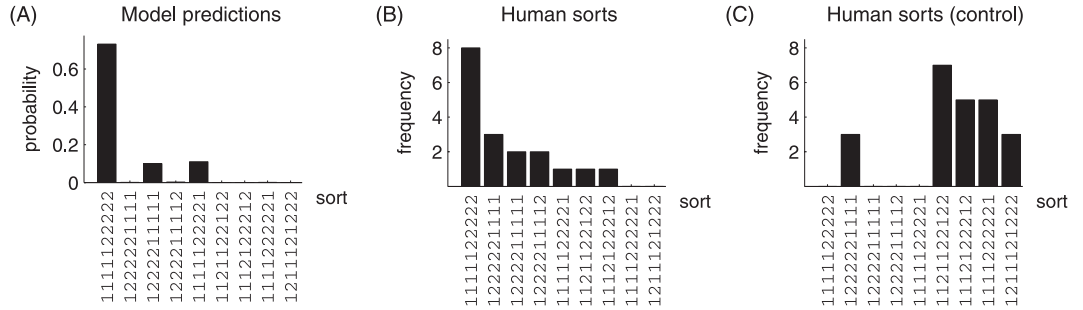


Fig. 17. Sorts for Experiment 3. (A) Relative probabilities of nine sorts according to the schema-learning model. Each sort is represented as a vector that specifies category assignments for the 10 objects in Fig. 15. The model prefers the family resemblance sort. (B) Sorts chosen by participants. Any sort not shown was chosen by at most one participant. (C) Sorts chosen in the control condition when no causal information was available.

Predictions about the sorting task are summarized in Fig. 17A. The top few sorts are included, and the most probable solution according to the model is the family resemblance sort. Although the model allows sorts with any number of categories (including one, three or more), the probabilities shown in Fig. 17A are calculated with respect to the class of all two-category solutions.

### *7.4. Results*

Mean responses for the two test blocks are shown in the right column of Fig. 16A. Both plots are U-shaped curves, suggesting that participants realize that some blocks activate the machine rarely and others activate the machine often, but that few blocks activate the machine half the time. As predicted, the curves are skewed in opposite directions, indicating that $o^+$ is considered more likely to activate the machine than $o^-$. We ran a two-factor ANOVA which compared ratings for the first (0–20) and last (80–100) intervals across the two test blocks. There is no main effect of interval [$F(1,23) = 0.056$, $p > .5$] or of test block [$F(1,23) = 1.50$, $p > .1$], but there is a significant interaction between interval and test block [$F(1,23) = 6.90$, $p < .05$]. Follow up paired-sample $t$ tests support the claim that both plots in Fig. 16 show skewed U-shaped curves. In the case of object $o^-$, the 0.1 bar is significantly greater than the 0.9 bar ($p < .05$, one-sided) and the difference between the 0.9 bar and the 0.5 bar is marginally significant ($p = .07$, one-sided). In the case of object $o^-$, the 0.9 bar is significantly greater than the 0.1 bar ($p < .05$, one-sided) and the difference between the 0.1 bar and the 0.5 bar is marginally significant ($p = .07$, one-sided).

We ran an additional 24 participants in a control task that was identical to Experiment 3 except that all blocks were perceptually identical. Mean responses are shown in the right column of Fig. 16B, and participants now generate U-shaped curves that are close to symmetric. The ANOVA analysis described in the previous paragraph now indicates that there is no main effect of interval or test block, and no significant interaction between interval and test block. Although both human curves in Fig. 16B are higher on the left than the right, paired-sample $t$ tests indicate that there is no significant difference between the 0.1 bar and

the 0.9 bar in either case ($p > .2$). The differences between the 0.9 bar and the 0.5 bar fail to reach significance in both cases, but the 0.1 bars are significantly greater than the 0.5 bars in both cases ($p < .05$, one-sided).

The U-shaped curves in Fig. 16A,B resolve a question left open by Experiment 1. Responses to the $f = \{0.1, 0.9\}$ condition of the first experiment did not indicate that participants had identified two categories, but the U-shaped curves in Fig. 16B suggest that participants recognized two categories of blocks. All of the blocks in Experiment 3 produce the effect sometimes, and the U-shaped curves suggest that participants can use probabilistic causal information to organize objects into categories. Two differences between Experiment 3 and the second condition of Experiment 1 seem particularly important. In Experiment 3, more blocks were observed for each category (4 rather than 3), and more trials were observed for each block (20 rather than 10). Experiment 3 therefore provides more statistical evidence that there are two categories of blocks.

Responses to the sorting task are summarized in Fig. 17B. The most popular sort organizes the blocks into the two family resemblance categories, and it is chosen by eight of the 24 participants. Studies of feature-based categorization have consistently found that family resemblance sorts are rare, and that participants prefer instead to sort objects according to a single dimension (e.g., size or color) (Medin, Wattenmaker, & Hampson, 1987). To confirm that this standard result applies in our case, we ran a control task where no causal observations were available and participants were asked to sort the blocks into categories on the basis of their perceptual features. The results in Fig. 17C show that none of 24 participants chose the family resemblance sort. A chi-square test confirms that the family resemblance sort was chosen significantly more often in the causal task than the control task ($p < .01$, one sided). Our results therefore suggest that the causal information provided in Experiment 3 overcomes the strong tendency to form categories based on a single perceptual dimension.

Regardless of the sort that they chose, most participants explained their response by stating that one category included ''strong activators'' or blocks that often lit up the machine, and that the other included weak activators. For example, one participant wrote that the first category ''activates approximately 10% of the time'' and the second category ''activates approximately 90% of the time.'' Although most participants seem to have realized that there were two qualitatively different kinds of blocks, only 13 of the 24 assigned the ''strong activators'' (blocks $o_1$ through $o_4$) to one category and the ''weak activators'' (blocks $o_5$ through $o_8$) to the other category. Some of the remaining participants may have deliberately chosen an alternative solution, but others gave explanations suggesting that they had lost track of the training trials. Note that the sorting task is relatively demanding, and that participants who do not organize the blocks carefully as they go along are likely to forget how many times each block activated the machine.

## 8. Discovering causal interactions between categories

Our approach so far captures some kinds of interactions between categories. For example, the schema in Fig. 1 captures interactions between categories of drugs and categories of

people—alpha blockers tend to produce headaches, but only in A-people. This schema, however, does not capture interactions between categories of drugs, and it makes no predictions about what might happen when alpha blockers and beta blockers are simultaneously ingested. Drugs may interact in surprising ways—for example, two drugs may produce a headache when combined even though each one is innocuous on its own. We now extend our model to handle cases of this kind where each event (e.g., ingestion) can involve varying numbers of objects (e.g., drugs).

The first step is to extend our notation for domain-level problems to allow sets of objects. The domain-level problem for the drugs and headaches example now becomes

$$\texttt{ingests}(\texttt{person}, \{\texttt{drug}\}) \xrightarrow{?} \texttt{headache}(\texttt{person})$$

where each cause event now specifies that a person ingests a set of drugs. Following Novick and Cheng (2004) we will decompose each cause event into subevents, one for each subset of the set of drugs. For example, the object-level problem

$$\texttt{ingests}(\texttt{Alice}, \{\texttt{Doxazosin}, \texttt{Acebutolol}\}) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{9}$$

can be viewed as a combination of four subproblems

$$\texttt{ingests}(\texttt{Alice}, []) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{10a}$$

$$\texttt{ingests}(\texttt{Alice}, [\texttt{Doxazosin}]) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{10b}$$

$$\texttt{ingests}(\texttt{Alice}, [\texttt{Acebutolol}]) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{10c}$$

$$\texttt{ingests}(\texttt{Alice}, [\texttt{Doxazosin}, \texttt{Acebutolol}]) \xrightarrow{?} \texttt{headache}(\texttt{Alice}) \tag{10d}$$

The difference between the curly brackets in Eq. (9) and the square brackets in Eq. (10d) is significant. The subproblem in Eq. (10d) refers to a causal relationship that depends exclusively on the interaction between Doxazosin and Acebutolol. In other words, the properties of this causal relationship depend only on the causal power of the pair of drugs, not on the causal power of either drug taken in isolation. The problem in Eq. (9) refers to the overall relationship that results from combining all instances in Eqs. (10a–d). In other words, the overall effect of taking Doxazosin and Acebutolol may depend on the base rate of experiencing headaches, the effect of Doxazosin alone, the effect of Acebutolol alone, and the effect of combining the two drugs.

Building on the approach described in previous sections, we introduce a causal model for each subproblem that depends on three parameters. The first indicates whether the subevent is causally related to the effect, the second indicates the polarity of this causal relationship, and the third indicates the strength of this relationship. As before, we organize the drugs into categories, and we assume that object-level causal models are generated from category-level models that capture the causal powers of each category acting in isolation. Now, however,

we introduce additional category-level models that capture interactions between categories. For instance, if Acebutolol and Atenolol are assigned to the same category, then the causal models for the subproblems

$$\texttt{ingests}(\texttt{Alice}, [\texttt{Doxazosin}, \texttt{Acebutolol}]) \xrightarrow{?} \texttt{headache}(\texttt{Alice})$$

$$\texttt{ingests}(\texttt{Alice}, [\texttt{Doxazosin}, \texttt{Atenolol}]) \xrightarrow{?} \texttt{headache}(\texttt{Alice})$$

will be generated from the same category-level model. This approach captures the intuition that members of the same category (e.g., Acebutolol and Atenolol) are expected to interact with Doxazosin in a similar way.

To formalize these ideas, we extend the $\Psi$ in Eq. 6 to include an arrow $a$, a polarity $g$ and a strength $s$ for each combination of objects. We extend the schema in a similar fashion and include category-level models for each combination of categories. As before, the parameters for each object-level causal model are generated from the parameters ($\bar{a}$, $\bar{g}$, and $\bar{s}$) for the corresponding category-level model. For instance, Fig. 18 shows how the causal model for the $o_9$ + $o_{18}$ pair is generated from a category-level model that states that categories $c_A$ and $c_B$ interact to generate the effect.

Our main remaining task is to specify how the object-level models for the subinstances in 10 combine to influence the probability that Alice develops a headache after ingesting Doxazosin and Acebutolol. We use the sequential interaction model of Novick and Cheng (2004) and assume that subevents combine according to a network of noisy-OR and noisy-AND-NOT gates (Fig. 19). To capture the idea that the causal powers of a set of objects can be very different from the causal powers of the objects taken individually, we assume that subevents involving small sets of objects {e.g., ingests(Alice, [Doxazosin])} act first and can be overruled by subevents involving larger sets {e.g., ingests(Alice, [Doxazosin, Atenolol])}. Although the sequential interaction model seems appropriate for our purposes, the

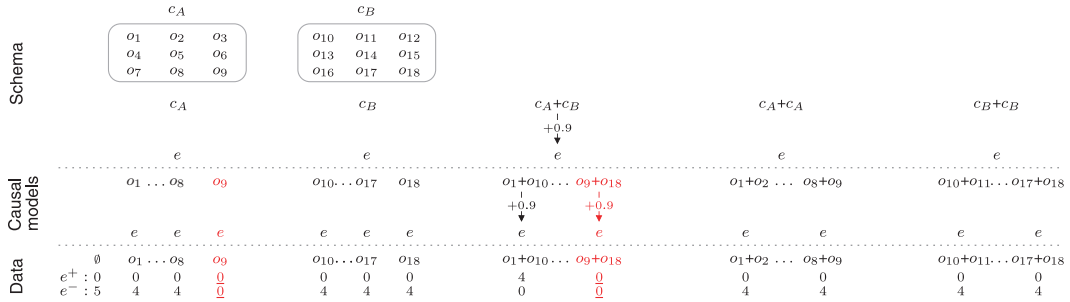


Fig. 18. Learning about interactions between objects. The schema includes category-level models for each individual category and for each pair of categories. The schema shown here has two categories: Individual objects of either category do not produce the effect, but any pair including objects from both categories will produce the effect. The collection of object-level causal models includes a model for each object and each pair of objects. Note that the schema supports inferences about sparsely observed individual objects (e.g., $o_9$) and about pairs that have never been observed to interact (e.g., $o_9$ and $o_{18}$, counts underlined in red).

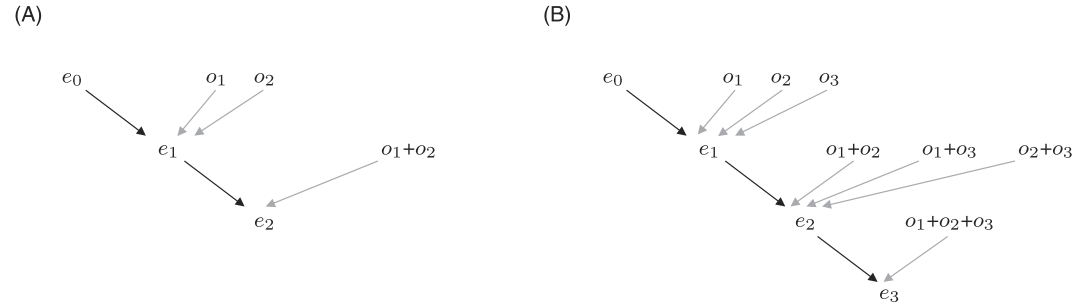


Fig. 19. The sequential interaction model of Novick and Cheng (2004). (A) Network for the case where two objects may interact to influence an effect event $e$. Event $e_i$ indicates whether the effect would have occurred based on interactions of up to $i$ objects. For example, $e_o$ indicates whether the background cause is active, and $e_1$ indicates whether the effect would have occurred as a result of combining the background cause with the causal contributions of each object taken individually. Variable $e_2$ indicates whether the effect event actually occurs when the interaction between the two objects is taken into account. The black arrows are generative with strength 1, and the gray arrows may or may not exist, may be generative or preventive, and may have any causal strength. Once the status of each gray arrow is specified, all events in the network combine according to a noisy-OR/noisy-AND-NOT model. (B) The same general approach can handle interactions among any number of cause events. Shown here is a case where three objects may interact to influence an effect event. In this case, variable $e_3$ indicates whether the effect event actually occurs.

general framework we have developed allows room for accounts of schema learning that incorporate alternative models of interaction.

## 9. Experiment 4: Causal interactions between categories

We designed an experiment to explore schema learning in a setting where pairs of objects may interact to produce a cause. Our formal framework can now handle several kinds of data, including contingency data for single objects, contingency data for pairs of objects, and perceptual features of the objects. In real-world settings, these different kinds of data will often reinforce each other and combine to pick out a single set of categories. Here, however, we explore whether information about pairwise interactions alone is sufficient for learners to discover causal schemata.

Experiment 4 used the same scenario developed for our previous experiments, but now participants were able to place up to two blocks inside the machine. Unlike Experiments 1 through 3, the individual causal powers of the blocks were identical, and unlike Experiment 3, the blocks were perceptually indistinguishable. The blocks, however, belonged to categories, and these categories determined the pairwise interactions between blocks. In the *pairwise activation* condition, the machine never activated when it contained a single block or two blocks from the same category, but always activated whenever it contained one block from each category (Fig. 18). In the *pairwise inhibition* condition the machine always activated when it contained a single block or two blocks from the same category, but never activated when it contained one block from each category. Experiment 4 explores whether

participants could infer the underlying causal categories based on pairwise interactions alone and could use this knowledge to rapidly learn causal models for novel objects.

Our experiment builds on the work of Kemp, Tenenbaum, Niyogi, and Griffiths (2010), who demonstrated that people can use relationships between objects to organize these objects into categories.[3] These authors considered interactions between objects, but their stimuli did not allow for the possibility that individual objects might produce the effect in isolation. We therefore designed a new experiment that relies on the same scenario used in Experiments 1 through 3.

### *9.1. Participants*

Thirty-two members of the CMU community participated for pay or course credit.

### *9.2. Stimuli and design*

Experiment 4 used the same graphical interface developed for Experiment 1. All of the blocks were perceptually indistinguishable. The experiment included two conditions and sixteen participants were assigned to each condition. In the pairwise activation condition, the machine never activated on its own and never activated when it contained a single block. The blocks, however, belonged to two categories, and the machine always activated on trials when it contained an A-block and a B-block. In the pairwise inhibition condition, the machine always activated when it contained a single block or two blocks from the same category, but it always failed to activate when it contained two blocks from different categories.

### *9.3. Procedure*

The experiment was divided into several phases. During phase 0, participants observed five trials where the empty machine failed to activate. Three blocks were added to the screen at the start of phase 1. Unknown to the participants, two blocks were A-blocks ($o_1$ and $o_2$) and the third was a B-block ($o_{10}$). Participants observed four trials for each individual block, and the machine never activated (pairwise activation condition) or always activated (pairwise inhibition condition). Before observing any interactions, participants predicted what would happen when $o_1$ and $o_{10}$ were simultaneously placed in the machine. The wording of the question was taken from our previous experiments: Participants imagined 100 trials when the machine contained the two blocks, and they rated the probability that the total number of activations would fall within each of five intervals. Participants then saw two trials for each pair of blocks. Phase 1 finished with a period of ‘‘free experimentation,’’ where participants were given the opportunity to carry out as many of their own trials as they wished.

Phases 2 through 6 were identical in structure. In each phase, three new blocks were added to the screen, and one of these blocks served as the ‘‘test block.’’ In some phases the test block was an A-block, and in others the test block was a B-block. Before observing any trials involving the new blocks, participants were given a pretest which required them to

Table 1
Design for Experiment 4

| Phase | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Blocks added | $o_1$, $o_2$, $o_{10}$ | $o_3$, $o_{11}$, $o_{12}$ | $o_4$, $o_{13}$, $o_{14}$ | $o_5$, $o_6$, $o_{15}$ | $o_7$, $o_8$, $o_{16}$ | $o_9$, $o_{17}$, $o_{18}$ | $o_A$, $o_B$ |
| Test blocks | | $o_{11}$ | $o_4$ | $o_5$ | $o_{16}$ | $o_{17}$ | $o_A$, $o_B$ |
| Probe blocks (i) | | $o_2$ | $o_{12}$ | $o_3$ | $o_{15}$ | $o_8$ | $o_{11}$, $o_4$ |
| Probe blocks (ii) | | $o_2$ | $o_{12}$ | $o_3$ | $o_{15}$ | $o_8$ | $o_4$, $o_{11}$ |
| Probe blocks (iii) | | $o_2$ | $o_{12}$ | $o_3$ | $o_6$ | $o_{16}$ | $o_{11}$, $o_4$ |
| Probe blocks (iv) | | $o_2$ | $o_{12}$ | $o_3$ | $o_6$ | $o_{16}$ | $o_4$, $o_{11}$ |

*Note*. Blocks $o_1$ through $o_9$ belong to category $c_A$ and blocks $o_{10}$ through $o_{18}$ belong to category $c_B$. In each pretest and posttest, participants make predictions about interactions between the test block and $o_1$ (an A-block) and between the test block and $o_{10}$ (a B-block). Between each pretest and posttest, participants observe a single trial where the test block is paired with a probe block. Probe blocks for the four groups of participants are shown.

predict how the test block would interact with two of the blocks already on screen, one ($o_1$) from category $c_A$ and the other ($o_{10}$) from category $c_B$. Participants then observed a single trial where the test block was paired with one of the blocks already on screen (the probe block). Armed with this single piece of information, participants completed a posttest that was identical to the pretest. The phase then finished with a period of free experimentation.

A complete specification of the design is shown in Table 1. The experiment involves 20 blocks in total: blocks $o_1$ through $o_9$ belong to category $c_A$, blocks $o_{10}$ through $o_{18}$ belong to category $c_B$, and there are two test blocks in the final phase ($o_A$ and $o_B$). The first and second rows of Table 1 list the blocks that are added to the screen in each phase, and the block that serves as the test block in each phase. Participants were randomly assigned to one of four groups (i through iv), and the probe blocks used for each group are shown in the final four rows of Table 1. No significant differences between these groups were observed, and we will collapse across these groups when reporting our results.

The final phase was very similar to phases 2 through 6, but only two new blocks were added to the screen. One block ($o_A$) was an A-block, and the second ($o_B$) was a B-block. In phases 2 through 6 only one of the new blocks served as the test block, but in the final phase both $o_A$ and $o_B$ served as test blocks. In the pretest for phase 7, participants made predictions about how $o_A$ and $o_B$ would interact with $o_1$ and $o_{10}$ before observing any pairwise trials involving the test blocks. Participants then observed a single trial involving each test block and responded to a posttest that was identical to the pretest. After providing these predictions, participants were asked to sort the blocks into two categories ''according to their effect on the machine'' and to ''describe how the blocks and machine work.''

### *9.4. Model predictions*

Although participants made inferences during each phase of the experiment, our main question is whether they had learned a causal schema by the end of the experiment. We therefore compare inferences about the first and last phases of the experiment. Kemp et al. (2010) describe a very similar task and show learning curves that include data from all phases of the experiment.

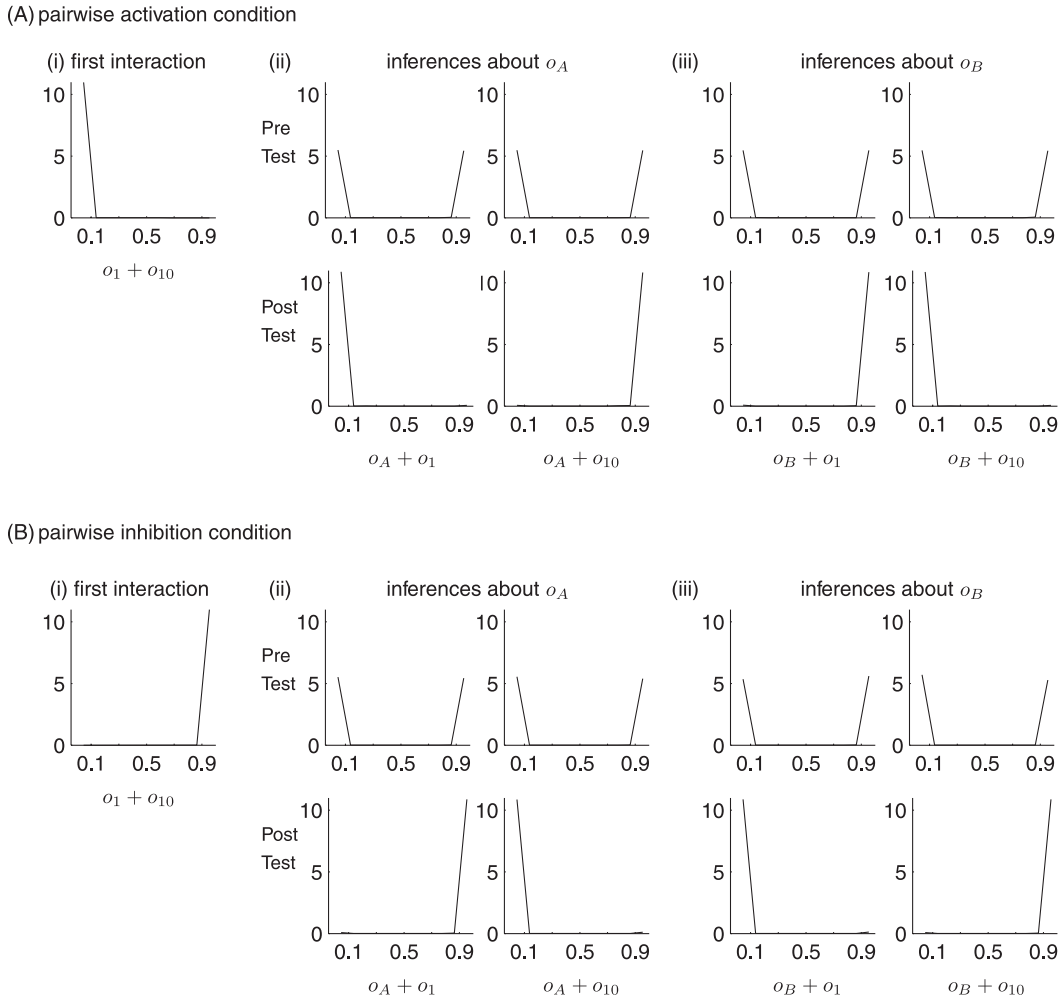


Fig. 20. Model predictions for Experiment 4. (A) Pairwise activation condition. (i) Before any pairwise trials have been observed, the model predicts that pairs of objects are unlikely to activate the machine. (ii) Inferences about test block $o_A$. Before observing any trials involving this block, the model is uncertain about whether it will activate the machine when paired with $o_1$ or $o_{10}$. After observing that $o_A$ activates the machine when paired with $o_{18}$ (a B-block), the model infers that $o_A$ will activate the machine when paired with $o_{10}$ but not $o_1$. (iii) Inferences about test block $o_B$ show a similar pattern: The model is uncertain during the pretest, but one observation involving $o_B$ is enough for it to make confident predictions on the posttest. (B) Pairwise inhibition condition. The prediction in (i) and the posttest predictions in (ii) and (iii) are the opposite of the corresponding predictions for the pairwise activation condition.

Figs. 20A.i, B.i show predictions about a pair of blocks before any pairwise trials have been observed. In the pairwise activation condition, the model has learned by this stage that individual blocks tend not to produce the effect, and the default expectation captured by the interaction model is that pairs of blocks will also fail to produce the effect. The model

allows for several possibilities: There may or may not be a conjunctive cause corresponding to any given pair of blocks, and this conjunctive cause (if it exists) may be generative or preventive and may have high or low strength. Most of these possibilities lead to the prediction that the pair of blocks will be unlikely to activate the machine. The machine is only likely to activate if the pair of blocks corresponds to a conjunctive cause with high strength, and this possibility receives a relatively low probability compared to the combined probability assigned to all other possibilities. Similarly, in the pairwise inhibition condition the model has learned that individual blocks tend to produce the effect, and the default expectation captured by the interaction model is that pairs of blocks will also produce the effect.

After observing several pairwise interactions, the model discovers that the default expectation does not apply in all cases, and that some pairs of blocks activate the machine when combined. By the final phase of the task, the model is confident that the blocks can be organized into two categories, where blocks $o_1$ through $o_9$ belong to category $c_A$ and blocks $o_{10}$ through $o_{18}$ belong to category $c_B$. The model, however, is initially uncertain about the category assignments of the two test blocks (blocks $o_A$ and $o_B$) and cannot predict with confidence whether either block will activate the machine when paired with $o_1$ or $o_{10}$ (Fig. 20ii–iii). Recall that the two categories have no distinguishing features, and that blocks $o_A$ and $o_B$ cannot be categorized before observing how they interact with one or more previous blocks. After observing a single trial where $o_A$ is paired with one of the previous blocks, the model infers that $o_A$ probably belongs to category $A$. In the pairwise activation condition, the model therefore predicts that the pair $\{o_A, o_{10}\}$ will probably activate the machine but that the pair $\{o_A, o_1\}$ will not (Fig. 20A.ii–iii). Similarly, in the pairwise activation condition, a single trial involving $o_B$ is enough for the model to infer that $\{o_B, o_1\}$ will probably activate the machine although the pair $\{o_B, o_{10}\}$ will not.

### *9.5. Results*

Figs. 21A.i, B.i show mean inferences about a pairwise interaction before any pairwise trials have been observed. As expected, participants infer that two blocks which fail to activate the machine individually will fail to activate the machine when combined (pairwise activation condition), and that two blocks which individually activate the machine will activate the machine when combined (pairwise inhibition condition). A pair of $t$ tests indicates that the 0.1 bar is significantly greater than the 0.9 bar in Fig. 21A.i ($p < .001$, one-sided) but that the 0.9 bar is significantly greater than the 0.1 bar in Fig. 21B.i ($p < .001$, one-sided). These findings are consistent with the idea that learners assume by default that multiple causes will act independently of one another.

By the end of the experiment, participants were able to use a single trial involving a novel block to infer how this block would interact with other previously observed blocks. The mean responses in Fig. 21 match the predictions of our model and show that one-shot learning is possible even in a setting where any two blocks taken in isolation appear to have identical causal powers. A series of paired-sample $t$ tests indicates that the difference between the 0.1 and the 0.9 bars is not significant for any of the pretest plots in Fig. 21 ($p > .3$ in all

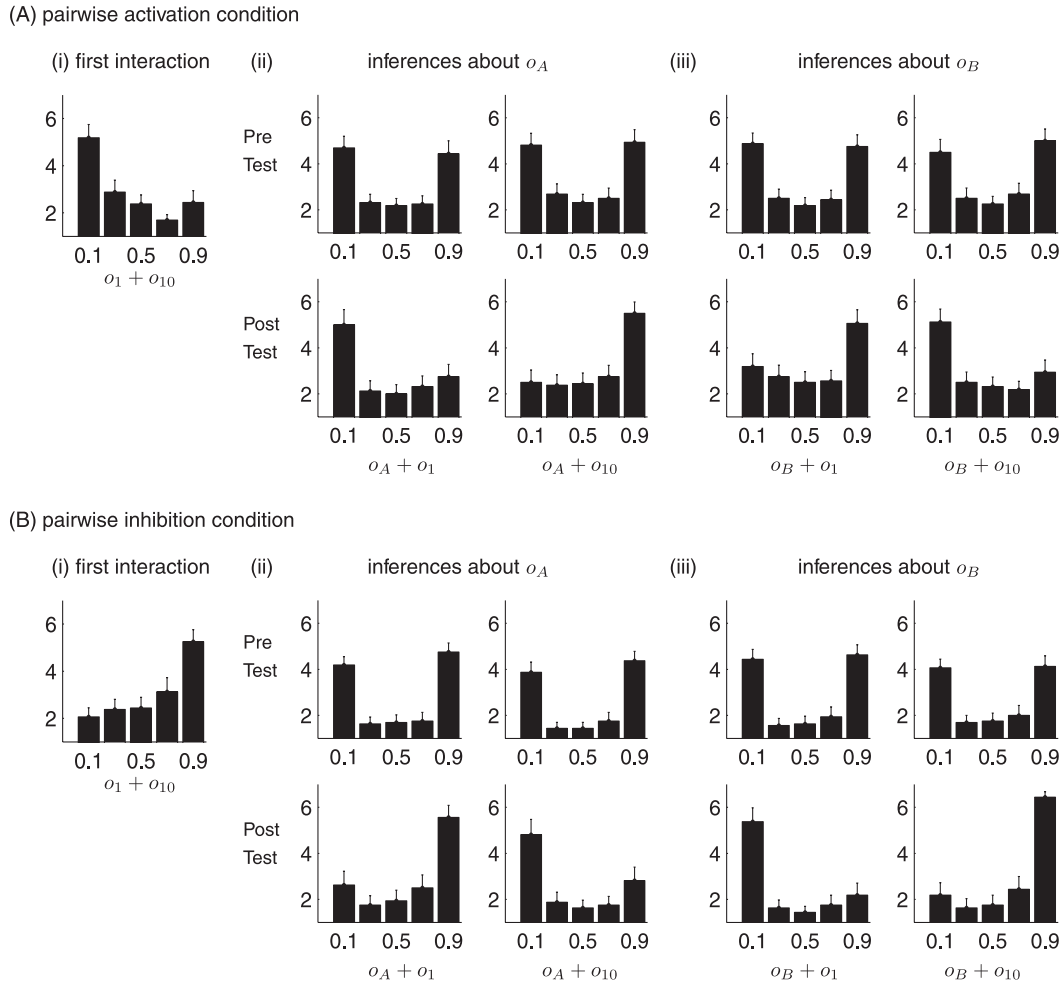


Fig. 21. Data for Experiment 4. All inferences are qualitatively similar to the model predictions in Fig. 20.

cases), but the difference between these bars is significant for each posttest plot ($p < .05$ in all cases). Although the model predictions are broadly consistent with our data, the model is often extremely confident in cases where the mean human response appears to be a U-shaped curve. In all of these cases, however, few individuals generate U-shaped curves, and the U-shaped mean is a consequence of averaging over a majority of individuals who match the model and a minority who generate curves that are skewed in the opposite direction.

Responses to the sorting task provided further evidence that participants were able to discover a causal schema based on interaction data alone. In each condition, the most common sort organized the 18 blocks into the two underlying categories. In the pairwise activation condition, five of the 16 participants chose this response, and an additional three gave responses that were within three moves of this solution. In the pairwise inhibition condition,

nine of the 16 participants chose this response, and an additional two gave responses that were within three moves of this solution. The remaining sorts appeared to vary idiosyncratically, and no sort other than the most common response was chosen by more than one participant. As in Experiment 3, the sorting task is relatively challenging, and participants who did not organize the blocks as they went found it difficult to sort them into two categories at the end of the experiment. Several participants gave explanations suggesting that they had lost track of the observations they had seen.

Other explanations, however, suggested that some participants had discovered an explicit causal schema. In the pairwise activation condition, one participant sorted the blocks into categories that she called ''activators'' and ''partners,'' and wrote that ''the machine requires both an activator and a partner to work.'' In the pairwise inhibition condition, one participant wrote the following:

> The machine appears to take two different types of blocks. Any individual block turns on the machine, and any pair of blocks from the same group turns on the machine. Pairing blocks from different groups does not turn on the machine.

An approach similar to the exemplar model described earlier will account for people's inferences about test blocks $o_A$ and $o_B$. For example, if $o_A$ is observed to activate $o_{18}$ in the pairwise activation condition, the exemplar model will assume that $o_A$ is similar to other blocks that have previously activated $o_{18}$, and will therefore activate $o_{11}$ but not $o_1$. Note, however, that the exemplar model assumes that learners have access to the observations made for all previous blocks, and we propose that this information can only be maintained if learners choose to sort the blocks into categories. The exemplar model also fails to explain the results of the sorting task, and the explanations that mention an underlying set of categories. Finally, Experiment 2 of Kemp et al. (2010) considers causal interactions, and it was specifically designed to compare approaches like the exemplar model with approaches that discover categories. The results of this experiment rule out the exemplar model, but they are consistent with the predictions of our schema-learning framework.

## 10. Children's causal knowledge and its development

We proposed that humans learn to learn causal models by acquiring abstract causal schemata, and our experiments confirm that adults are able to learn and use abstract causal knowledge. Some of the most fundamental causal schemata, however, are probably acquired early in childhood, and learning abstract schemata may itself be a key component of cognitive development. Although our experiments focused on adult learning, this section shows how our approach helps to account for children's causal learning.

Our experiments explored three learning challenges: grouping objects into categories with similar causal powers (Fig. 6 and Experiments 1 and 2), categorizing objects based on their causal powers and their perceptual features (Fig. 14 and Experiment 3), and forming categories to explain causal interactions between objects (Fig. 18 and Experiment 4). All

three challenges have been explored in the developmental literature, and we consider each one in turn.

### *10.1. Categories and causal powers*

The developmental literature on causal learning includes many studies that address the relationship between categorization and causal reasoning. Researchers have explored whether children organize objects into categories with similar causal powers, and whether their inferences rely more heavily on causal powers or perceptual features. Many studies that address these questions have used the blicket detector paradigm (Gopnik & Sobel, 2000; Nazzi & Gopnik, 2000; Sobel, Sommerville, Travers, Blumenthal, & Stoddard, 2009), and we will show how our model accounts for several results that have emerged from this paradigm.

In a typical blicket detector study, children are shown a set of blocks and a detector. Some blocks are *blickets* and will activate the detector if placed on top of it. Other blocks are inert and have no effect on the detector. Many questions can be asked using this setup, but for now we consider the case where all blocks are perceptually identical and the task is to organize these blocks into categories after observing their interactions with the detector. Gopnik and Sobel (2000) and others have established that young children can accurately infer whether a given block is a blicket given only a handful of relevant observations. For example, suppose that the detector activates when two blocks (A and B) are simultaneously placed on top of it, but fails to activate when A alone is placed on top of it. Given these outcomes, 3-year-olds correctly infer that block B must be a blicket.

Our formal approach captures many of the core ideas that motivated the original blicket detector studies, including the idea that objects have causal powers and the idea that objects with similar causal powers are organized into categories. Our work also formalizes the relationship between object categories (e.g., categories of blocks) and event data (e.g., observations of interactions between blocks and the blicket detector). In particular, we propose that children rely on an intermediate level of knowledge which specifies the causal powers of individual objects, and that they understand that the outcome of a causal event depends on the causal powers of the specific objects (e.g., blocks) involved in that event.

Several previous authors have presented Bayesian analyses of blicket-detector experiments (Gopnik, Glymour, Sobel, Schulz, Kushnir, & Danks, 2004), and it is generally accepted that the results of these experiments are consistent with a Bayesian approach. Typically, however, the Bayesian models considered do not incorporate all of the intuitions about causal kinds that are captured by our framework. A standard approach used by Gopnik et al. (2004) and others is to construct a Bayes net where there is a variable for each block indicating whether it is on the detector, an additional variable indicating whether the detector activates, and an arrow from each block variable to the detector variable only if that block is a blicket. This simple approach provides some insight but fails to capture key aspects of knowledge about the blicket detector setting. For example, if the experimenter introduces a new block and announces that it is a blicket, the network must be extended by adding a new variable that indicates whether the new block is on the detector and by draw-

ing an arrow between this new variable and the detector variable. Knowing how to modify the network in this way is critical, but this knowledge is not captured by the original network. More precisely, the original network does not explicitly capture the idea that blocks can be organized into categories, and that there is a predictable relationship between the category membership of a block and the outcome of events involving that block.

To address these limitations of a basic Bayes net approach, Danks (2007) and Griffiths and Tenenbaum (2007) proposed formalisms that explicitly rely on distinct causal models for blickets and nonblickets. Both of these approaches assume that all blickets have the same causal strength, but our model is more flexible and allows objects in the same category to have different causal strengths. For example, in the $p = \{0, 0.5\}$ condition of Experiment 1, block $o_6$ activates the machine 4 times out of 10 and block $o_7$ activates the machine 6 times out of 10. Our model infers that $o_7$ has a greater causal strength than $o_6$, and the means of the strength distributions for these blocks are 0.49 and 0.56, respectively. Although the blocks vary in strength, the model is 90% certain that the two belong to the same category. To our knowledge, there are no developmental experiments that directly test whether children understand that blocks in the same category can have different causal strengths. This prediction of our model, however, is supported by two existing results. Kushnir and Gopnik (2005) found that 4-year-olds track the causal strengths of individual blocks, and Gopnik, Sobel, Shulz, and Glymour (2001) found that 3-year-olds will categorize two objects as blickets even if one activates the machine more often (three of three trials) than the other (two of three trials). Combining these results, it seems likely that 4-year-olds will understand that two objects have different causal strengths but recognize that the two belong to the same category.

Although most blicket detector studies present children with only a single category of interest (i.e., blickets), our model makes an additional prediction that children should be able to reason about multiple categories. In particular, our model predicts that children will distinguish between categories of objects that have similar causal powers but very different causal strengths. Consider a setting, for example, where there are three kinds of objects: blickets, wugs, and inert blocks. Each blicket activates the detector 100% of the time, and each wug activates the detector between 20% and 30% of the time. Our model predicts that young children will understand the difference between blickets and wugs, and will be able to organize novel blocks into these categories after observing their effects on the detector.

### *10.2. Categories, causal powers, and features*

This section has focused so far on problems where the objects to be categorized are perceptually identical, but real-world object categories often vary in their perceptual properties as well as their causal powers. A central theme in the developmental literature is the relationship between perceptual categorization (i.e., categorization on the basis of perceptual properties) and conceptual or theory-based categorization (i.e., categorization on the basis of nonobservable causal or functional properties). Many researchers have compared these two kinds of categorization and have explored how the tradeoff between the two varies with age. One influential view proposes that infants initially form perceptual categories and only

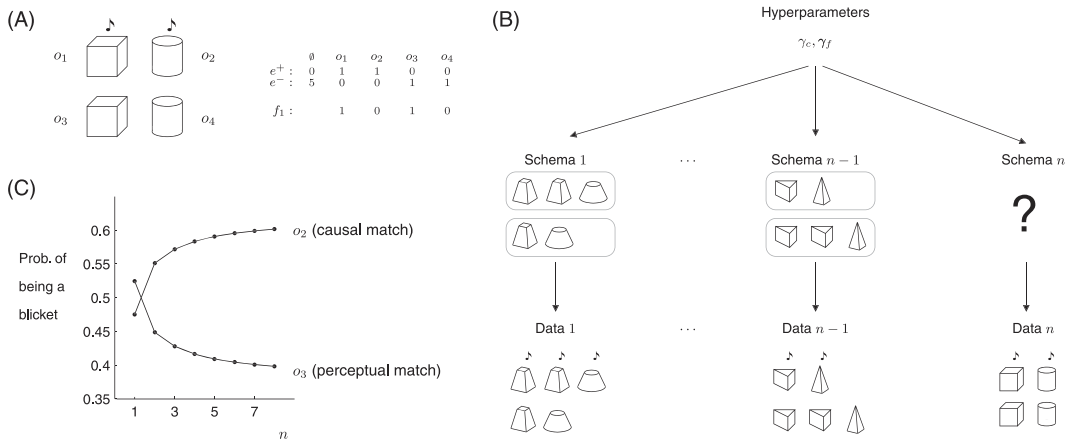


Fig. 22. Modeling the shift from perceptual to causal categorization. (A) The four objects in the Gopnik and Sobel (2000) conflict task. The two objects with the power to activate the blicket detector are marked with musical notes. Note that object $o_1$ could be grouped with a causal match ($o_2$) or a perceptual match ($o_3$). The table shows how the causal and perceptual data are provided as input to our model, and it includes a single feature $f_1$ which indicates whether the objects are cubes or cylinders. (B) Our hierarchical Bayesian framework can be extended to handle multiple systems of objects. Note that a single set of hyperparameters which specifies the relative weights of causal ($\gamma_c$) and perceptual ($\gamma_f$) information is shared across all systems. Our model observes how the objects in the first $n - 1$ systems are organized into categories, and it learns that in each case the categories are better aligned with the causal observations than the feature data. The model must now infer how the objects in the final system are organized into categories. (C) After learning that object $o_1$ in the final system is a blicket, the model infers whether $o_2$ and $o_3$ are likely to be blickets. Relative probabilities of these two outcomes are shown. The curves show a shift from perceptual categorization ($o_3$ preferred) to causal categorization ($o_2$ preferred).

later come to recognize categories that rely on nonobservable causal properties. Keil (1989) refers to this position as ‘‘Original Sim,’’ and he and others have explored its implications.

The blicket detector paradigm can be used to explore a simple version of the tradeoff between perceptual and causal categorization. Gopnik and Sobel (2000) considered a conflict task where the blocks to be categorized had different perceptual features, and where these perceptual features were not aligned with the causal powers of these blocks. One task used four blocks, where two blocks activated the blicket detector but two did not (Fig. 22A). Each block therefore had a causal match, and each block was also perceptually identical to exactly one other block in the set. Crucially, however, the perceptual match and the causal match for each block were different. Children were told that one of the blocks that activated the detector was a blicket and were asked to pick out the other blicket. Consistent with the ‘‘Original Sim’’ thesis, 2-year-olds preferred the perceptual match. Three- and four-year-olds relied more heavily on causal information and were equally likely to choose the perceptual and the causal match. A subsequent study by Nazzi and Gopnik (2000) used a similar task and found that 4.5-year-olds showed a small but reliable preference for the causal match. Taken together, these results provide evidence for a developmental shift from perceptual to causal categorization.

Unlike previous Bayes net models of blicket detector tasks, our approach can be applied to problems like the conflict task where causal information and perceptual information are both available. As demonstrated in our third experiment, a causal schema can specify information about the appearance and the causal powers of the members of a given category, and our schema learning model can exploit both kinds of information. In the conflict task of Gopnik and Sobel (2000), the inference made by our model will depend on the relative values of two hyperparameters: $\gamma_c$ and $\gamma_f$, which specify the extent to which the blocks in a given category are expected to have different causal powers ($\gamma_c$) and different features ($\gamma_f$). For modeling our adult experiments we set $\gamma_c$ to a smaller value than $\gamma_f$ ($\gamma_c = 0.1$ and $\gamma_f = 0.5$), which captures the idea that adults view causal information as a more reliable guide to category membership than perceptual information. As initially configured, our model therefore aims to capture causal knowledge at a stage after the perceptual to causal shift has occurred.

A natural next step is to embed our model in a framework where the hyperparameters $\gamma_c$ and $\gamma_f$ are learned from experience. The resulting approach is motivated by the idea that the developmental shift from perceptual to causal categorization may be explained in part as a consequence of rational statistical inference. Given exposure to many settings where causal information provides a more reliable guide to category membership than perceptual information, a child may learn to rely on causal information in future settings. To illustrate this idea, we describe a simple simulation based on the Gopnik and Sobel (2000) conflict task.

Fig. 22B shows how our schema learning framework can be extended to handle multiple systems of objects. We consider a simple setting where each system has two causal categories and up to six objects. Fig. 22B shows that the observations for the final test system are consistent with the Gopnik and Sobel (2000) conflict task: objects $o_1$ and $o_2$ activate the detector but the remaining objects do not, and object $o_1$ is perceptually identical to $o_3$ (both have feature $f_1$) but not $o_2$ or $o_4$. We assume that causal and feature data are available for each previous system, that the category assignments for each previous system are observed, and that these category assignments are always consistent with the causal data rather than the feature data. Two of these previous systems are shown in Fig. 22B.

Fig. 22B indicates that the category assignments for the test system are unobserved, and that the model must decide whether $o_1$ is more likely to be grouped with $o_2$ (the causal match) or $o_3$ (the perceptual match). If the test system is the first system observed (i.e., if $n = 1$), Fig. 22C shows that the model infers that the perceptual match ($o_3$) is more likely to be a blicket. Given experience with several systems, however, the model now infers that the causal match ($o_2$) is more likely to be a blicket.

The developmental shift in Fig. 22C is driven by the model's ability to learn appropriate values of the hyperparameters $\gamma_c$ and $\gamma_f$ given the first $n - 1$ systems of objects. The hierarchy in Fig. 22B indicates that a single pair of hyperparameters is assumed to characterize all systems, and the prior distribution used for each parameter is a uniform distribution over the set $\{2^{-6}, 2^{-5}, \ldots, 2^{3}\}$. Although the model begins with a symmetric prior over these hyperparameters, it initially prefers categories that match the features rather than the causal observations. The reason is captured by Fig. 3D, which indicates that the features are directly generated from the underlying categories but that the event data are one step removed from

these categories. The model assumes that causal powers rather than causal events are directly generated from the categories, and it recognizes that a small set of event data may not accurately reflect the causal powers of the objects involved. Given experience with several previous systems, however, the model infers that $\gamma_c$ is smaller than $\gamma_f$, and that causal observations are a more reliable guide to category membership than perceptual features. A similar kind of learning is discussed by Kemp et al. (2007), who describe a hierarchical Bayesian model that learns that shape tends to be a more reliable guide to category membership than other perceptual features such as texture and color.

The simulation results in Fig. 22C are based on a simple artificial scenario, and the proposal that statistical inference can help to explain the perceptual to conceptual shift needs to be explored in more naturalistic settings. Ultimately, however, this proposal may help to resolve a notable puzzle in the developmental literature. Many researchers have discussed the shift from perceptual to conceptual categorization, but Mandler (2004) writes that ''no one … has shown how generalization on the basis of physical appearance gets replaced by more theory-based generalization'' (p. 173). We have suggested that this shift might be explained as a consequence of learning to learn, and that hierarchical Bayesian models like the one we developed can help to explain how this kind of learning is achieved.

Although this section has focused on tradeoffs between perceptual and causal information, in many cases children rely on both kinds of information when organizing objects into categories. For example, children may learn that balloons and pins have characteristic features (e.g., balloons are round and pins are small and sharp) and that there is a causal relationship between these categories (pins can pop balloons). Children must also combine perceptual and causal information when acquiring the concept of animacy: Animate objects have characteristic features, including eyes (Jones, Smith & Landau, 1991), but they also share causal powers like the ability to initiate motion (Massey & Gelman, 1988). Understanding how concepts like animacy emerge over development is a challenging puzzle, but models that combine both causal and perceptual information may contribute to the solution.

### *10.3. Causal interactions*

Children make inferences about the causal powers of individual objects but also understand how these causal powers combine when multiple objects act simultaneously. The original blicket detector studies included demonstrations where multiple objects were placed on the detector, and 4-year-olds correctly assumed that these interactions were consistent with an OR function (i.e., that the detector would activate if one or more blickets were placed on top of it). Consistent with these results, our model assumes by default that causal interactions are governed by a noisy-OR function, but Experiment 4 demonstrates that both adults and our model are able to learn about other kinds of interactions. Lucas and Griffiths (2010) present additional evidence that adults can learn about a variety of different interactions, and future studies can test the prediction that this ability is available relatively early in development.

Our modeling approach relies on the idea that causal interactions between individual objects can be predicted using abstract laws that specify how categories of objects are expected to interact. Recent work of Schulz, Goodman, Tenenbaum, and Jenkins (2008)

supports the idea that young children can learn abstract laws, and they can do so on the basis of a relatively small number of observations. These authors introduced preschoolers to a set of seven blocks that included two red blocks, two yellow blocks, two blue blocks, and one white block. Some pairs of blocks produced a sound whenever they came into contact—for example, a train sound was produced whenever a red block and a blue block came into contact, and a siren sound was produced whenever a yellow block and a blue block came into contact (Fig. 23A). Other pairs of blocks produced no sound—for example, red blocks and yellow blocks never produced a sound when paired. Here we consider two conditions that differ only in the role played by the white block. In condition 1, the white block produced the train sound when paired with a red block, but in condition 2 the white block produced the train sound when paired with a blue block. No other observations involved the white block—in particular, children never observed the white block come into contact with a yellow block.

Using several dependent measures, Schulz and colleagues found that children in condition 1 expected the white block to produce the siren sound when paired with a yellow block, but that children in condition 2 did not. Our model accounts for this result. The evidence in condition 1 is consistent with the hypothesis that white blocks and blue blocks belong to the

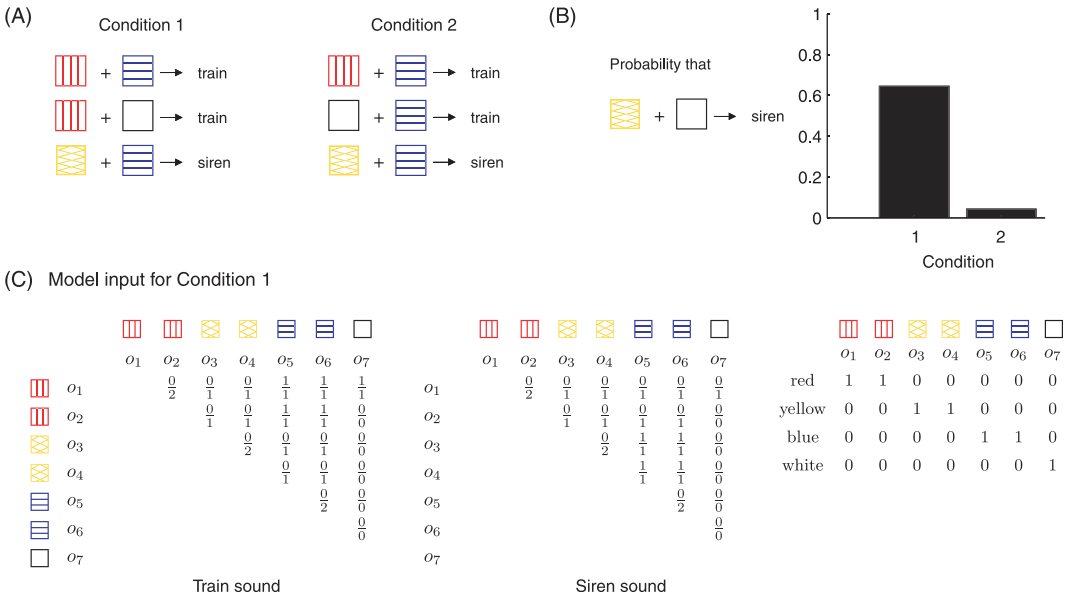


Fig. 23. (A) Evidence provided in conditions 1 and 2 of Schulz et al. (2008). (B) Model predictions about an interaction between a yellow block and a white block. Like preschoolers, the model infers that this combination is likely to produce a siren noise in condition 1 but not in condition 2. (C) Input data used to generate the model prediction for condition 1. Each entry in the first matrix shows the number of times that two blocks were touched and the number of times that the train sound was heard. For example, the red blocks came into contact twice, and the train sound was produced on neither trial. The second matrix specifies information about the siren sound, and the third matrix captures the perceptual features of the seven blocks. The input data for condition 2 are similar but not shown here.

same causal category—the category of WB blocks, say. Because the evidence suggests that yellow blocks produce the siren sound when paired with WB blocks, our model infers that the combination of a yellow block and a white block will probably produce the siren sound (Fig. 23B). In condition 2, however, the evidence supports the hypothesis that white blocks and red blocks belong to a category—the category of WR blocks. Because the evidence suggests that WR blocks and yellow blocks produce no sound when paired, the model infers that the combination of a yellow block and a white block will probably fail to produce the siren sound. The input data used to generate the model predictions for condition 1 are shown in Fig. 23C. The data include a matrix of observations for each effect (train sound and siren sound) and a matrix of perceptual features that specifies the color of each block.

The result in Fig. 23B follows directly from the observation that white blocks are just like blue blocks in condition 1, but that white blocks are just like red blocks in condition 2. This observation may seem simple, but Schulz and colleagues point out that it cannot be captured by the standard Bayes net approach to causal learning. The standard approach will learn a Bayes net defined over variables that represent events, such as a contact event involving a red block and a white block. The standard approach, however, has no basis for making predictions about novel events such as a contact event involving a yellow block and a white block. Our model overcomes this limitation by learning categories of objects and recognizing that the outcome of a novel event can be predicted given information about the category membership of the objects involved. The work of Schulz et al. suggests that young children are also able to learn causal categories from interaction data and to use these categories to make inferences about novel events.

We have now revisited three central themes addressed by our experiments—causal categorization, the tradeoff between causal and perceptual information, and causal interactions—and showed how each one is grounded in the literature on cognitive development. We described how our model can help to explain several empirical results, but future developmental experiments are needed to test our approach in more detail. Causal reasoning has received a great deal of attention from the developmental community in recent years, but there are still few studies that explore learning to learn. We hope that our approach will stimulate further work in this area, and we expect in turn that future empirical results will allow us to improve our approach as a model of children's learning.

## 11. Discussion

This paper developed a computational model that can handle multiple inductive tasks, and that learns rapidly about later tasks given experience with previous tasks from the same family. Our approach is motivated by the idea that learning to learn can be achieved by acquiring abstract knowledge that is relevant to all of the inductive tasks within a given family. A hierarchical Bayesian approach helps to explain how abstract knowledge can be learned after experience with the first few tasks in a family, and how this knowledge can guide subsequent learning. We illustrated this idea by developing a hierarchical Bayesian model of causal learning.

The model we described includes representations at several levels of abstraction. Near the top of the hierarchy is a schema that organizes objects into categories and specifies the causal powers and characteristic features of these categories. We showed that schemata of this kind support top-down learning and capture background knowledge that is useful when learning causal models for sparsely observed objects. Our model, however, also supports bottom-up learning, and we showed how causal schemata can be learned given perceptual features and contingency data.

Our experiments suggest that our model matches the abilities of human learners in several respects. Experiment 1 explored one-shot causal learning and suggested that people learn schemata which support confident inferences given very sparse data about a new object. Experiment 2 explored a case where people learn a causal model for an object that is qualitatively different from all previous objects. Strong inductive constraints are critical when data are sparse, but Experiment 2 showed that people (and our model) can overrule these constraints when necessary. Experiment 3 focused on ''zero-shot causal learning'' and showed that people make inferences about the causal powers of an object based purely on its perceptual features. Experiment 4 suggested that people form categories that are distinguished only by their causal interactions with other categories.

Our experiments used two general strategies to test the psychological reality of the hierarchy used by our model. One strategy focused on inferences at the bottom level of the hierarchy. Experiments 1, 3, and 4 considered one-shot or zero-shot causal learning and suggested that the upper levels of the model explain how people make confident inferences given very sparse data about a new object. A second strategy is to directly probe what people learn at the upper levels of the hierarchy. Experiments 3 and 4 asked participants to sort objects into categories, and the resulting sorts provide evidence about the representations captured by the schema level of our hierarchical model. A final strategy that we did not explore is to directly provide participants with information about the upper levels of the hierarchy, and to test whether this information guides subsequent inferences. Consider, for instance, the case of a science student who is told that ''pineapple juice is an acid, and acids turn litmus paper red.'' When participants are sensitive to abstract statements of this sort, we have additional evidence that their mental representations capture some of the same information as the hierarchies used by our model.

### *11.1. Related models*

Our work is related to three general areas that have been explored by previous researchers: causal learning, categorization, and learning to learn. This section compares our approach to some of the formal models that have been developed in each area.

#### *11.1.1.Learning to learn*

The hierarchical Bayesian approach provides a general framework for explaining learning to learn, and it has been explored by researchers from several communities. Statisticians and machine learning researchers have explored the theoretical properties of hierarchical Bayesian models (Baxter, 1998) and have applied them to challenging real-world problems (Blei,

Ng, & Jordan, 2003; Gelman, Carlin, Stern, & Rubin, 2003; Good, 1980). Psychologists have suggested that these models can help to explain human learning, and they have used them to explore how children learn to learn words (Kemp et al., 2007) and feature-based categories (Perfors & Tenenbaum, 2009).

Our work is motivated by many of the same general considerations as these previous approaches, but it represents one of the first attempts to explore learning to learn in a causal context. Our work also helps to demonstrate the flexibility of the hierarchical Bayesian approach to learning. Previous hierarchical approaches in the psychological literature often use hierarchies where the knowledge at the top level is very simple—for example, where this knowledge is captured by one or two parameters (Kemp et al., 2007). Our work illustrates that the same basic approach can explain how richer kinds of abstract knowledge can be acquired. We showed, for example, how causal schemata can be learned, where each schema is a system that includes causal categories along with category-level causal models that specify causal relationships between these categories.

#### *11.1.2. Causal learning*

Although there are few accounts of learning to learn in a causal setting, there are many previous models of causal learning and reasoning. Like many of these models (Gopnik & Glymour, 2002; Griffiths & Tenenbaum, 2005; Pearl, 2000), our work uses Bayesian networks to capture causal knowledge. For example, each object-level causal model in our framework is formalized as a causal Bayesian network. Note, however, that our approach depends critically on a level of representation that is more abstract than causal networks. We suggest that human inferences rely on causal schemata or systems of knowledge that capture expectations about object-level causal models.

#### *11.1.3. Categorization*

A causal schema groups a set of objects into categories, and our account of schema learning builds on two previous models of categorization. Our approach assumes that the category assignments of two objects will predict how they relate to each other, and the same basic assumption is made by the infinite relational model (Kemp et al., 2006), a probabilistic approach that organizes objects into categories that relate to each other in predictable ways. We also assume that objects belonging to the same category will tend to have similar features, and we formalize this assumption using the same probabilistic machinery that lies at the heart of Anderson's rational approach to categorization (Anderson, 1991). Our model can therefore be viewed as an approach that combines these two accounts of categorization with a Bayesian network account of causal reasoning. Because all of these accounts work with probabilities, it is straightforward to bring them together and create a single integrated framework for causal reasoning.

#### *11.1.4. Categorization and causal learning*

Previous authors have studied the relationship between categorization and causal reasoning (Waldmann & Hagmayer, 2006), and Lien and Cheng (2000) present a formal model that combines these two aspects of cognition. These authors consider a setting

similar to our third experiment where learners must combine contingency data and perceptual features to make inferences about sparsely observed objects. Their approach assumes that the objects of interest can be organized into one or more hierarchies, and that there are perceptual features which pick out each level in each hierarchy. Each perceptual feature is assumed to be a potential cause of effect $e$, and the *probabilistic contrast* for each cause $c$ with respect to the effect is $P(e^+ \mid c^+) - P(e^+ \mid c^-)$. Lien and Cheng suggest that the best explanation of the effect is the cause with maximal probabilistic contrast.

Although related to our own approach, the theoretical problem addressed by the principle of maximal contrast is different from the problem of discovering causal schemata. In our terms, Lien and Cheng assume that a learner already knows about several overlapping categories, where each category corresponds to a subtree of one of the hierarchies. They do not discuss how these categories might be discovered in the first place, but they provide a method for identifying the category that best explains a novel causal relation. We have focused on a different problem: Our schema-learning model does not assume that the underlying categories are known in advance, but it shows how a single set of nonoverlapping categories can be discovered.

Our work goes beyond the Lien and Cheng approach in several respects. Our model accounts for the results of Experiments 1, 2, and 4, which suggest that people organize perceptually identical objects into causal categories. In contrast, the Lien and Cheng model has no way to address problems where all objects are perceptually identical. In their own experiments, Lien and Cheng apply their model to several problems where causal information and perceptual features are both available, and where a subset of the perceptual features pick out the underlying causal categories. Experiment 3, however, exposes a second important difference between our model and their approach. Our model handles cases like Fig. 14 where the features provide a noisy indication of the underlying causal categories, but the Lien and Cheng approach can only handle causal categories that correlate perfectly with a perceptual feature. Experiment 3 supports our approach by demonstrating that people can discover categories in settings where perceptual features correlate roughly with the underlying categories, but where there is no single feature that perfectly distinguishes these categories.

Although the Lien and Cheng model will not account for the results of any of our experiments, it goes beyond our work in one important respect. Lien and Cheng suggest that potential causes can be organized into hierarchies—for example, ''eating cheese'' is an instance of ''eating dairy products'' which in turn is an instance of ''eating animal products.'' Different causal relationships are best described at different levels of these hierarchies—for example, a certain allergy might be caused by ''eating dairy products,'' and a vegan may feel sick at the thought of ''eating animal products.'' Our model does not incorporate the notion of a causal hierarchy—objects are grouped into categories, but these categories are not grouped into higher-level categories. As described in the next section, however, it should be possible to develop extensions of our approach where object-level causal models and features are generated over a hierarchy rather than a flat set of categories.

### *11.2. Learning and prior knowledge*

Any inductive learner must rely on prior knowledge of some kind and our model is no exception. This section highlights the prior knowledge assumed by our approach and discusses where this knowledge might come from. Understanding the knowledge assumed by our framework is especially important when considering its developmental implications. The ultimate goal should be to situate our approach in a developmental sequence that helps to explain the origin of each of its components, and we sketch some initial steps towards this goal.

The five shaded nodes in Fig. 3D capture much of the knowledge assumed by our approach. Consider first the nodes that represent domains (e.g., people) and events (e.g., `ingests`$(\cdot,\cdot)$). Domains can be viewed as categories in their own right, and these categories might emerge as the outcome of prior learning. For example, our approach could help to explain how a learner organizes the domain of physical objects into animate and inanimate objects, and how the domain of animate objects is organized into categories like people and animals. As these examples suggest, future extensions of our approach should work with hierarchies of categories and explore how these hierarchies are learned. It may be possible, for example, to develop a model that starts with a single, general category (e.g., physical objects) and that eventually develops a hierarchy which indicates that people are animate objects and that animate objects are physical objects. There are several probabilistic approaches that work with hierarchies of categories (Kemp, Griffiths, Stromsten, & Tenenbaum, 2004; Kemp & Tenenbaum, 2008; Schmidt, Kemp, & Tenenbaum, 2006), and it should be relatively straightforward to combine one of these approaches with our causal framework.

Although our model helps to explain how categories of objects are learned, it does not explain how categories of events might emerge. There are several probabilistic approaches that explore how event categories could be learned (Buchsbaum, Griffiths, Gopnik, & Baldwin, 2009; Goodman, Mansinghka, & Tenenbaum, 2007), and it may be possible to combine these approaches with our framework. Ultimately researchers should aim for models that can learn hierarchies of event categories—for example, touching is a kind of physical contact, and physical contact is a kind of event.

The third shaded node at the top of Fig. 3D represents a domain-level problem. Our framework takes this problem for granted but could potentially learn which problems capture possible causal relationships. Given a set of domains and events, the learner could consider a hypothesis space that includes all domain-level problems constructed from these elements, and the learner could identify the problems that seem most consistent with the available data. Different domain-level problems may make different assumptions about which events are causes and which are effects, and intervention data and temporal data are likely to be especially useful for resolving this issue: Effect events can be changed by intervening on cause effects, but not vice versa, and event effects usually occur some time after cause effects.

In many cases, however, the domain-level problem will not need to be learned from data, but will be generated by inheritance over a hierarchy of events and a hierarchy of domains.

For example, suppose that a learner has formulated a domain-level problem which recognizes that acting on a physical object can affect the state of that object:

$$\texttt{action}(\texttt{physical object 1}, \texttt{physical object 2}) \xrightarrow{?} \texttt{state}(\texttt{physical object 2})$$

If the learner knows that a touching is an action, that people and toys are both physical objects, and that emitting sound is a state, then the learner can use domain and event inheritance to formulate a domain-level problem which recognizes that humans can make toys emit sound by touching them:

$$\texttt{touch}(\texttt{person}, \texttt{toy}) \xrightarrow{?} \texttt{emits\_sound}(\texttt{toy})$$

A domain-level problem identifies a causal relationship that might exist, but additional evidence is needed to learn a model which specifies whether this relationship exists in reality. The distinction between domain-level problems and causal models is therefore directly analogous to the distinction between possibility statements (this toy could be made out of wood) and truth statements (this toy is actually made out of plastic). Previous authors have suggested that possibility statements are generated by inheritance over ontological hierarchies (Keil, 1979; Sommers, 1963), and that these hierarchies can be learned (Schmidt et al., 2006). Our suggestions about the origins of domain-level problems are consistent with these previous proposals.

The final two shaded nodes in Fig. 3D represent the event and feature data that are provided as input to our framework. Like most other models, our current framework takes these inputs for granted, but it is far from clear how a learner might convert raw sensory input into a collection of events and features. We can begin to address this question by adding an additional level at the bottom of our hierarchical Bayesian model. The information observed at this level might correspond to sensory primitives, and a learner given these observations might be able to identify the events and features that our current approach takes for granted. Goodman et al. (2007) and Austerweil and Griffiths (2009) describe probabilistic models that discover events and features given low-level perceptual primitives, and the same general approach could be combined with our framework.

Even if a learner can extract events and features from the flux of sensory experience, there is still the challenge of deciding which of these events and features are relevant to the problem at hand. We minimized this challenge in our experiments by exposing our participants to simple settings where the relevant features and events were obvious. Future analyses can consider problems where many features and events are available, some of which are consistent with an underlying causal schema, but most of which are noisy. Machine learning researchers have developed probabilistic methods for feature selection that learn a weight for each feature and are able to distinguish between features that carry useful information and those that are effectively random (George & McCulloch, 1993; Neal, 1996). It should be possible to combine these methods with our framework, and the resulting model may help to explain how children and adults extract causal information from settings that are noisy and complex.

We have now discussed how several components of the framework in Fig. 3D could be learned rather than specified in advance. Although our model could be extended in several directions, note that there are fundamental questions about the origins of causal knowledge that it does not address. For example, our model suggests how a schema learner might discover the schema that accounts best for a given domain, but it does not explain how a learner might develop the ability to think about schemata in the first place. Similarly, our model can learn about the causal powers of novel objects, but it does not explain how a precausal learner might develop the ability to think about causal powers. There are two possible solutions to developmental questions like these: Either concepts like causal schema and causal power could be innate, or one or both of these concepts could emerge as a consequence of early learning. Our work is compatible with both possible solutions, and future modeling efforts may help to suggest which of the two is closer to the truth.

## 12. Conclusion

We developed a hierarchical Bayesian framework that addresses the problem of learning to learn. Given experience with the causal powers of an initial set of objects, our framework helps to explain how learners rapidly learn causal models for subsequent objects from the same family. Our approach relies on the acquisition and use of causal schemata, or systems of abstract causal knowledge. A causal schema organizes a set of objects into categories and specifies the causal powers and characteristic features of each categories. Once acquired, these causal schemata support rapid top-down inferences about the causal powers of novel objects.

Although we focused on causal learning, the hierarchical Bayesian approach can help to explain learning to learn in other domains, including word learning, visual learning, and social learning. The hierarchical Bayesian approach accommodates both abstract knowledge and learning, and it provides a convenient framework for exploring two fundamental questions about cognitive development: how abstract knowledge is acquired, and how this knowledge is used to support subsequent learning. Answers to both questions should help to explain how learning accelerates over the course of cognitive development, and how this accelerated learning can bridge the gap between knowledge in infancy and adulthood.

## Notes

1. We will assume that $g$ and $s$ are defined even if $a = 0$ and there is no causal relationship between $o$ and $e$. When $a = 0$, $g$ and $s$ could be interpreted as the polarity and strength that the causal relationship between $o$ and $e$ would have if this relationship actually existed. Assuming that $g$ and $s$ are always defined, however, is primarily a mathematical convenience.
2. Unlike Experiment 1, the background rate is nonzero, and these probability distributions are not equivalent to distributions on the causal power of a test block.

3. In particular, the pairwise activation condition of Experiment 4 is closely related to the symmetric regular condition described by Kemp et al. (2010).

## Acknowledgments

An early version of this work was presented at the Twenty Ninth Annual Conference of the Cognitive Science Society. We thank Bobby Han for collecting the data for Experiment 4, and Art Markman and several reviewers for valuable suggestions. This research was supported by the William Asbjornsen Albert memorial fellowship (C. K.), the James S. McDonnell Foundation Causal Learning Collaborative Initiative (N. D. G., J. B. T), and the Paul E. Newton Chair (J. B. T.)

## References

Aldous, D. (1985). Exchangeability and related topics. In P. L. Hennequin (Ed.), *École d'Été de Probabilités de Saint-Flour, XIII—1983* (pp. 1–198). Berlin: Springer.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409–429.

Austerweil, J., & Griffiths, T. L. (2009). Analyzing human feature learning as nonparametric Bayesian inference. In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in neural information processing systems 21* (pp. 97–104).

Baxter, J. (1998). Theoretical models of learning to learn. In S. Thrun & L. Pratt (Eds.), *Learning to learn* (pp. 71–94). Norwell, MA: Kluwer.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.

Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Buchsbaum, D., Griffiths, T. L., Gopnik, A., & Baldwin, D. (2009). Learning from actions and their consequences: Inferring causal variables from continuous sequences of human action. In N. A. Taatgen & H. Van Rijn (Eds.), *Proceedings of the 31st annual conference of the Cognitive Science Society* (pp. 2493–2498). Austin, TX: Cognitive Science Society.

Caruana, R. (1997). Multitask learning. *Machine Learning*, *28*, 41–75.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367–405.

Danks, D. (2007). Theory unification and graphical models in human categorization. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation*. (pp. 173–189). Oxford, England: Oxford University Press.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). New York: Chapman & Hall.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In E. M. Keramida (Ed.), *Computing science and statistics: Proceedings of the 23rd symposium interface* (pp. 156–163). Fairfax Station, VA: Interface Foundation.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Good, I. J. (1980). Some history of the hierarchical Bayesian methodology. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian statistics* (pp. 489–519). Valencia, Spain: Valencia University Press.

Goodman, N. D., Mansinghka, V. K., & Tenenbaum, J. B. (2007). Learning grounded causal models. In D. S. Mc Namara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 305–310). Austin, TX: Cognitive Science Society.

Gopnik, A., & Glymour, C. (2002). Causal maps and Bayes nets: A cognitive and computational account of theory-formation. In P. Carruthers, S. Stich & M. Siegal (Eds.), *The cognitive basis of science* (pp. 117–132). Cambridge, England: Cambridge University Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.

Gopnik, A., & Sobel, D. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*, 1205–1222.

Gopnik, A., Sobel, D. M., Shulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two, three, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology*, *37*, 620–629.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 354–384.

Griffiths, T. L., & Tenenbaum, J. B. (2007). Two proposals for causal grammars. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 323–346). Oxford, England: Oxford University Press.

Harlow, H. F. (1949). The formation of learning sets. *Psychological Review*, *56*, 51–65.

Jain, S., & Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet Process mixture model. *Journal of Computational and Graphical Statistics*, *13*, 158–182.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*, 499–516.

Keil, F. C. (1979). *Semantic and conceptual development*. Cambridge, MA: Harvard University Press.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.

Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: perceiving the causes of behavior* (pp. 151–174). Morristown, NJ: General Learning Press.

Kemp, C. (2008). *The acquisition of inductive constraints*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Kemp, C., Griffiths, T. L., Stromsten, S., & Tenenbaum, J. B. (2004). Semi-supervised learning with trees. In S. Thrun, L. Saul & B. Schölkopt (Eds.), *Advances in neural information processing systems 16* (pp. 257–264). Cambridge, England, MA: MIT Press.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., & Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In Y. Gil & R. J. Mooney (Eds.), *Proceedings of the 21st national conference on artificial intelligence* (pp. 381–388). Menlo park, CA: AAAI Press.

Kemp, C., Tenenbaum, J. B., Niyogi, S., & Griffiths, T. L. (2010). A probabilistic model of theory formation. *Cognition*, *114*(2), 165–196.

Kushnir, T., & Gopnik, A. (2005). Children infer causal strength from probabilities and interventions. *Psychological Science*, *16*, 678–683.

Lagnado, D., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 856–876.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*, 87–137.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review*, *115*(4), 955–984.

Lucas, C. G., & Griffiths, T. L. (2010). Learning the form of causal relationships using hierarchical Bayesian models. *Cognitive Science 34*, 113–147.

Mandler, J. M. (2004). *The foundations of mind: origins of conceptual thought*. New York: Oxford University Press.

Massey, C., & Gelman, R. (1988). Preschoolers' ability to decide whether pictured unfamiliar objects can move themselves. *Developmental Psychology*, *24*, 307–317.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness and category construction. *Cognitive Psychology*, *19*, 242–279.

Nazzi, T., & Gopnik, A. (2000). A shift in children's use of perceptual and causal cues to categorization. *Developmental Science*, *3*(4), 389–396.

Neal, R. M. (1996). *Bayesian learning for neural networks (No. 118)*. New York: Springer-Verlag.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal inference. *Psychological Review*, *111*, 455–485.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

Perfors, A. F., & Tenenbaum, J. B. (2009). Learning to learn categories. In N. A. Taatqer & H. Van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 136–141). Austin, TX: Cognitive Science Society.

Sakamoto, Y., & Love, B. C. (2004). Schematic influences on category learning and recognition memory. *Journal of Experimental Psychology: General*, *133*(4), 534–553.

Schmidt, L. A., Kemp, C., & Tenenbaum, J. B. (2006). Nonsense and sensibility: Discovering unseen possibilities. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 744–749). Mahwah, NJ: Erlbaum.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162–176.

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. (2008). Going beyond the evidence: abstract laws and preschoolers' responses to anomalous data. *Cognition*, *109*(2), 211–223.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(4), 405–415.

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.

Sobel, D. M., Sommerville, J. A., Travers, L. V., Blumenthal, E. J., & Stoddard, E. (2009). The role of probability and intentionality in preschoolers' causal generalizations. *Journal of Cognition and Development*, *10*(4), 262–284.

Sommers, F. (1963). Types and ontology. *Philosophical Review*, *72*, 327–363.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*, 431–445.

Stevenson, H. W. (1972). *Children's learning*. New York: Appleton-Century-Crofts.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E. J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453–489.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Science*, *10*(7), 309–318.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261.

Thrun, S. (1998). Lifelong learning algorithms. In S. Thrun & L. Pratt (Eds.), *Learning to learn* (pp. 181–209). Norwell, MA: Kluwer.

Thrun, S., & Pratt, L. (Eds.) (1998). *Learning to learn*. Norwell, MA: Kluwer.
Waldmann, M. R., & Hagmayer, Y. (2006). Categories and causality: The neglected direction. *Cognitive Psychology*, *53*, 27–58.
Yerkes, R. M. (1943). *Chimpanzees: A laboratory colony*. New Haven, CT: Yale University Press.

## Appendix: A schema learning model

This appendix describes some of the mathematical details needed to specify our schema-learning framework in full.

### *Learning a single object-level causal model*

Consider first the problem of learning a causal model that captures the relationship between a cause event and an effect event. We characterize this relationship using four parameters. Parameters $a$, $g$, and $s$ indicate whether a causal relationship exists, whether it is generative, and the strength of this relationship. We assume that there is a generative background cause of strength $b$.

We place uniform priors on $a$ and $g$, and we assume that the strength parameter $s$ is drawn from a logistic normal distribution:

$$\begin{aligned} \text{logit}(s) &\sim \mathcal{N}(\mu, \sigma^2) \\ \mu &\sim \mathcal{N}(\eta, \tau\sigma^2) \\ \sigma^2 &\sim \text{Inv-gamma}(\alpha, \beta) \end{aligned} \tag{11}$$

The priors on $\mu$ and $\sigma^2$ are chosen to be conjugate to the Gaussian distribution on logit($s$), and we set $\alpha = 2$, $\beta = 0.3$, $\eta = 1$, and $\tau = 10$. The background strength $b$ is drawn from the same distribution as $s$, and all hyperparameters are set to the same values except for $\eta$ which is set to $-1$. Setting $\eta$ to these different values encourages $b$ to be small and $s$ to be large, which matches standard expectations about the likely values of these variables (Lu et al., 2008). As for all other hyperparameters in our model, $\alpha = 2$, $\beta = 0.3$, $\eta = 1$, and $\tau = 10$ were not tuned to fit our experimental results but were assigned to values that seemed plausible a priori. We expect that the qualitative predictions of our model are relatively insensitive to the precise values of these hyperparameters provided that they capture the expectation that $b$ should be small and $s$ should be large.

### *Learning multiple object-level causal models*

Consider now the problem of simultaneously learning multiple object-level models. The example in Fig. 1A includes two sets of objects (people and drugs), but we initially consider the case where there is just one person and we are interested in problems like

$$\begin{aligned}\texttt{ingests}(\texttt{Alice},\texttt{Doxazosin}) &\xrightarrow{?} \texttt{headache}(\texttt{Alice})\\ \texttt{ingests}(\texttt{Alice},\texttt{Prazosin}) &\xrightarrow{?} \texttt{headache}(\texttt{Alice})\\ &\vdots\end{aligned}$$

which concern the effects of different drugs on Alice.

As described in the main text, our model organizes the drugs into categories and assumes that the object-level model for each drug is generated from a corresponding causal model at the category level. Our prior $P(z)$ on category assignments is induced by the Chinese Restaurant Process (CRP, Aldous, 1985). Imagine building a partition by starting with a single category including a single object, and adding objects one by one until every object is assigned to a category. Under the CRP, each category attracts new members in proportion to its size, and there is some probability that a new object will be assigned to a new category. The distribution over categories for object $i$, conditioned on the category assignments for objects 1 through $i-1$ is

$$P(z_i = a \mid z_1, \ldots, z_{i-1}) = \begin{cases} \frac{n_a}{i-1+\gamma}, & n_a > 0 \\ \frac{\gamma}{i-1+\gamma}, & a \text{ is a new category} \end{cases} \tag{12}$$

where $z_i$ is the category assignment for object $i$, $n_a$ is the number of objects previously assigned to category $a$, and $\gamma$ is a hyperparameter (we set $\gamma = 0.5$). Because the CRP prefers to assign objects to categories which already have many members, the resulting prior $P(z)$ favors partitions that use a small number of categories.

When learning causal models for multiple objects, the parameters for each model can be organized into three vectors $\boldsymbol{a}$, $\boldsymbol{g}$, and $\boldsymbol{s}$. Let $\Psi$ be the tuple ($\boldsymbol{a}$, $\boldsymbol{g}$, $\boldsymbol{s}$, $b$) which includes all of these parameters along with the background strength $b$. Similarly, let $\bar{\Psi}$ be the tuple ($\bar{\boldsymbol{a}}$, $\bar{\boldsymbol{g}}$, $\bar{\boldsymbol{s}}$, $\bar{b}$) that specifies the parameters of the causal-models at the category level.

Our prior $P(\bar{\Psi})$ assumes that the entries in $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{g}}$ are independently drawn from a Beta($\gamma_c$,$\gamma_c$) distribution. Unless mentioned otherwise, we set $\gamma_c = 0.1$ in all cases. Each entry in $\bar{\boldsymbol{s}}$ is a pair that specifies a mean $\mu$ and a variance $\sigma^2$. We assume that these means and variances are independently drawn from the conjugate prior in Eq. 11 where $\eta = 1$. The remaining parameter $\bar{b}$ is a pair that specifies the mean and variance of the distribution that generates the background strength $b$. We assume that $\bar{b}$ is drawn from the conjugate prior specified by Eq. 11 where $\eta = -1$.

Suppose now that we are working in a setting (Fig. 1A) that includes two sets of objects—people and drugs. We introduce partitions $\boldsymbol{z}_{\text{people}}$ and $\boldsymbol{z}_{\text{drugs}}$ for both sets, and we place independent CRP priors on both partitions. We introduce a category-level causal model for each combination of a person category and a drug category, and we assume that each object-level causal model is generated from the corresponding category-level model. As before, we assume that the category-level parameters $\bar{a}$, $\bar{g}$, and $\bar{s}$ are generated independently for each category-level model. The same general strategy holds when working with problems that involve three or more sets of objects. We assume that each set is organized into a partition drawn from a CRP prior, introduce category level models for each

combination of categories, and assume that the parameters for these category-level models are independently generated from the distributions already described.

*Features*

To apply Eq. 8 we need to specify a prior distribution $P(\bar{F})$ on the feature matrix $\bar{F}$. We assume that all entries in the matrix are independent draws from a Beta($\gamma_f$,$\gamma_f$) distribution. Unless mentioned otherwise, we set $\gamma_f = 0.5$ in all cases. Our feature model is closely related to the Beta-Bernoulli model used by statisticians (Gelman et al., 2003) and is appropriate for problems where the features are binary. Some features, however, are categorical (i.e., they can take many discrete values), and others are continuous. Our approach can handle both cases by replacing the Beta-Bernoulli component with a Dirichlet-multinomial model, or a Gaussian model with conjugate prior.

*Inference*

Our model can be used to learn a schema (top level of Fig. 1), to learn a set of object-level causal models (middle level of Fig. 1), or to make predictions about future events involving a set of objects (bottom level of Fig. 1). All three kinds of inferences can be carried out using a Markov chain Monte Carlo (MCMC) sampler. Because we use conjugate priors on the model parameters at the category level ($\bar{\Psi}$ and $\bar{F}$), it is straightforward to integrate out these parameters and sample directly from $P(z,\Psi|V)$. To sample the schema assignments in $z$, we combined Gibbs updates with the split-merge scheme described by Jain and Neal (2004). We used Metropolis-Hasting updates on the parameters $\Psi$ of the object-level models and found that mixing improved when the three parameters for a given object $i$ ($a_i$, $g_i$ and $s_i$) were updated simultaneously. To further facilitate mixing, we used Metropolis-coupled MCMC: We ran several Markov chains at different temperatures and regularly considered swaps between the chains (Geyer, 1991).

We evaluated our model by comparing two kinds of distributions against human responses. Figs. 8, 10, 16, and 20 show posterior distributions over the activation strength of a given block, and Fig. 17 shows a posterior distribution over category assignments. In all cases except Fig. 20ii,iii we computed model predictions by drawing a bag of MCMC samples from $P(z,\Psi \mid V,F)$. We found that our sampler did not mix well when directly applied to the setting in Experiment 4 and therefore used importance sampling to generate the predictions in Fig. 20ii,iii. Let a partition $z$ be *plausible* if it assigns objects $o_1$ through $o_9$ to the same category and $o_{10}$ through $o_{18}$ to the same category. There are 15 plausible partitions, and we define a distribution $P_1(\cdot)$ that is uniform over these partitions:

$$P_1(z) = \begin{cases} \frac{1}{15}, & \text{if } z \text{ is plausible} \\ 0, & \text{otherwise} \end{cases}$$

For each plausible partition $z$ we used a separate MCMC run to draw 20,000 samples from $P(\Psi \mid V,z)$. When aggregated, these results can be treated as a single large sample from a distribution $q(z,\Psi)$ where

$$q(z, \Psi) \propto P(\Psi \mid V, z)P_1(z).$$

We generated model predictions for Fig. 20ii,iii using $q(\cdot,\cdot)$ as an importance sampling distribution. The importance weights required take the form $P(z)P(V \mid z)$, where $P(z)$ is induced by Eq. 12 and $P(V \mid z) = \int P(V|\Psi,z)P(\Psi \mid z)d\Psi$ can be computed using a simple Monte Carlo approximation for each plausible $z$.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1244–1286

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01129.x

# From Perceptual Categories to Concepts: What Develops?

Vladimir M. Sloutsky

*Center for Cognitive Science, The Ohio State University*

Received 24 December 2008; received in revised form 10 November 2009; accepted 12 November 2009

**Abstract**

People are remarkably smart: They use language, possess complex motor skills, make nontrivial inferences, develop and use scientific theories, make laws, and adapt to complex dynamic environments. Much of this knowledge requires concepts and this study focuses on how people acquire concepts. It is argued that conceptual development progresses from simple perceptual grouping to highly abstract scientific concepts. This proposal of conceptual development has four parts. First, it is argued that categories in the world have different structure. Second, there might be different learning systems (subserved by different brain mechanisms) that evolved to learn categories of differing structures. Third, these systems exhibit differential maturational course, which affects how categories of different structures are learned in the course of development. And finally, an interaction of these components may result in the developmental transition from perceptual groupings to more abstract concepts. This study reviews a large body of empirical evidence supporting this proposal.

*Keywords:* Cognitive development; Category learning; Concepts; Conceptual development; Cognitive neuroscience

## 1. Knowledge acquisition: Categories and concepts

People are remarkably smart: They use language, possess complex motor skills, make nontrivial inferences, develop and use scientific theories, make laws, and adapt to complex dynamic environments. At the same time, they do not exhibit evidence of this knowledge at birth. Therefore, one of the most interesting and exciting challenges in the study of human cognition is to gain an understanding of how people acquire this knowledge in the course of development and learning.

A critical component of knowledge acquisition is the ability to use acquired knowledge across a variety of situations, which requires some form of abstraction or generalization.

Correspondence should be sent to Vladimir M. Sloutsky, Cognitive Development Laboratory, Center for Cognitive Science, The Ohio State University, 208C Ohio Stadium East, 1961 Tuttle Park Place, Columbus, OH 43210. E-mail: sloutsky.1@osu.edu

Examples of abstraction are ample. People can recognize the same object under different viewing conditions. They treat different dogs as members of the same class and expect them to behave in fundamentally similar ways. They learn words uttered by different speakers. Upon learning a hidden property of an item, they extend this property to other similar items. And they apply ways of solving familiar problems to novel problems. In short, people can generalize or form equivalence classes by focusing only on some aspects of information while ignoring others.

This ability to form equivalence classes or categories is present in many nonhuman species (see Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008 for a review); however, only humans have the ability to acquire concepts—lexicalized groupings that allow ever-increasing levels of abstraction (e.g., Cat → Animal → Living thing → Object). These lexicalized groupings may include both observable and unobservable properties. For example, although prelinguistic infants can acquire a category ''cat'' by strictly perceptual means (Quinn, Eimas, & Rosenkrantz, 1993), the concept ''cat'' may include many properties that have to be inferred rather than observed directly (e.g., ''mating only with cats, but not with dogs,'' ''being able to move in a self-propelled manner,'' ''having insides of a cat,'' etc.). Often such properties are akin to latent variables—they are inferred from patterns of correlations among observable properties (Rakison & Poulin-Dubois, 2001). These properties can also be lexicalized, and when lexicalized, they allow nontrivial generalizations (e.g., ''plants and animals are alive'' or ''plants and animals reproduce themselves''). Although the existence of prelinguistic concepts is a matter of considerable debate, it seems rather noncontroversial to define those lexicalized properties that have to be inferred (rather than observed) as *conceptual* and lexicalized categories that include such properties as *concepts*.

Concepts are central to human intelligence as they allow uniquely human forms of expression, such as many forms of reasoning. For example, counterfactuals (e.g., ''if the defendant were at home at the time of the crime, she could not have been at the crime scene at the same time'') would be impossible without concepts. According to the present proposal, most concepts develop from perceptual categories and most conceptual properties are inferred from perceptual properties.[1] Therefore, although categories comprise a broader class than concepts (i.e., there are many categories that are not lexicalized and are not based on conceptual properties), there is no fundamental divide between category learning and concept acquisition.

Most of the examples presented in this study deal with ''thing'' concepts (these are lexicalized by ''nominals''), whereas many other concepts, such as actions, properties, quantities, and conceptual combinations are left out. This is because nominals are often most prevalent in the early vocabulary (Gentner, 1982; Nelson, 1973) and entities corresponding to nominals are likely to populate the early experience. Therefore, these concepts appear to be a good starting point in thinking about conceptual development.

The remainder of the study consists of four parts. First, I consider what may develop in the course of conceptual development. Second, I consider some of the critical components of category learning: the structure of input, the multiple competing learning systems, and the asynchronous developmental time course of these systems. Third, I consider evidence

for interactions among these components in category learning and category representation. And, finally, I consider how conceptual development may proceed from perceptual groupings to abstract concepts.

## 2. The origins of conceptual knowledge

In an attempt to explain developmental origins of conceptual knowledge, a number of theoretical accounts have been proposed. Some argue that input is dramatically underconstrained to enable the acquisition of complex knowledge and some knowledge has to come a priori from the organism, thus constraining future knowledge acquisition. Others suggest that there is much regularity (and thus many constraints) in the environment, with additional constrains stemming from biological specifications of the organism (e.g., limited processing capacity, especially early in development). In the remainder of this section, I review these theoretical approaches.

### *2.1. Skeletal principles, core knowledge, constraints, and biases*

According to this proposal, structured knowledge cannot be recovered from perceptual input because the input is too indeterminate to enable such recovery (Gelman, 1990). This approach is based on an influential idea that was originally proposed for the case of language acquisition but was later generalized to some other aspects of cognitive development, including conceptual development. The original idea is that linguistic input does not have enough information to enable the learner to recover a particular grammar, while ruling out alternatives (Chomsky, 1980). Therefore, some knowledge has to be innate to enable fast, efficient, and invariable learning under the conditions of impoverished input. This argument (known as the Poverty of the Stimulus argument) has been subsequently generalized to perceptual, lexical, and conceptual development. If input is too impoverished to constrain possible inductions and to license the concepts that we have, the constraints have to come from somewhere. It has been proposed that these constraints are internal—they come from the organism, and they are a priori and top-down (i.e., they do not come from data). A variety of such constraints have been proposed, including, but not limited to, innate knowledge within ''core'' domains (Carey, 2009; Carey & Spelke, 1994, 1996; Spelke, 2000; Spelke & Kinzler, 2007), skeletal principles (e.g., Gelman, 1990), ontological knowledge (Keil, 1979; Mandler, Bauer, & McDonough, 1991; Pinker, 1984; Soja, Carey, & Spelke, 1991), conceptual assumptions (Gelman, 1988; Gelman & Coley, 1991; Markman, 1989), and word-learning biases (Markman, 1989; see also Golinkoff, Mervis, & Hirsh-Pasek, 1994).

However, there are several lines of evidence challenging (a) the explanatory machinery of this account with respect to language (Chater & Christiansen, this issue) and (b) the existence of particular core abilities (e.g., Twyman & Newcome, this issue). Furthermore, although the Poverty of the Stimulus argument is formally valid, its premises and therefore its conclusions are questionable. Most importantly, very little is known about the information value of input with respect to knowledge in question. Therefore, it is not clear whether

input is in fact as impoverished as it has been claimed. In addition, there are several lines of evidence suggesting that input might be richer than it is expected under the Poverty of the Stimulus assumption.

First, the fact that infants, great primates, monkeys, rats, and birds all can learn a variety of basic-level perceptual categories (Cook & Smith, 2006; Quinn et al., 1993; Smith, Redford, & Haas, 2008; Zentall et al., 2008) strongly indicates that perceptual input (at least for basic-level categories) is not impoverished. Otherwise, one would need to assume that all these species have the same constraints as humans, which seems implausible given vastly different environments in which these species live.

In addition, there is evidence that perceptual input (Rakison & Poulin-Dubois, 2001) or a combination of perceptual and linguistic input (Jones & Smith, 2002; Samuelson & Smith, 1999; Yoshida & Smith, 2003) can jointly guide the acquisition of broad ontological classes. Furthermore, cross-linguistic evidence suggests that ontological boundaries exhibit greater cross-linguistic variability than could be expected if they were fixed (Imai & Gentner, 1997; Yoshida & Smith, 2003). Therefore, there might be enough information in the input for the learner to form both basic-level categories and broader ontological classes. There is also modeling work (e.g., Gureckis & Love, 2004; Rogers & McClelland, 2004) offering a mechanistic account of how coherent covariation in the input could guide the acquisition of broad ontological classes as well as more specific categories.

In short, there are reasons to doubt that input is in fact impoverished, and if it is not impoverished, then a priori assumptions are not necessary. Therefore, to understand conceptual development, it seems reasonable to shift the focus away from a priori constraints and biases and toward the input and the way it is processed.

### *2.2. Similarity, correlations, and attentional weights*

According to an alternative approach, conceptual knowledge as well as some of the biases and assumptions are a product rather than a precondition of learning (see Rogers & McClelland, 2004, for a connectionist implementation of these ideas). Early in development, cognitive processes are grounded in powerful learning mechanisms, such as statistical and attentional learning (French, Mareschal, Mermillod, & Quinn, 2004; Mareschal, Quinn, & French, 2002; Rogers & McClelland, 2004; Saffran, Johnson, Aslin, & Newport, 1999; Sloutsky, 2003; Sloutsky & Fisher, 2004a; Smith, 1989; Smith, Jones, & Landau, 1996).

According to this view, input is highly regular and the goal of learning is to extract these regularities. For example, category learning could be achieved by detecting multiple commonalities, or similarities, among presented entities. In addition, not all commonalities are the same—features may differ in salience and usefulness for generalization, with both salience and usefulness of a feature reflected in its attentional weight. However, unlike the a priori assumptions, attentional weights are not fixed and they can change as a result of learning: Attentional weights of more useful features increase, whereas these weights decrease for less useful features (Kruschke, 1992; Nosofsky, 1986; Opfer & Siegler, 2004; Sloutsky & Spino, 2004; see also Hammer & Diesendruck, 2005).

There are several lines of research presenting evidence that both basic-level categories (e.g., dogs) and broader ontological classes (e.g., animates vs. inanimates) have multiple perceptual within-category commonalities and between-category differences (French et al., 2004; Rakison & Poulin-Dubois, 2001; Samuelson & Smith, 1999). Some researchers argue that additional statistical constraints come from language in the form of syntactic cues, such as count noun and mass noun syntax (Samuelson & Smith, 1999). Furthermore, cross-linguistic differences in the syntactic cues (e.g., between English and Japanese) can push ontological boundaries in speakers of respective languages (Imai & Gentner, 1997; Yoshida & Smith, 2003). Finally, different categories could be organized differently (e.g., living things could be organized by multiple similarities, whereas artifacts could be organized by shape), and there might be multiple correlations between category structure, perceptual cues, and linguistic cues. All this information could be used to distinguish between different kinds. As children acquire language, they may become sensitive to these correlations, which may affect their attention to shape in the context of artifacts versus living things (Jones & Smith, 2002).

This general approach may offer an account of conceptual development that does not posit a priori knowledge structures. It assumes that input is sufficiently rich to enable the young learner to form perceptual groupings. Language provides learners with an additional set of cues that allow them to form more abstract distinctions. Finally, lexicalization of such groupings as well as of some unobservable conceptual features could result in the acquisition of concepts at progressively increasing levels of abstraction. In the next section, I will outline how conceptual development could proceed from perceptual groupings to abstract concepts.

### *2.3. From percepts to concepts: What develops?*

If people start out with perceptual groupings, how do they end up with sophisticated conceptual knowledge? According to the proposal presented here, conceptual development hinges on several critical steps. These include the ability to learn similarity-based unimodal categories, the ability to integrate cross-modal information, the lexicalization of learned perceptual groupings, the lexicalization of conceptual features, and the development of executive function. The latter development is of critical importance for acquiring abstract concepts that are not grounded in similarity. Examples of such concepts are unobservables (e.g., *love*, *doubt*, *thought*), relational concepts (e.g., *enemy* or *barrier*), as well as a variety of rule-based categories (e.g., *island*, *uncle*, or *acceleration*). Because these concepts require focusing on unobservable abstract features, their acquisition may depend on the maturity of executive function.

This developmental time course is determined in part by an interaction of several critical components. These components include the following: (a) the structure of the to-be-learned category, (b) the competing learning systems that might subserve learning categories of different structures, and (c) developmental course of these learning systems. First, categories differ in their structure. For example, some categories (e.g., most of natural kinds, such as *cat* or *dog*) have multiple intercorrelated features relevant for category membership. These

features are jointly predictive, thus yielding a highly redundant (or statistically dense) category. These categories often have graded membership (i.e., a typical dog is a better member of the category than an atypical dog) and fuzzy boundaries (i.e., it is not clear whether a cross between a dog and a cat is a dog). At the same time, other categories are defined by a single dimension or a relation between or among dimensions. Members of these categories have very few common features, with the rest of the features varying independently and thus contributing to irrelevant or ''surface'' variance. Good examples of such sparse categories are mathematical and scientific concepts. Consider the two situations: (a) increase in a population of fish in a pond and (b) interest accumulation in a bank account. Only a single commonality—exponential growth—makes both events instances of the same mathematical function. All other features are irrelevant for membership in this category and can vary greatly.

Second, there might be multiple systems of category learning (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998) evolved to learn categories of different structures. In particular, a compression-based system may subserve category learning by reducing perceptually rich input to a more basic format. As a result of this compression, features that are common to category members (but not to nonmembers) become a part of representation, whereas idiosyncratic features get washed out. In contrast, the selection-based system may subserve category learning by shifting attention to category-relevant dimension(s) and away from irrelevant dimension(s). Such selectivity may require the involvement of brain structures associated with executive function. The compression-based system could have an advantage for learning dense categories, which could be acquired by mostly perceptual means. At the same time, the selection-based system could have an advantage for learning sparse categories, which require focusing on few category-relevant features (Kloos & Sloutsky, 2008; see also Blair, Watson, & Meier, 2009, for a discussion).

The involvement of each system may also affect what information is encoded in the course of category learning, and, subsequently, how a learned category is represented. In particular, the involvement of the compression-based system may result in a reduced yet fundamentally perceptual representation of a category, whereas the involvement of the selection-based system may result in a more abstract (e.g., lexicalized) representation. Given that many real-life categories (e.g., dogs, cats, or cups) are acquired by perceptual means and later undergo lexicalization, there are reasons to believe that these categories combine perceptual representation with a more abstract lexicalized representation. These abstract lexicalized representations are critically important for the ability to reason and form arguments that could be all but impossible to form by strictly perceptual means. For example, it is not clear how purely perceptual representation of constituent entities would support a counterfactual of the form ''If my grandmother were my grandfather…''

And third, the category-learning systems and associated brain structures may come online at different points in development, with the system subserving learning of dense categories coming online earlier than the system subserving learning of sparse categories. In particular, there is evidence that many components of executive function critical for learning sparse categories exhibit late developmental onset (e.g., Davidson, Amso, Anderson, & Diamond, 2006). If this is the case, then able learning and representation of dense categories should

precede that of sparse categories. Under this view, ''conceptual'' assumptions do not have to underlie category learning, as most categories that children acquire are spontaneously dense and can be acquired implicitly, without a teaching signal or supervision. At the same time, some of these ''conceptual'' assumptions could be a product of later development.

The current proposal of conceptual development has three parts (see Sections 3–5). In Section 3, I consider in detail components of category learning: category structure, the multiple competing learning systems, and the potentially different maturational course of these systems. I suggest that categories in the world differ in their structure and consider ways of quantifying this structure. I then present another argument that there might be different learning systems (subserved by different brain mechanisms) that evolved to learn categories of differing structures. Finally, I argue that these systems exhibit differential maturational course, which affects how categories of different structures are learned in the course of development. Then, in Section 4, I consider an interaction of these components. This interaction is important because it may result in the developmental transition from perceptual groupings to abstract concepts. These arguments point to a more nuanced developmental picture (presented in Section 5), in which learning of perceptual categories, cross-modal integration, lexicalization, learning of conceptual properties, the ability to focus and shift attention, and the development of lexicalized concepts are logical steps in conceptual development.

## 3. Components of category learning: Input, learning system, and the learner

### *3.1. Characteristics of input: Category structure*

It appears almost self-evident that categories differ in their structure. Some categories are coherent: Their members have multiple overlapping features and are often similar (e.g., cats or dogs are good examples of such categories). Other categories seem to be less coherent: Their members have few overlapping features (e.g., square things). These differences have been noted by a number of researchers who pointed to different category structures between different levels of ontology (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) and between animal and artifact categories (Jones & Smith, 2002; Jones, Smith, & Landau, 1991; Markman, 1989). Category structure can be captured formally and one such treatment of category structure has been offered recently (Kloos & Sloutsky, 2008). The focal idea of this proposal is that category structure can be measured by statistical density of a category. Statistical density is a function of within-category compactness and between-category distinctiveness, and it may have profound effects on category learning. In what follows, I will elaborate this idea.

#### *3.1.1. Statistical density as a measure of category structure*

Any set of items can have a number of possible dimensions (e.g., color, shape, size), some of which might vary and some of which might not. Categories that are statistically dense have multiple intercorrelated (or covarying) features relevant for category membership, with only a few features being irrelevant. Good examples of statistically dense

categories are basic-level animal categories such as cat or dog. Category members have particular distributions of values on a number of dimensions (e.g., shape, size, color, texture, number of parts, type of locomotion, type of sounds they produce, etc.). These distributions are jointly predictive, thus yielding a dense (albeit probabilistic) category. Categories that are statistically sparse have very few relevant features, with the rest of the features varying independently. Good examples of sparse categories are dimensional groupings (e.g., ''round things''), relational concepts (e.g., ''more''), scientific concepts (e.g., ''accelerated motion''), or role-governed concepts (e.g., cardinal number; see Markman & Stilwell, 2001, for a discussion of role-governed categories).

Conceptually, statistical density is a ratio of variance relevant for category membership to the total variance across members and nonmembers of the category. Therefore, density is a measure of statistical redundancy (Shannon & Weaver, 1948), which is an inverse function of relative entropy.

Density can be expressed as

$$D = 1 - \frac{H_{\text{within}}}{H_{\text{between}}}, \tag{1}$$

where $H_{\text{within}}$ is the entropy observed within the target category, and $H_{\text{between}}$ is the entropy observed between target and contrasting categories.

A detailed treatment of statistical density and ways of calculating it is presented elsewhere (Kloos & Sloutsky, 2008); thus, only a brief overview of statistical density is presented below. Three aspects of stimuli are important for calculating statistical density: variation in stimulus dimensions, variation in relations among dimensions, and attentional weights of stimulus dimensions.

First, a stimulus dimension may vary either within a category (e.g., members of a target category are either black or white) or between categories (e.g., all members of a target category are black, whereas all members of a contrasting category are white). Within-category variance decreases the density, whereas between-category variance increases the density.

Second, dimensions of variation may be related (e.g., all items are black circles), or they may vary independently of each other (e.g., items can be black circles, black squares, white circles, or white squares). Covarying dimensions result in smaller entropy than dimensions that vary independently. It is not unreasonable to assume that only dyadic relations (i.e., relations between two dimensions) are detected spontaneously, whereas relations of higher arity (e.g., a relation among color, shape, and size) are not (Whitman & Garner, 1962). Therefore, only dyadic relations are included in the calculation of entropy.

The total entropy is the sum of the entropy due to varying dimensions ($H^{\text{dim}}$), and the entropy due to varying relations among the dimensions ($H^{\text{rel}}$). More specifically,

$$H_{\text{within}} = H^{\text{dim}}_{\text{within}} + H^{rel}_{\text{within}}, \text{and} \tag{2a}$$

$$H_{\text{between}} = H^{\text{dim}}_{\text{between}} + H^{\text{rel}}_{\text{between}}. \tag{2b}$$

The concept of entropy was formalized by the information theory (Shannon & Weaver, 1948), and we use these formalisms here. First consider the entropy due to dimensions. This within- and between-category entropy is presented in Eqs. 3a and 3b, respectively.

$$H^{\text{dim}}_{\text{within}} = -\sum_{i=1}^{M} w_i \left[ \sum_{j=0,1} {}_{\text{within}}(p_j \log_2 p_j) \right] \tag{3a}$$

$$H^{\text{dim}}_{\text{between}} = -\sum_{i=1}^{M} w_i \left[ \sum_{j=0,1} {}_{\text{between}}(p_j \log_2 p_j) \right] \tag{3b}$$

where $M$ is the total number of varying dimensions, $w_i$ is the attentional weight of a particular dimension (the sum of attentional weights equals to a constant), and $p_j$ is the probability of value $j$ on dimension $i$ (e.g., the probability of a color being white). The probabilities could be calculated within a category or between categories.

The attentional-weight parameter is of critical importance—without this parameter, it would be impossible to account for learning of sparse categories. In particular, when a category is dense, even relatively small attentional weights of individual dimensions add up across many dimensions. This makes it possible to learn the category without supervision. Conversely, when a category is sparse, only few dimensions are relevant. If attentional weights of each dimension are too small, supervision could be needed to direct attention to these relevant dimensions.

Next, consider entropy that is due to a relation between dimensions. To express this entropy, we need to consider the co-occurrences of dimensional values. If dimensions are binary, with each value coded as 0 or 1 (e.g., white = 0, black = 1, circle = 0, and square = 1), then the following four co-occurrence outcomes are possible: 00 (i.e., white circle), 01 (i.e., white square), 10 (i.e., black circle), and 11 (i.e., black square). The within- and between-category entropy that is due to relations is presented in Eqs. 4a and 4b, respectively.

$$H^{\text{rel}}_{\text{within}} = -\sum_{k=1}^{o} w_k \left[ \sum_{\substack{m=0,1 \\ n=0,1}} {}_{\text{within}}(p_{mn} \log_2 p_{mn}) \right], \tag{4a}$$

$$H^{\text{rel}}_{\text{between}} = -\sum_{k=1}^{o} w_k \left[ \sum_{\substack{m=0,1 \\ n=0,1}} {}_{\text{between}}(p_{mn} \log_2 p_{mn}) \right], \tag{4b}$$

where $o$ is the total number of possible dyadic relations among the varying dimensions, $w_k$ is the attentional weight of a particular relation (again, the sum of attentional weights equals to a constant), and $p_{mn}$ is the probability of a co-occurrence of values $m$ and $n$ on a binary relation $k$ (which conjoins two dimensions of variation).

*3.1.2. Density, salience, and similarity*

The concept of density is closely related to the ideas of salience and similarity, and it is necessary to clarify these relations. First, density is a function of *weighted* entropy, with attentional weight corresponding closely to the salience of a feature. Therefore, feature salience can affect density by affecting the attentional weight of the feature in question. Of course, as mentioned earlier, attentional weights are not fixed and they can change as a result of learning. Second, perceptual similarity is a sufficient, but not necessary condition of density—all categories bound by similarity are dense, but not all dense categories are bound by similarity. For example, some categories could have multiple overlapping relations rather than overlapping features (e.g., members of a category have short legs and short neck or long legs and long neck). It is conceivable that such nonlinearly separable (NLS) categories could be relatively dense, yet not bound by similarity.

*3.1.3. Category structure and early learning*

Although it is difficult to precisely calculate the density of categories surrounding young infants, some estimates can be made. It seems that many of these categories, while exhibiting within-category variability in color (and sometime in size), have similar within-category shape, material, and texture (*ball*, *cup*, *bottle*, *shoe*, *book*, or *apple* are good examples of such categories); these categories should be relatively dense. As I show next, dense categories can be learned implicitly, without supervision. Therefore, it is possible that prelinguistic infants implicitly learn many of the categories surrounding them. Incidentally, the very first noun words that infants learn denote these dense categories (see Dale & Fenson, 1996; Nelson, 1973). Therefore, it is possible that some of early word learning consists of learning lexical entries for already known dense categories. This possibility, however, is yet to be tested empirically.

*3.1.3.1. Characteristics of the learning system: Multiple competing systems of category learning:* The role of category structure in category learning has been a focus of the neuroscience of category learning. Recent advances in that field suggest that there might be multiple systems of category learning (e.g., Ashby et al., 1998; Cincotta & Seger, 2007; Nomura & Reber, 2008; Seger, 2008; Seger & Cincotta, 2002) and an analysis of these systems may elucidate how category structure interacts with category learning. I consider these systems in this section.

There is an emerging body of research on brain mechanisms underlying category learning (see Ashby & Maddox, 2005; Seger, 2008, for reviews). Although the anatomical localization and the involvement of specific circuits remain a matter of considerable debate, there is substantial agreement that ''wholistic'' or ''similarity-based'' categories (which are typically dense) and ''dimensional'' or ''rule-based'' categories (which are typically sparse) could be learned by different systems in the brain.

There are several specific proposals identifying brain structures that comprise each system of category learning (Ashby et al., 1998; Cincotta & Seger, 2007; Nomura & Reber, 2008; Seger, 2008; Seger & Cincotta, 2002). Most of the proposals involve three major hierarchical structures: cortex, basal ganglia, and thalamus. There is also evidence for the

involvement of the medial temporal lobe (MTL) in category learning (e.g., Nomura et al., 2007; see also Love & Gureckis, 2007). However, because the maturational time course of the MTL is not well understood (Alvarado & Bachevalier, 2000), I will not focus here on this area of the brain.

One influential proposal (e.g., Ashby et al., 1998) posited two cortical–striatal–pallidal–thalamic–cortical loops, which define two acting in parallel circuits. The circuit responsible for learning of similarity-based categories originates in extrastriate visual areas of the cortex (such as inferotemporal [IT] cortex) and includes the posterior body and tail of the caudate nucleus. In contrast, the circuit responsible for the learning of rule-based categories originates in the prefrontal and anterior-cingulated cortices (ACCs) and includes the head of the caudate (Lombardi et al., 1999; Rao et al., 1997; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000).

In a similar vein, Seger and Cincotta (2002) proposed the *visual* loop, which originates in the inferior temporal areas and passes through the tail of the caudate nucleus in the striatum, and the *cognitive* loop, which passes through the prefrontal cortex (PFC) and the head of the caudate nucleus. The visual loop has been shown to be involved in visual pattern discrimination in nonhuman animals (Buffalo et al., 1999; Fernandez-Ruiz, Wang, Aigner, & Mishkin, 2001; Teng, Stefanacci, Squire, & Zola, 2000), and Seger and Cincotta (2002) have proposed that this loop may subserve learning of similarity-based visual categories. The cognitive loop has been shown to be involved in learning of rule-based categories (e.g., Rao et al., 1997; Seger & Cincotta, 2002; see also Seger, 2008).

There is also evidence that category learning is achieved differently in the two systems. The critical feature of the visual system is the reduction of information or *compression*, with only some but not all stimulus features being encoded. Therefore, I will refer to this system as the *compression-based* system of category learning. A schematic representation of processing in this system is depicted in Fig. 1A. The feature map in the top layer gets compressed in the bottom layer, with only some features of the top layer represented in the bottom layer.

This compression is achieved by many-to-one projections of the visual cortical neurons in the IT cortex onto the neurons of the tail of the caudate (Bar-Gad, Morris, & Bergman, 2003; Wilson, 1995). In other words, many cortical neurons converge on an individual caudate neuron. As a result of this convergence, information is compressed to a more basic form, with redundant and highly probable features likely to be encoded (and thus learned) and idiosyncratic and rare features likely to be filtered out.

Category learning in this system results in a reduced (or compressed) yet fundamentally perceptual representation of stimuli. If every stimulus is compressed, then those features and feature relations that are frequent in category members should survive the compression, whereas rare or unique features/relations should not. Because compression does not require selectivity, compression-based learning could be achieved implicitly, without supervision (such as feedback or even more explicit forms of training), and it should be particularly successful in the learning of dense categories.

In short, there is a critical feature of the compression-based system—it can learn dense categories without supervision. Under some conditions, the compression-based system may

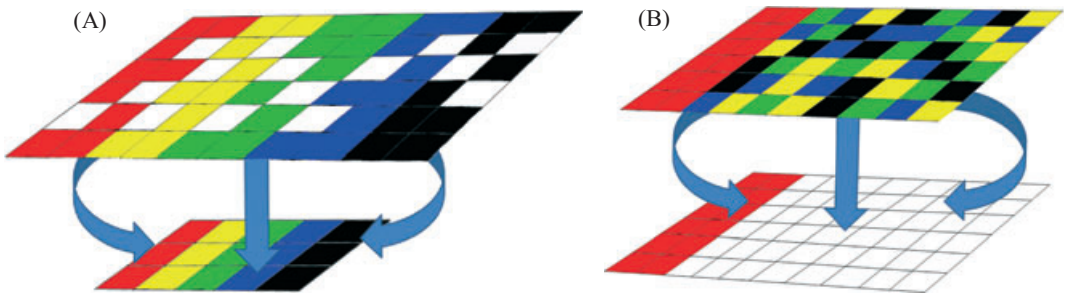


Fig. 1. (A) Schematic depiction of the compression-based system. The top layer represents stimulus encoding in inferotemporal cortex. This rich encoding gets compressed to a more basic form in the striatum represented by the bottom layer. Although some of the features are left out, much perceptual information present in the top layer is retained in the bottom layer. (B) Schematic depiction of the selection-based system. The top layer represents selective encoding in the prefrontal cortex. The selected dimension is then projected to the striatum represented by the bottom layer. Only the selected information is retained in the bottom layer.

also learn structures defined by a single dimension of variation (e.g., color or shape). For example, when there is a small number of dimensions of variation (e.g., color and shape, with shape distinguishing among categories), compression may be sufficient for learning a category-relevant dimension. However, if categories are sparse, with only few relevant dimensions and multiple irrelevant dimensions, learning of the relevant dimensions by compression could be exceedingly long or not possible at all.

The critical aspect of the second system of category learning is the cognitive loop, which involves (in addition to the striatum) the dorsolateral PFC and the ACC—the cortical areas that subserve attentional selectivity, working memory, and other aspects of executive function (Posner & Petersen, 1990). I will therefore refer to this system as *selection-based*. The selection-based system enables attentional learning—allocation of attention to some stimulus dimensions and ignoring others (e.g., Kruschke, 1992, 2001; Mackintosh, 1975; Nosofsky, 1986). Unlike the compression-based system where learning is driven by reduction and filtering of idiosyncratic features (while retaining features and feature correlations that recur across instances), learning in the selection-based system could be driven by error reduction. As schematically depicted in Fig. 1B, attention is shifted to those dimensions that predict error reduction and away from those that do not (e.g., Kruschke, 2001; but see Blair et al., 2009).

Given that attention has to be shifted to a relevant dimension, the task of category learning within the selection-based system should be easier when there are fewer relevant dimensions (see Kruschke, 1993, 2001, for related arguments). This is because it is easier to shift attention to a single dimension than to allocate it to multiple dimensions. Therefore, the selection-based system is better suited to learn sparse categories (recall that the compression-based system is better suited to learn dense categories). For example, Kruschke (1993) describes an experiment where participants learned a category in a supervised manner, with feedback presented on every trial. For some categories, a single dimension was relevant, whereas for other categories, two related dimensions were relevant for categorization.

Participants were shown to learn better in the former than in the latter condition. Given that learning was supervised (i.e., category learning and stimulus dimensions that might be relevant for categorization were mentioned explicitly, and feedback was given on every trial), it is likely that the selection-based system was engaged.

The selection-based system depends critically on prefrontal circuits because these circuits enable the selection of a relevant stimulus dimension (or rule), while inhibiting irrelevant dimensions. The selected (and perhaps amplified) dimension is likely to survive the compression in the striatum, whereas the nonselected (and perhaps weakened) dimensions may not. Therefore, there is little surprise that young children (whose selection-based system is still immature) tend to exhibit successful categorization performance when categories are based on multiple dimensions than when they are based on a single dimension (e.g., Smith, 1989).

How are the systems deployed? Although the precise mechanism remains unknown, several ideas have been proposed. For example, Ashby et al. (1998) posited competition between the systems, with the selection-based system being deployed by default. This idea stems from evidence that participants exhibited more able learning when categories were based on a single dimension than when categories are based on multiple dimensions (e.g., Ashby et al., 1998; Kruschke, 1993). However, it is possible that the selection-based system was triggered by feedback and explicit learning regime, whereas in the absence of supervision the compression-based system is a default (Kloos & Sloutsky, 2008). Furthermore, it seems unlikely that the idea of the default deployment of the selection-based system describes accurately what happens early in development. As I argue in the next section, because some critical cortical components of the selection-based system mature relatively late, it is likely that early in development the competition is weakened (or even absent), thus making the compression-based system default.

If the compression-based system is deployed by default early in development (and, when supervision is absent, it is deployed by default in adults as well), this default deployment may have consequences for category learning. In particular, if a category is sparse, the compression-based system may fail to learn it due to a low signal-to-noise ratio in the sparse category. In contrast, the selection-based system may have the ability to increase the signal-to-noise ratio by shifting attention to the signal, thus either amplifying the signal or by inhibiting noise.

The idea of multiple systems of category learning has been supported by both fMRI and neuropsychological evidence. In one neuroimaging study reported by Nomura et al. (2007), participants were scanned while learning two categories of sine wave gratings. The gratings varied on two dimensions: spatial frequency and orientation of the lines. In the rule-based condition, category membership was defined only by the spatial frequency of the lines (see Fig. 2A), whereas in the ''wholistic'' condition, both frequency and orientation determined category membership (see Fig. 2B). Note that each point in Fig. 2 represents an item and the colors represent distinct categories. Rule-based categorization showed greater differential activation in the hippocampus, the ACC, and medial frontal gyrus. At the same time, the wholistic categorization exhibited greater differential activation in the head and tail of the caudate.

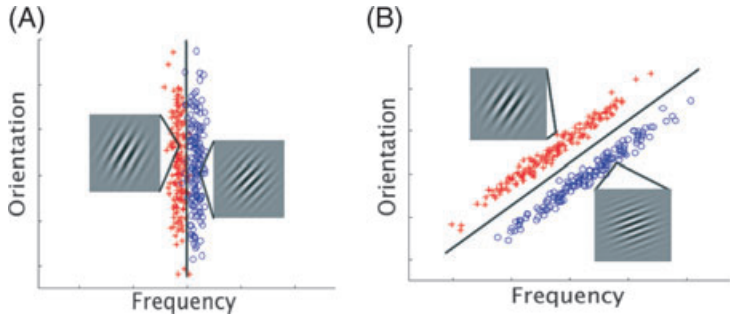


Fig. 2. RB (A) and II stimuli (B) (after Nomura & Reber, 2008). Each point represents a distinct Gabor patch (sine-wave) stimulus defined by orientation (tilt) and frequency (thickness of lines). In both stimulus sets, there are two categories (red and blue points). RB categories are defined by a vertical boundary (only frequency is relevant for categorization), whereas II categories are defined by a diagonal boundary (both orientation and frequency are relevant). In both RB and II stimuli, there are examples of a stimulus from each category. II, information integration; RB, rule based.

Some evidence for the possibility of the two systems of category learning stem from neuropsychological research. One of the most frequently studied populations are patients with Parkinson's disease (PD), because the disease often affects frontal cortical areas in addition to striatal areas (e.g., van Domburg & ten Donkelaar, 1991). As a result, these patients often exhibit impairments in both the compression-based and the selection-based systems of category learning. Therefore, this group provides only indirect rather than clear-cut evidence for the dissociation between the systems. For example, impairments of the compression-based system in PD were demonstrated in a study by Knowlton, Mangels, and Squire (1996), in which patients with PD (which affects the release of dopamine in the striatum) had difficulty in learning probabilistic categories that were determined by co-occurrence of multiple perceptual cues. Impairments of the selection-based learning system have been demonstrated in patients with damage to the PFC (which also often include PD patients). Specifically, in the multiple studies using the Wisconsin Card Sorting Test (WCST: Berg, 1948; Brown & Marsden, 1988; Cools, van den Bercken, Horstink, van Spaendonck, & Berger, 1984), it was found that the patients often exhibit impaired learning of categories based on verbal rules, as well as impairments in shifting attention from successfully learned rules to new rules (see Ashby et al., 1998, for a review).

In the WCST, participants have to discover an experimenter-defined matching rule (e.g., ''objects with the same shape go together'') and respond according to the rule. In the middle of the task, the rule may change and participants must sort according to the new rule. Two aspects of the task are of interest, rule learning and response shifting, with both being likely to be subserved by the selection-based system (see Ashby et al., 1998, for a discussion). There are several types of shifts, with two being of particular interest for understanding of the selection-based system—the reversal shift and the extradimensional shift.

The reversal shift consists of a reassignment of a dimension to a response. For example, a participant could initially learn that ''if Category A (say the color is green), then press

button 1, and if Category B (say the color is red), then press button 2.'' The reversal shift requires a participant to change the pattern of responding, such that ''if Category A, then press button 2, and if Category B, then press button 1.'' In contrast, the extradimensional shift consists of change in which dimension is relevant. For example, if a participant initially learned that ''if Category A (say the color is green), then press button 1, and if Category B (say the color is red), then press button 2,'' the extradimensional shift would require a different pattern of responding: ''if Category K (say the size is small), then press button 1, and if Category M (say the size is large), then press button 2.'' Findings indicate that patients with lesions to PFCs had substantial difficulties with extradimensional, but not with the reversal shifts on the WCST (e.g., Rogers et al., 2000). Therefore, these patients did not have a difficulty in inhibiting the previously learned pattern of responding but rather had difficulty in shifting attention to a formerly irrelevant dimension, which is indicative of a selection-based system impairment.

In sum, there is evidence that the compression-based and the selection-based system may be dissociated in the brain. Furthermore, although both systems involve parts of the striatum, they differ with respect to other areas of the brain. Whereas the selection-based system relies critically on the PFC and the ACC, the compression-based system relies on IT cortex. As I argue in the next section, the IT and the PFCs may exhibit differential maturational time course. The relative immaturity of PFCs early in development coupled with a relative maturity of the IT cortex and the striatum should result in young children having a more mature compression-based than selection-based system and thus being more efficient in learning dense than sparse categories (Smith, 1989; Smith & Kemier-Nelson, 1984).

### *3.2. Characteristics of the learner: Differential maturational course of brain systems underlying category learning*

Many vertebrates have a brain structure analogous to the IT cortex and the striatum, whereas only mammals have a developed PFC (Striedter, 2005). Studies of normal brain maturation (Jernigan, Trauner, Hesselink, & Tallal, 1991; Pfefferbaum et al., 1994; Caviness, Kennedy, Richelme, Rademacher, & Filipek, 1996; Giedd et al., 1996a, 1996b; Sowell & Jernigan, 1999; Sowell, Thompson, Holmes, Batth, Jernigan, and Toga, 1999, Sowell, Thompson, Holmes, Jernigan, and Toga, 1999) have indicated that brain morphology continues to change well into adulthood. As noted by Sowell, Thompson, Holmes, Batth, et al. (1999), maturation progresses in a programmed way, with phylogenetically more primitive regions of the brain (e.g., brain stem and cerebellum) maturing earlier, and more advanced regions of the brain (e.g., the association circuits of the frontal lobes) maturing later. In addition to the study of brain development focused on the anatomy, physiology, and chemistry of the changing brain, researchers have studied the development of function that is subserved by particular brain areas.

Given that the two learning systems differ primarily with respect to the cortical structures involved (the basal ganglia structures are involved in both systems), I will focus primarily on the maturational course of these cortical systems. I will first review data pertaining to the

maturational course of IT and associated visual recognition functions and then pertaining to the PFC and associated executive function.

*3.2.1. Maturation of the IT cortex*

Maturation of the IT cortex has been extensively studied in monkeys using single-cell recording techniques. As demonstrated by several researchers (e.g., Rodman, 1994; Rodman, Skelly, & Gross, 1991), many fundamental properties of IT emerge quite early. Most importantly, as early as 6 weeks, neurons in this cortical area exhibit adult-like patterns of responsiveness. In particular, researchers presented subjects with different images (e.g., monkey faces and objects varying in spatial frequency), while recording electrical activity of IT neurons. They found that, in both infant and adult monkeys, IT neurons exhibited a pronounced form of tuning, with different neurons responding selectively to different types of stimuli. These and similar findings led researchers to conclude that the IT cortex is predisposed to rapidly develop major neural circuitry necessary for basic visual processing. Therefore, although some aspects of the IT circuitry may exhibit a more prolonged development, the basic components develop relatively early. These findings contrast sharply with findings indicating a lengthy developmental time course of PFCs (e.g., Bunge & Zelazo, 2006).

*3.2.2. Maturation of the PFC*

There is a wide range of anatomical, neuroimaging, neurophysiological, and neurochemical evidence indicating that the development of the PFC continues well into adolescence (e.g., Sowell, Thompson, Holmes, Jernigan, et al., 1999; see also Luciana & Nelson, 1998; Rueda et al., 2004; Davidson et al., 2006, for extensive reviews).

The maturational course of the PFC has been studied in conjunction with research on executive function—the cognitive function that depends critically on the maturity of the PFC (Davidson et al., 2006; Diamond & Goldman-Rakic, 1989; Fan, McCandliss, Sommer, Raz, & Posner, 2002; Goldman-Rakic, 1987; Posner & Petersen, 1990). Executive function comprises a cluster of abilities such as holding information in mind while performing a task, switching between tasks or between different demands of a task, inhibiting a dominant response, deliberate selection of some information and ignoring other information, selection among different responses, and resolving conflicts between competing stimulus properties and competing responses.

There is a large body of behavioral evidence that, early in development, children exhibit difficulties in deliberately focusing on relevant stimuli, inhibiting irrelevant stimuli, and switching attention between stimuli and stimulus dimensions (Diamond, 2002; Kirkham, Cruess, & Diamond, 2003; Napolitano & Sloutsky, 2004; Shepp & Swartz, 1976; Zelazo, Frye, & Rapus, 1996; Zelazo, Muller, Frye, & Marcovitch, 2003; see also Fisher, 2007, for a more recent review).

Maturation of the prefrontal structures in the course of individual development results in progressively greater efficiency of executive function, including the ability to deliberately focus on what is relevant while ignoring what is irrelevant. This is a critical step in acquiring the ability to form abstract, similarity-free representations of categories and use these

representations in both category and property induction. Therefore, the development of relatively abstract category-based generalization may hinge on the development of executive function. As suggested above, while the selection-based system could be deployed by default in adults when learning is supervised (e.g., Ashby et al., 1998), it could be that, early in development, it is the compression-based system that is deployed by default.

Therefore, there are reasons to believe that the cortical circuits that subserve the compression-based learning system (i.e., IT) come online earlier than the cortical circuits that subserve the selection-based learning system (i.e., PFC). Thus, it seems likely that, early in development, children would be more efficient in learning dense, similarity-bound categories (as these could be efficiently learned by the compression-based system) than sparse, similarity-free ones (as these require the involvement of the selection-based system).

In sum, understanding category learning requires understanding an interaction of at least three components: (a) the structure of the input, (b) the learning system that evolved to process this input, and (c) the characteristics of the learner in terms of the availability and maturity of each of the system. Understanding the interaction among these components leads to several important predictions. First, dense categories should be learned more efficiently by the nondeliberate, compression-based system, whereas sparse categories should be learned more efficiently by the more deliberate selection-based system. Second, because the critical components of the selection-based system develop late (both phylo- and ontogenetically) relative to the compression-based system, learning of dense categories should be more universal, whereas learning of sparse categories should be limited to those organisms that have a developed PFC. Third, because the selection-based system of category learning undergoes a more radical developmental transformation, learning of sparse categories should exhibit greater developmental change than learning of dense categories. Fourth, young children can spontaneously learn dense categories that are based on multiple overlapping features, whereas they should have difficulty in spontaneously learning sparse categories that have few relevant features or dimensions and multiple irrelevant features. Note that the critical aspect here is not whether a category is defined by a single dimension or by multiple dimensions, but whether the category is dense or sparse. For example, it should be less difficult to learn a color-based categorization if color is the only dimension that varies across the categories, whereas it should be very difficult to learn a color-based categorization if items vary on multiple irrelevant dimensions. And finally, given the immaturity of the selection-based system of category learning and of executive function, it seems implausible that, early in development, children can spontaneously use a single predictor as a category marker overriding all other predictors. In particular, this immaturity casts doubt on the ability of babies or even young children to spontaneously use linguistic labels as category markers in category representation. Because the issue of the role of category labels in category representation is of critical importance for understanding of conceptual development, I will focus on it in one of the sections below.

In what follows, I review empirical evidence that has been accumulated over the years, with particular focus on research generated in my lab. Although many issues remain unknown, I will present two lines of evidence supporting these predictions. First, I present evidence that category structure, learning system, and developmental characteristics of the

learner interact in category learning and category representation. In particular, early in development, the compression-based system exhibits greater efficiency than the selection-based system. In addition, early in development, categories are represented perceptually, and only later do participants form more abstract, dimensional, rule-based, or lexicalized representations of categories. And second, the role of words in category learning is not fixed; rather, it undergoes developmental change: Words initially affect processing of visual input, and only gradually they become category markers.

## 4. Interaction among category structure, learning system, and characteristics of the learner: Evidence from category learning and category representation

Recall that I hypothesized an interaction among (a) the structure of the category (in particular, its density), (b) the learning system that evolved to process this input, and (3) the characteristics of the learner in terms of the availability and maturity of each system. In what follows, I consider components of this interaction with respect to category learning and category representation.

### *4.1. Category learning*

As discussed earlier, there are reasons to believe that, in the course of individual development, the compression-based system comes online earlier than the selection-based system (i.e., due to the protracted immaturity of the executive function that subserves the selection-based system). Therefore, it seems plausible that, at least early in development, the compression-based system is deployed by default, whereas the selection-based system has to be triggered explicitly (see Ashby et al., 1998 for arguments that this may not be the case in adults). It is also possible that there are experimental manipulations that could trigger the nondefault system. In particular, the selection-based system could be triggered by explicit supervision or an error signal.

If the systems are dissociated, then sparse categories that depend critically on selective attention (as they require focusing on a few relevant dimensions, while ignoring irrelevant dimensions) may be learned better under the conditions triggering the selection-based system. At the same time, dense categories that have much redundancy may be learned better under the conditions of implicit learning. Finally, because dense categories could be efficiently learned by the compression-based system, which is more primary, both phylo- and ontogenetically, learning of dense categories should be more universal than learning of sparse categories. In what follows, I review evidence exemplifying these points.

#### *4.1.1. Interactions between category structure and the learning system*

In a recent study (Kloos & Sloutsky, 2008), we demonstrated that category structure interacts with the learning system as well as with characteristics of the learner. In this study, 5-year-olds and adults were presented with a category learning task where they learned

either dense or sparse categories. These categories consisted of artificial bug-like creatures that had a number of varying features: sizes of tail, wings, and fingers; the shadings of body, antenna, and buttons; and the numbers of fingers and buttons (see Fig. 3, for examples of categories). Category learning was administered under either an unsupervised, spontaneous learning condition (i.e., participants were merely shown the items) or under a supervised, deliberate learning condition (i.e., participants were told the category inclusion rule). Recall that the former learning condition was expected to trigger the compression-based system of category learning, whereas the latter was expected to trigger the selection-based system. If category structure interacts with the learning system, then implicit, unsupervised learning should be more optimal for learning dense categories, whereas explicit, supervised learning should be more optimal for learning sparse categories. This is exactly what was found: For both children and adults, dense categories were learned better under the unsupervised, spontaneous learning regime, whereas sparse categories were learned more efficiently under the supervised learning regime. Critical data from this study are presented in Fig. 4. The figure presents categorization accuracy (i.e., the proportion of hits, or correct identification of category members minus the proportion of false alarms, or confusion of nonmembers for members) after the category learning phase.

These findings dovetail with results reported by Yamauchi, Love, and Markman (2002) and Yamauchi and Markman (1998) in adults. In these studies, participants completed a category learning task that had two learning conditions: classification and inference. In the classification condition, participants learned categories by predicting category membership of each study item. In the inference condition, participants learned categories by predicting a feature shared by category members. Across the conditions, results revealed a category structure by learning condition interaction. In particular, NLS categories (which are

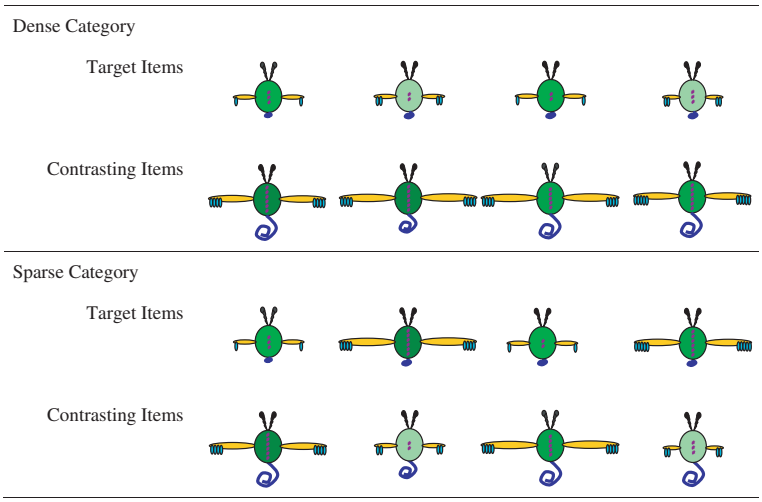


Fig. 3. Examples of items used from Kloos and Sloutsky (2008), Experiment 1. In the dense category, items are bound by similarity, whereas in the sparse category, the length of the tale is the predictor of the category membership.

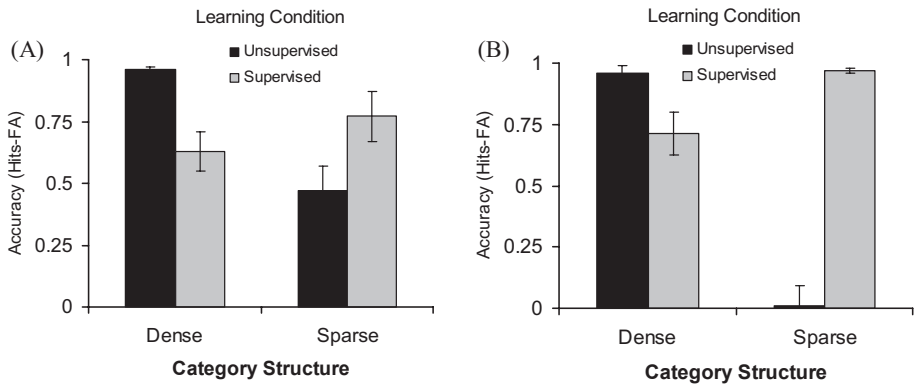


Fig. 4. Mean accuracy scores by category type and learning condition in adults (A) and in children (B). In this and all other figures, error bars represent standard errors of the mean. For the dense category, $D = 1$, and for the sparse category, $D = 0.17$.

typically sparser) were learned better in the classification condition, whereas prototype-based categories (which are typically denser) were learned better in the inference condition.

The interaction between the category structure and the learning system has been recently demonstrated by Hoffman and Rehder (2010), with respect to the cost of selectivity in category learning. Similar to Yamauchi and Markman (1998), participants learned categories either by classification or by feature inference. In the classification condition, participants were presented with two categories (e.g., A and B). On each trial, they saw an item and their task was to predict whether the item in question is a member of A or B. In the inference condition, participants were also presented with categories A and B. On each trial, they saw an item that had one missing feature and their task was to predict whether it was a feature common to A or common to B. In both conditions, upon responding, participants received feedback.

Each category had three binary dimensions whose values were designated as 0 or 1. There were two learning phases. In Phase 1, participants learned two categories A and B, with Dimensions 1 and 2 distinguishing between the categories and Dimension 3 being fixed across the categories (e.g., all items had a value of 0 on the fixed Dimension 3). In Phase 2, participants learned two other Categories C and D, with Dimensions 1 and 2 again distinguishing between the categories and Dimension 3 being fixed again (e.g., now items had a value of 1 on the fixed Dimension 3). After the two training phases, participants were given categorization trials involving contrasts between categories that were not paired during training (e.g., A vs. C). Note that correct responding on these novel contrasts required attending to Dimension 3, which had been previously irrelevant during training. If participants attend selectively to dimensions, their attention should have been allocated to Dimensions 1 and 2 during learning, which should have attenuated attention to Dimension 3. This attenuated attention represents the cost of selectivity. Alternatively, if no selectivity is involved, there should be little or no attenuation, and therefore, little or no cost. It was found that the cost was higher for classification learners than for inference learners, thus

suggesting that classification learning, but not inference learning, engages the selection-based system.

*4.1.2. Developmental primacy of the compression-based system*

Zentall et al. (2008) present an extensive literature review indicating that although birds, monkeys, apes, and humans are capable of learning categories consisting of highly similar yet discriminable items (i.e., dense categories), only some apes and humans could learn sparse relational categories, such as ''sameness'' when an equivalence class consisted of dissimilar items (e.g., a pair of red squares and a pair of blue circles are members of the same sparse category). However, even here it is not clear that subjects were learning a sparse category. As shown by Wasserman, Young, and Cook (2004), nonhuman animals readily distinguish situations with no variability in the input (i.e., zero entropy) from situations where input has stimulus variability (i.e., nonzero entropy). Therefore, it is possible that learning was based on the distinction between zero entropy in each of the ''same'' displays and nonzero entropy in each of the ''different'' displays.

The idea of the developmental primacy of the compression-based system is supported by data from Kloos and Sloutsky (2008) reviewed earlier. In particular, data presented in Fig. 4 clearly indicate that, for both children and adults, sparse categories were learned better under the explicit, supervised condition, whereas dense categories were learned better under the implicit, unsupervised condition. Also note that adults learned the sparse category even in the unsupervised condition, whereas young children exhibited no evidence of learning. These data support the contention that the compression-based system is the default in young children.

In addition, data from Kloos and Sloutsky (2008) indicate that although both children and adults exhibited able spontaneous learning of a dense category, there were marked developmental differences in spontaneous learning of sparse categories. Categorization accuracy in the spontaneous condition by category density and age is presented in Fig. 5. Two aspects of these data are worth noting. First, there was no developmental difference in spontaneous

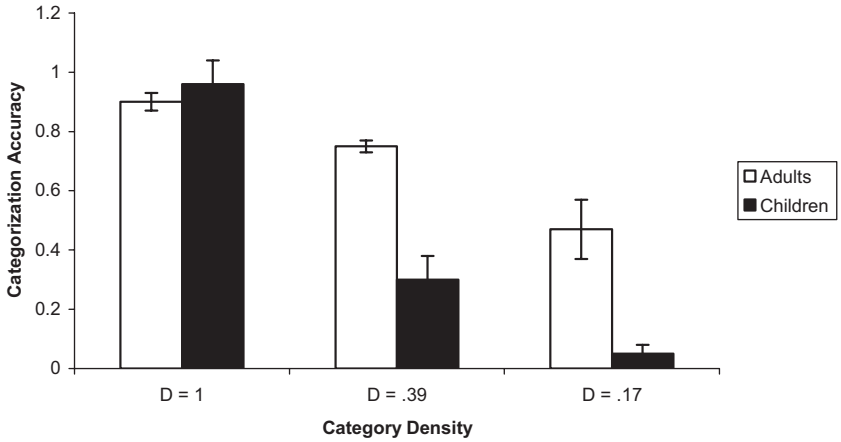


Fig. 5. Unsupervised category learning by density and age group from Kloos and Sloutsky (2008).

learning of the very dense category, which suggests that the compression-based system of category learning exhibits the adult level of functioning in 4- to 5-year-olds. And second, there were substantial developmental differences in spontaneous learning of sparser categories, which suggests that adults, but not young children, may spontaneously deploy the selection-based system of category learning. Therefore, the marked developmental differences pertain mainly to the deployment and functioning of the selection-based system, but not of the compression-based system (see also Hammer, Diesendruck, Weinshall, & Hochstein, 2009, for related findings).

Additional evidence for the developmental primacy of the compression-based learning system stems from research demonstrating that young children can learn complex contingencies implicitly, but not explicitly (Sloutsky & Fisher, 2008). The main idea behind the Sloutsky and Fisher (2008) experiments was that implicit (and perhaps compression-based) learning of complex contingencies might underlie seemingly selective generalization behaviors of young children. There is much evidence suggesting that, even early in development, people's generalization could be selective—depending on the situation, people may rely on different kinds of information. This selectivity has been found in a variety of generalization tasks, including lexical extension, categorization, and property induction. For example, in a lexical extension task (Jones et al., 1991), 2- and 3-year-olds were presented with a named target (i.e., ''this is a dax''), and then were asked to find another dax among test items. Children extended the label by shape alone when the target and test objects were presented without eyes. However, they extended the label by shape and texture when the objects were presented with eyes.

Similarly, in a categorization task, 3- and 4-year-olds were more likely to group items on the basis of color if the items were introduced as food, but group on the basis of shape if the items were introduced as toys (Macario, 1991). More recently, Opfer and Bulloch (2007) examined flexibility in lexical extension, categorization, and property induction tasks. It was found that across these tasks, 4- to 5-year-olds relied on one set of perceptual predictors when the items were introduced as ''parents and offspring,'' whereas they relied on another set of perceptual predictors when items were introduced as ''predators and prey.'' These findings pose an interesting problem—is this putative selectivity subserved by the selection-based system or by the compression-based system? Given critical immaturities of the selection-based system early in development, the latter possibility seems more plausible. Sloutsky and Fisher's (2008) study supported this possibility.

A key idea is that many stimulus properties intercorrelate, such that some clusters of properties co-occur with particular outcomes, and other clusters co-occur with different outcomes, thus resulting in a dense ''context–outcome'' structures (cf., with the idea of ''coherent covariation'' presented in Rogers & McClelland, 2004). Learning these correlations may result in differential allocation of attention to different stimulus properties in different situations or contexts, with flexible generalizations being a result of this learning. In particular, participants could learn the following set of contingencies: In Context 1, Dimension 1 (say, color) was predictive, but Dimension 2 (say, shape) was not, whereas the reverse is true in Context 2. If, as argued earlier, the system of implicit compression-based learning is fully functioning even early in development, then the greater the number of contextual

variables correlating with the relevant dimension (i.e., the greater the density), the greater the likelihood of learning. However, if learning is selection-based the reverse may be the case. This is because the larger the number of relevant dimensions, the more difficult it could be to formulate a contingency as a simple rule.

These possibilities have been tested in multiple experiments reported in Sloutsky and Fisher (2008). In these experiments, 5-year-olds were presented with triads of geometric objects differing in color and shape. Each triad consisted of a Target and two Test items. Participants were told that a prize was hidden behind the Target and their task was to determine the Test item that had a prize behind it. Children were trained that, in Context 1, shape of an item was predictive of an outcome, whereas in Context 2 color was predictive. Context was defined as the color of the background on which stimuli appeared and the location of the stimuli on the screen. Therefore, in Context 1, training stimuli appeared on a yellow background in the upper-right corner of the computer screen, and on a green background in the bottom-left corner of the computer screen in Context 2. Training stimuli were triads each consisting of a target and two test items. Participants were given information about a target item and they had to generalize this information to one of the test items. Each participant was given three training blocks. In one training block, only color was predictive, in another training block, only shape was predictive, whereas the third block was a mixture of the former two blocks. Participants were then presented with testing triads that had an important difference from training triads. Whereas training triads were ''unambiguous'' in that only one dimension of variation (either color or shape) was predictive and only one test item matched the target on the predictive dimension, this was not the case for testing triads. In particular, testing triads were ''ambiguous'' in that one test item matched the target on one dimension and the other test item matched the target on the other dimension. The only disambiguating factor was the context.

It was found that participants had no difficulty in learning the contingency between the context and the predictive dimension when there were multiple contextual variables correlating with the predictive dimension. In particular, children tested in Context 1 primarily relied on shape and those in Context 2 primarily relied on color. Learning, however, attenuated markedly when the number of contextual variables was reduced, which should not have happened if learning was selection based. And finally, when presented with testing triads and explicitly given a simple rule (e.g., children were asked to make choices by focusing either on color or on shape), they were unable to focus on the required dimension. These findings present further evidence for the developmental asynchrony of the two learning systems: Although 5-year-old children could readily perform the task when relying on the compression-based learning system, they were unable to perform the task when they had to rely on the selection-based system.

In sum, there is emerging body of evidence from category learning suggesting an interaction between the category structure and the learning system, pointing to developmental asynchronies in the two systems. Future research should reexamine category structure and category learning in infancy. In particular, given the critical immaturity of the selection-based system, most (if not all) of category learning in infancy should be accomplished by the compression-based system.

### 4.2. Category representation

In the previous section, I reviewed evidence indicating that category learning is affected by an interaction among category structure, the learning systems processing this structure, and the characteristics of the learner. In this section, I will review evidence demonstrating components of this interaction for category representation. Most of the evidence reviewed in this section pertains to developmental asynchronies between the learning systems. Two interrelated lines of evidence will be presented: (a) the development of selection-based category representation and (b) the changing role of linguistic label in category representation.

#### 4.2.1. The development of selection-based category representation

If the compression-based and the selection-based learning systems mature asynchronously, such that early in development the former system exhibits greater maturity than the latter, then it is likely that most of the spontaneously acquired categories are learned implicitly by the compression-based learning system. If this is the case, it is unlikely that young children form abstract rule-based representations of spontaneously acquired categories, whereas they are likely to form perceptually rich representations. A representation of a category is abstract if category items are represented by either a category inclusion rule or by a lexical entry. A representation of a category is perceptually rich if category representation retains (more or less fully) perceptual detail of individual exemplars.

One way of examining category representation is focusing on what people remember about category members. For example, Kloos and Sloutsky (2008, Experiment 4B) presented 5-year-olds and adults with a category learning task. Similar to the above-described experiment by Kloos and Sloutsky (2008), there were two between-subjects conditions, with some participants learning a dense category and some learning a sparse category. Both categories consisted of the described above artificial bug-like creatures that had a number of varying features: sizes of tail, wings, and fingers; the shadings of body, antenna, and buttons; and the numbers of fingers and buttons. The relation between the two latter features defined the arbitrary rule: Members of the target category had either many buttons and many fingers or few buttons and few fingers. All the other features constituted the appearance features. Members of the target category had a long tail, long wings, short fingers, dark antennas, a dark body, and light buttons (target appearance $A_T$), whereas members of the contrasting category had a short tail, short wings, long fingers, light antennas, a light body, and dark buttons (contrasting appearance $A_C$). All participants were presented with the same set of items; however, in the sparse condition participants' attention was focused on the inclusion rule, whereas in the dense condition it was focused on appearance information. This was achieved by varying the description of items across the conditions. In the sparse-category condition, the description was as follows: ''Ziblets with many aqua fingers on each yellow wing have many buttons, and Ziblets with few aqua fingers on each yellow wing have few buttons.'' In the dense-category condition, in addition to the above-described rule, the appearance of exemplars was described. In both conditions, appearance features were probabilistically related to

category membership, whereas the rule was fully predictive. After training, participants were tested on their category learning and then presented with a surprise recognition task. During the recognition phase, they were presented with four types of recognition items: $A_TR_T$ (the items that had both the appearance and the rule of the Target category), $A_CR_C$ (the items that had both the appearance and the rule of the Contrast category), $A_TR_C$ (the items that had the appearance of Target category and the rule of the Contrast category), and $A_CR_T$ (the items that had the appearance of the Contrast category and the rule of the Target category). If participants learned the category, they should accept $A_TR_T$ items and reject $A_CR_C$ items. In addition, if participants' representation of the category is based on the rule, they may false alarm on $A_CR_T$, but not on $A_TR_C$ items. However, if participants' representation of the category is based on the appearance, they should false alarm on $A_TR_C$, but not on $A_CR_T$ items.

False alarm rates by age and test item type are presented in Fig. 6. As can be seen in the figure, adults were more likely to false alarm on same appearance items (i.e., $A_TR_C$) in the dense condition and on same rule items (i.e., $A_CR_T$) in the sparse condition. In contrast, young children were likely to false alarm on same appearance items (i.e., $A_TR_C$) in both conditions. These results suggest that, in adults, dense and sparse categories could be represented differently: The former are represented perceptually, whereas the latter are represented more abstractly. At the same time, 5-year-old children are likely to represent perceptually both dense and sparse categories. These data suggest that the representation of sparse (but not dense) categories changes in the course of development.

These findings, however, were limited to newly learned categories that were not lexicalized. What about the representation of lexicalized dense categories? One possibility is that lexicalized dense categories are also represented perceptually, similar to newly learned dense categories. In this case, there should be no developmental differences in the representation of lexicalized dense categories. However, representations of lexicalized dense categories may include the linguistic label (which could be the most reliable guide to category membership). In particular, it is possible that lexicalization of a perceptual grouping eventually results in an abstract label-based representation (in the limit, a member of a category

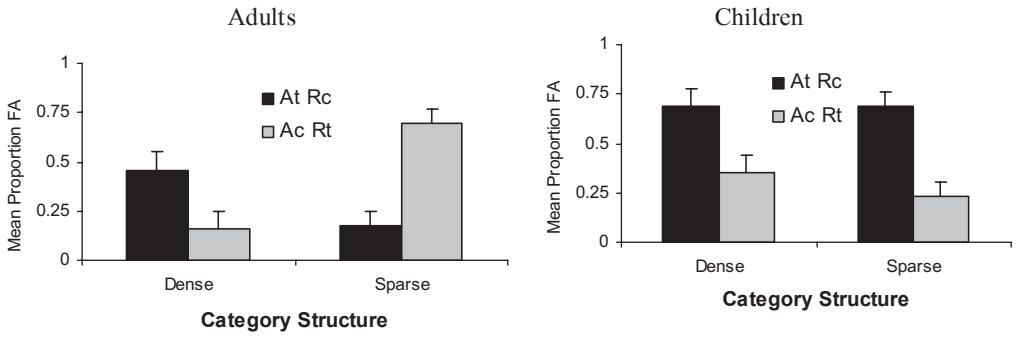


Fig. 6. False alarm rate by category structure and foil type in adults and children from Kloos and Sloutsky (2008), Experiment 4.

could be represented just by its label). If this is the case, then there should be substantial developmental differences in the representation of lexicalized dense categories. Furthermore, in this case, adults should differently represent highly familiar lexicalized dense categories (e.g., cat) and newly learned nonlexicalized dense categories (e.g., categories consisting of bug-like creatures). In particular, they should form an abstract representation of the former, but not the later.

These possibilities have been examined in a set of recognition memory experiments (e.g., Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2004a, 2004b). If participants form abstract representation of category items, then a task that prompts categorization of items may result in attenuated memory for appearance information. This reasoning is based on a long tradition of false memory research demonstrating that deep semantic processing of studied items (including grouping of items into categories) often increases memory intrusions—false recognition and recall of nonpresented ''critical lures'' or items semantically associated with studied items (e.g., Koutstaal & Schacter, 1997; Thapar & McDermott, 2001). Thus, ''deeper'' processing can lead to lower recognition accuracy when critical lures are semantically similar to studied items. In contrast to deep processing, focusing on perceptual details of pictorially presented information results in accurate recognition (Marks, 1991).

Therefore, if a recognition memory task is presented after a task that encourages access to the abstract representation of familiar categories, patterns of recognition errors may reveal information about how categories are represented. If participants processed items relatively abstractly as members of a category, then they would be more likely to have difficulty in discriminating studied targets from conceptually similar critical lures. If, on the other hand, they processed items more concretely, focusing on perceptual details, then they should discriminate relatively well.

In a set of experiments, Fisher and Sloutsky (2005) presented adults with one of two tasks. In the Baseline condition, the task was to remember items as accurately as possible, whereas in the Induction condition, the task was to generalize a property from a target item to each presented item. In both conditions, study phase items consisted of several categories, with multiple items per category. Following this study phase, participants in both conditions were presented with a surprise recognition task. Recognition items included Old Items (those presented during the Study phase), Critical Lures (novel items from studied categories), and Unrelated Items (novel items from new categories). If participants accept Old Items and Critical Lures, but reject Unrelated Items, then it is likely that they represented only abstract category information, not appearance information. However, if they accept only Old Items, but reject Critical Lures and Unrelated Items, then it is likely that they represented appearance information.

In one experiment reported by Fisher and Sloutsky (2005), adults were presented with familiar lexicalized dense categories (e.g., cats, bears, etc.), whereas in another condition, dense categories included artificial bug-like creatures, similar to those used by Kloos and Sloutsky (2008). Memory accuracy (which is a function of hits and false alarms on Critical Lures) by condition and category type in adults is presented in Fig. 7. Note that the dependent variable is A-prime (A-prime is a nonparametric analog of the signal-detection d-prime statistic), and the value of 0.5 represents no discrimination between Old Items and Critical

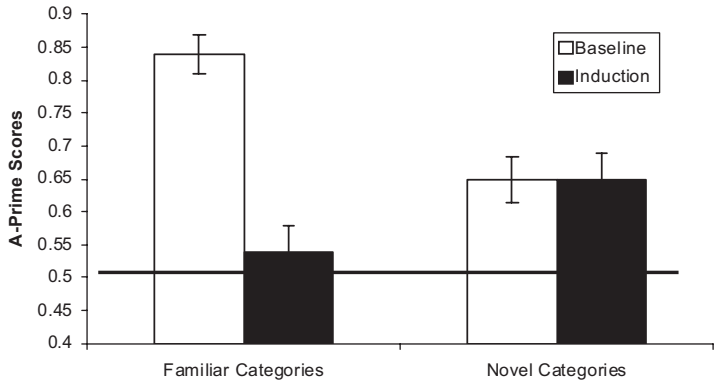


Fig. 7. Recognition accuracy in adults by category familiarity and study phase condition from Fisher and Sloutsky (2005).

Lures. When categories were familiar, adults were accurate in the Baseline condition, whereas they did not distinguish between Old Items and Critical Lures in the Induction condition. This *category processing effect* indicates that adults form a relatively abstract representation of familiar (and lexicalized) dense categories. It is also possible that category label plays an important role in such a representation (cf., findings reported by Tipper & Driver, 2000 on priming between pictures of objects and their labels in adults). At the same time, when categories were novel, adults were accurate in both the Baseline and Induction condition. Therefore, perceptual information plays an important role in the representation of novel dense categories in adults.

In contrast to adults, young children do not exhibit evidence of abstract representation of even familiar dense categories. As shown in Fig. 8, after performing induction with pictures

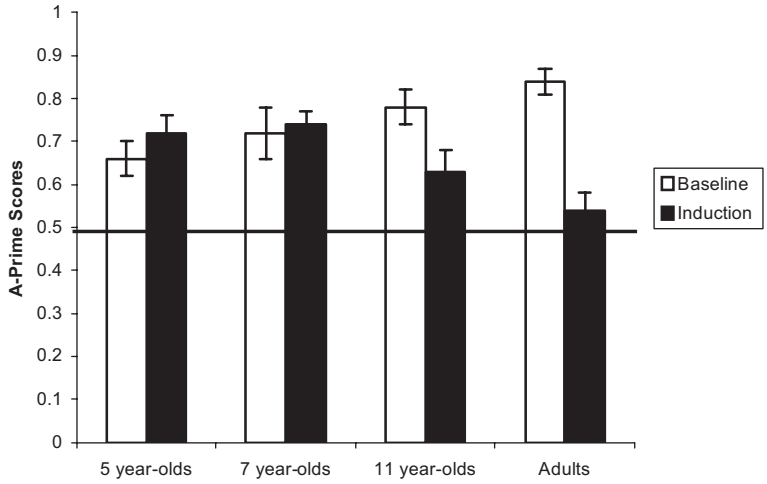


Fig. 8. Recognition accuracy by age and study phase condition from Fisher and Sloutsky (2005).

of members of familiar categories (e.g., cats), young children exhibited greater recognition accuracy than did adults, with recognition gradually decreasing with increasing age (Fisher & Sloutsky, 2005; Sloutsky & Fisher, 2004a, 2004b). The figure depicts A-prime scores across the conditions and the difference in A-prime score between the Baseline and Induction conditions reflects the ''category processing effect''—a decreased recognition of categorized items compared with the baseline. As shown in the figure, there is no evidence of the category processing effect early in development, and even in preadolescence the magnitude of the effect is smaller than that in adults. Recall that when adults were given the same task with novel items for which they did not have compressed category representation, their recognition accuracy increased to the level of young children (see Fig. 7).

These findings in conjunction with the relative immaturity of the executive function in 4- and 5-year-olds suggest that these participants, even if they learn a sparse rule-based category, would be unable to use this learned category in other tasks. It has been often argued that one of the most important roles of categories is to support inductive generalization. If one learns that an individual has a particular property (e.g., a particular dog likes bones), one could generalize this property to other members of this category. Although most transient properties (e.g., is awake) cannot be generalized, many stable properties can. Therefore, examining the pattern of inductive generalization could elucidate how categories are represented. If participants do not form an abstract representation of a sparse category, they would be unable to use the category in induction.

One way of addressing this issue is to teach participants a novel sparse rule-based category. Once participants learn the category, they could be presented with a property induction task, in which they could rely either on the rule or on appearance information, which is irrelevant for category membership. If young children represent the category by an abstract rule, they should use this representation when performing inductive generalization. Conversely, if they represent appearance of the items, then young children (even when they successfully learn the category) should rely on appearance information, while disregarding category membership information. These possibilities were tested in a set of experiments reported by Sloutsky, Kloos, and Fisher (2007). In these experiments, participants were first presented with a category learning task during which they learned two categories of artificial animals. Category membership was determined by a rule, whereas perceptual similarity was not predictive of category membership. Children were then given a categorization task with items that differed from those used during training. Participants readily acquired these categories and accurately sorted the items according to their category information. Then participants were presented with a triad induction task. Each triad consisted of a target and two test items, with one test item sharing the target's category membership, and the other test item being similar to the target (without sharing category membership). Participants were familiarized with a quasi-biological property of the target and asked to generalize this property to one of the test items. Finally, participants were given a final (i.e., postinduction) categorization task using the same items as the induction task. The results indicate that, although participants learned the category-inclusion rule, they did not use it in the course of induction, rather basing their induction on perceptual information.

In sum, early in development, similarity plays an important role in the representation of even sparse categories, whereas later in development categories may be represented in a more abstract manner. One possibility is that, later in development, labels begin to play a more central role in category representation.

*4.2.1.1 The developing role of linguistic labels in category representation:* In the previous section, I reviewed evidence that in young children (in contrast to adults) a category label does not figure prominently in category representation. This developmental change in the role of category labels represents another source of evidence for the developmental asynchronies between the two systems of category learning. In this section, I focus on the changing role of category labels in greater detail.

To examine the role of linguistic labels in category representation of adults, Yamauchi and colleagues conducted a series of studies supporting the idea that for adults a label is a symbol that represents a category (Yamauchi & Markman, 2000; Yamauchi & Yu, 2008). The overall reasoning behind this work is that if labels are category markers, they should be treated differently from the rest of features (such shape, color, size, etc.). However, this may not be the case if labels are features. Therefore, inferring a label when features are given (i.e., a classification task) should elicit different performance from a task of inferring a feature when the label is given (i.e., a feature induction task).

To test these ideas, Yamauchi and Markman (2000) used the above-described category learning task that was presented under either classification or feature induction learning condition. There were two categories, $C_1$ and $C_2$ denoted by two labels, $L_1$ and $L_2$. Stimuli were bug-like artificial creatures that varied on several dimensions, with one range of values determining $C_1$ and another range of values determining $C_2$. In the feature induction task, participants were shown a creature with one missing feature and were given a category label. Their task was to predict the missing feature. In the classification task, they were presented with a creature that was not labeled, and the task was to predict the category label. The critical condition was the case when an item was a member of $C_1$, but was similar to $C_2$, with the dependent variable being the proportion of $C_1$ responses. The results indicated that there were significantly more category-based responses in the induction condition (where participants could rely on the category label) than in the categorization condition (where participants had to infer the category label). It was concluded therefore that category labels differed from other features in that participants treated labels as category markers. These findings have been replicated in a series of follow-up studies (Yamauchi, Kohn, & Yu, 2007; Yamauchi & Yu, 2008; see also Markman & Ross, 2003, for a review). For example, Yamauchi et al. (2007) examined patterns of mouse-tracking (a procedure that is similar to eye tracking) to examine attention allocated to labels when labels were introduced as category markers (e.g., ''This is a dax'') or as denoting category features (e.g., ''This one has a dax''). Results indicated that participants viewed these visually presented labels more often in the former condition than in the latter condition. In sum, there is a body of evidence indicating that adults tend to treat the category label as a category marker rather than as a category feature.

However, the reliance on category labels in category representation requires the involvement of the selection-based system. At the same time, if the selection-based system exhibits

a slow developmental course, the ability to use category labels as category markers should be limited early in development. Furthermore, simultaneous processing of auditory and visual input (e.g., an object and corresponding sound) requires the ability to integrate information coming from different modalities. This ability also exhibits a relatively slow maturational course (see Robinson & Sloutsky, 2010, for a review) and is unlikely to be fully functional in infancy. In part, this slow maturational course in the ability to integrate cross-modal information could be related to a slow maturational course of neurons processing multisensory information. For example, there is evidence from animal models indicating that multisensory neurons located in the superior colliculus and at various cortical locations do not mature until the sufficient visual experience is accumulated (see Wallace, 2004, for a review).

If the contribution of labels to categorization and category learning hinges on (a) the ability to process cross-modal information and (b) the ability to attend selectively, with both abilities undergoing substantial developmental change, then the role linguistic labels play in categorization and category learning may change across development. In what follows, I review evidence indicating the changing role of category labels and consider possible mechanisms underlying these developmental changes.

As my colleagues and I have argued elsewhere, auditory input may affect attention allocated to corresponding visual input (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003; Sloutsky & Robinson, 2008), and these effects may change in the course of learning and development. In particular, linguistic labels may strongly interfere with visual processing in prelinguisitic children, but these interference effects may weaken when children start acquiring language (Sloutsky & Robinson, 2008; see also Robinson & Sloutsky, 2007a, 2007b).

In one experiment, Sloutsky and Robinson (2008) familiarized 10- and 16-month-olds with auditory–visual compounds. The familiarization compound consisted of a three-shape pattern and a word presented at the same time (both the word and the three-shape pattern were ably processed by infants of these age groups when presented unimodally). The familiarization phase was followed by the test phase, in which participants were presented with four different auditory–visual test items. One test item was the familiarization compound ($AUD_{Target}VIS_{Target}$), one had a changed visual component ($AUD_{Target}VIS_{New}$), one had a changed auditory component ($AUD_{New}VIS_{Target}$), and one had both components changed ($AUD_{New}VIS_{New}$).

The dependent variable was looking time at each test item. If participants considered a test item to be different from the familiarization item, looking time to this item should increase compared with the end of familiarization. Because the $AUD_{Target}VIS_{Target}$ is the familiarization item, it should elicit looking that is comparable with looking at the end of familiarization phase. Because the $AUD_{New}VIS_{New}$ is a novel item, it should elicit longer looking. At the same time, looking at $AUD_{Target}VIS_{New}$ and $AUD_{New}VIS_{Target}$ items should depend on whether participants processed auditory and visual components of the familiarization compound. If infants did, they should increase looking to both test items. If infants processed only the auditory component, they should increase looking only to $AUD_{New}VIS_{Target}$ item, whereas if they processed only the visual component, they should

increase looking only to $AUD_{Target}VIS_{New}$ item. Looking times to $AUD_{Target}VIS_{New}$, $AUD_{New}VIS_{Target}$, and $AUD_{New}VIS_{New}$ items compared with the $AUD_{New}VIS_{Target}$ item are presented in Fig. 9. These results clearly indicate that although 10-month-old infants failed to process the visual component, 16-month-old infants processed both components. It was concluded therefore that linguistic input interfered with processing of visual input at 10 months of age, but these interference effects weakened by 16 months of age.

In another experiment, Robinson and Sloutsky (2007a) presented 8- and 12-month-olds with a categorization task. Participants were familiarized with category exemplars under one of the three conditions: (a) all items were accompanied by the same label, (b) all items were accompanied by the same sound, or (c) all items were presented in silence. At test, participants were presented with two types of test trials: (a) recognition trials (i.e., a studied item was paired with a new item) and (b) categorization trials (i.e., a novel in-category exemplar was paired with a novel out-of-category exemplar). If participants recognize the studied item, they should prefer looking to the novel item, and if they learned the category, they should prefer looking to an out-of-category item. Results indicated that performance was significantly better in the silent condition, thus suggesting that both sounds and labels interfered with the categorization task. Similar results were reported for individuation tasks (Robinson & Sloutsky, 2008).

By the onset of word learning, children should start acquiring the ability to integrate linguistic and visual input (Robinson & Sloutsky, 2007b; Sloutsky & Robinson, 2008). However, even then cross-modal processing may not reach the full level of maturity and therefore linguistic labels may attenuate the processing of corresponding visual input. As

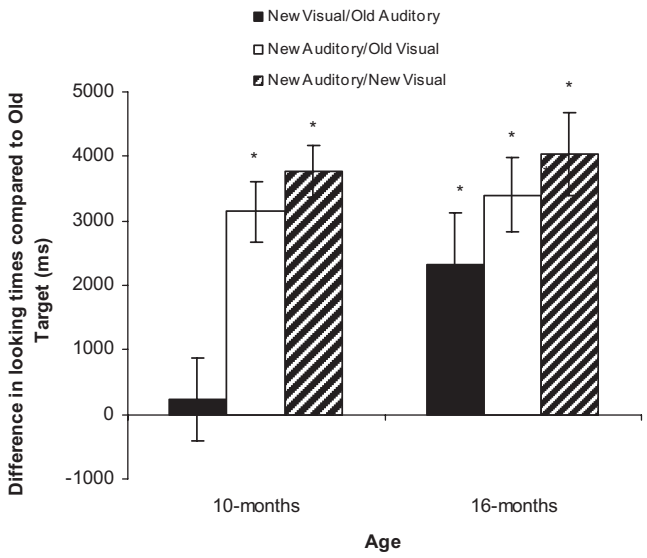


Fig. 9. Differences in looking times by Age and Test item type from Sloutsky and Robinson (2008). *Difference scores >0, $p < .05$.

discussed below, this attenuated processing may result in an increased similarity of entities that have the same label and thus in an increased tendency to group them together (e.g., Sloutsky & Fisher, 2004a; Sloutsky & Lo, 1999; Sloutsky, Lo, & Fisher, 2001).

Although interference effects attenuate with development, they do not disappear completely. This issue has been examined in depth in a series of recognition experiments (e.g., Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004; Sloutsky & Napolitano, 2003).

In these recognition experiments, 4-year-olds and adults were presented with a compound Target stimulus, consisting of simultaneously presented auditory and visual components ($AUD_{Target}VIS_{Target}$). These experiments were similar to the above-described experiment, except that no learning was involved. Participants were presented with a Target, which was followed immediately by a Test item and the task was to determine whether the Target and Test items were exactly the same.

There were four types of test items: (a) $AUD_{Target}VIS_{Target}$, which was the Old Target item; (b) $AUD_{Target}VIS_{New}$, which had the target auditory component and a new visual component; (c) $AUD_{New}VIS_{Target}$, which had the target visual component and a new auditory component; or (d) $AUD_{New}VIS_{New}$, which had a new visual component and a new auditory component. The task was to determine whether each presented test item was exactly the same as the Target (i.e., both the same auditory and visual components) or a new item (i.e., differed on one or both components).

Similar to the experiment with infants (Robinson & Sloutsky, 2004), it was reasoned that if participants process both auditory and visual stimuli, they should correctly respond to all items by accepting Old Target items and rejecting all other test items. Alternatively, if they fail to process the visual component, they should falsely accept $AUD_{Target}VIS_{New}$ items, while correctly responding to other items. Finally, if they fail to process the auditory component, they should falsely accept $AUD_{New}VIS_{Target}$ items, while correctly responding to other items. In one experiment (Napolitano & Sloutsky, 2004), speech sounds were paired with either geometric shapes or pictures of unfamiliar animals. Results indicated that although children ably processed either stimulus in the unimodal condition, they failed to process visual input in the cross-modal condition. Furthermore, a yet unpublished study by Napolitano and Sloutsky indicates that interference effects attenuate gradually in the course of development, with very little evidence of interference in adults.

There is also evidence that this dominance of auditory input is not under strategic control: Even when instructed to focus on visual input young children had difficulties doing so (Napolitano & Sloutsky, 2004; Robinson & Sloutsky, 2004). In one of the experiments described in Napolitano and Sloutsky (2004), 4-year-olds were explicitly instructed to attend to visual stimuli, with instructions repeated before each trial. However, despite the repeated explicit instruction to attend to visual stimuli, 4-year-olds continued to exhibit auditory dominance. These results suggest that auditory dominance is unlikely to stem from deliberate selective attention to a particular modality, but it is more likely to stem from automatic pulls on attention.

If linguistic labels attenuate visual processing, such that children ably process a label, but they do so to a lesser extent the corresponding visual input, then these findings can explain the role of labels in categorization tasks. In particular, items that share a label may appear

more similar than the same items presented without a label. In other words, early in development, labels may function as features contributing to similarity, and their role may change in the course of development. In fact, there is evidence supporting this possibility (e.g., Sloutsky & Fisher, 2004a; Sloutsky & Lo, 1999).

The key idea behind these experiments is if two items have a particular degree of visual similarity, then adding a common label would increase this similarity due to the above-described attenuated visual processing. These effects have been demonstrated with a frequently used forced choice task, where participants are expected to make either a similarity judgment (i.e., which one of the several test items looks more like the target) or a categorization judgment (i.e., which one of the several test items belongs to the same kind as the target).

In this case, the probability of selecting a particular test item is a function of a *ratio* of the similarity of a given test item to the Target to the summed similarity of other test items to the Target. In this case, the common label affects the similarity ratio. These ideas have been implemented in model SINC (for Similarity, Induction, Naming, and Categorization; Sloutsky & Lo, 1999; Sloutsky & Fisher, 2004a) that accurately predicted similarity and categorization judgment in young children when labels were and were not introduced.

In these experiments, young children were presented with triads of items (a Target and two Test items) and were asked which of the Test items looked more similar to the Target. One of the test items (e.g., Test A) was very similar to the Target, whereas similarity of the other test item (say Test B) varied across trials from very similar to very different. In the Baseline condition, labels were not provided, whereas in the Label condition, one of the Test items shared the label with the Target, whereas the other Test item did not. The labels were artificial bisyllabic count nouns. Proportions of selecting Test B as more similar to the Target by condition and similarity ratio (Test B–Target/Test A–Target) are presented in Fig. 10A. As can be seen in the figure, the presence of labels increased similarity for all levels of similarity. However, when the same task was given to adults (Fig. 10B), labels had no effect on similarity judgment.

Therefore, it seems that labels function differently across development: Whereas labels are likely to contribute to similarity of compared items in children (e.g., Sloutsky & Fisher,

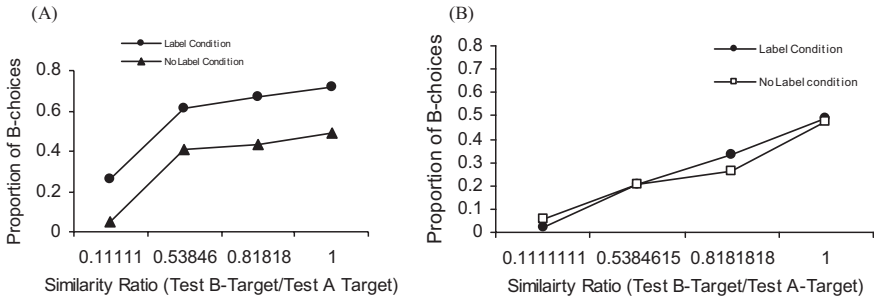


Fig. 10. Similarity judgment by similarity ratio and labeling condition in (A) children and (B) adults from Sloutsky and Fisher (2004a).

2004a; Sloutsky & Lo, 1999), they are not likely to do so in adults (Yamauchi & Markman, 2000).

There is also evidence that labels have similar effects on categorization—these effects are also graded rather than rule-like, with labels affecting, but not overriding perceptual similarity (e.g., Sloutsky & Fisher, 2004a). In several experiments conducted by Sloutsky and Fisher, 4- and 5-year-olds performed a match-to-sample categorization task. On each trial, they were presented with a triad of pictures, a target, and two test items. All items were labeled and only one of the test items shared the label with the target. Participants were asked to decide which of the test items belongs to the same kind as the Target. Strikingly similar patterns were observed for categorization and feature induction tasks in young children: Again, participants' categorization and induction responses were affected by the similarity ratio, with labels contributing to these effects of similarity rather than overriding them.

In yet another experiment, Sloutsky and Fisher (2004a, 2004b) used items that had been previously used by Gelman and Markman (1986), which turned out to vary widely in terms of appearance similarity. Again, there was little evidence that, in their induction responses, 4- and 5-year-olds relied exclusively on linguistic labels.

In short, the reviewed evidence supports the idea that young children treat labels as perceptual features that contribute to similarity of compared entities. It seems that these effects of labels stem from critical immaturities of cross-modal processing coupled with immaturities of selective attention. Further development of cross-modal processing and the selection-based system, coupled with acquired knowledge that a category label is highly predictive of category membership, may result in category labels becoming category markers in adults (e.g., Yamauchi & Markman, 2000; Yamauchi & Yu, 2008; see also Markman & Ross, 2003). However, additional research is needed to establish a detailed understanding of the changing role of linguistic labels in category representation.

### *4.3. Summary*

In this section, I considered interactions among category structure, the learning system, and characteristics of the learner in category learning and category representation. First, I reviewed evidence demonstrating that dense categories could be learned efficiently by the compression-based system, whereas sparse categories require the involvement of the selection-based system. Second, although the compression-based system exhibits able functioning even early in development, the selection-based system undergoes developmental transformations. As a result, early in development learning subserved by the compression-based system exhibits greater efficiency than learning subserved by the selection-based system. Third, representation of sparse categories changes in the course of development: Although adults form an abstract representation of sparse categories, young children form similarity-based representations of sparse categories. Fourth, there are developmental differences in the representation of dense lexicalized categories: Adults, but not young children, can represent these categories abstractly. And finally, there is evidence that the role of category labels in category representation changes in the course of development; not until late

in development do labels become category markers (although see Waxman & Markow, 1995; Welder & Graham, 2001; Xu, 2002).

## 5. Conceptual development: From perceptual categories to abstract concepts

On the basis of the formulated characteristics of the input, of the learning systems, and of the learner, we can propose a rough sketch of how conceptual development proceeds. The early functioning of the compression-based system suggests that even young infants should ably learn dense perceptual categories. The ability to learn perceptual categories from relatively dense input has been demonstrated in nonhuman animals as well as in 3- and 4-month-old human infants (Cook & Smith, 2006; Quinn et al., 1993; Smith et al., 2008; Zentall et al., 2008). Although some of these perceptual categories (e.g., cats, dogs, or food) will undergo lexicalization, others (e.g., some categories of speech sounds) will not.

The next critical step is the development of the ability to integrate cross-modal information that may subserve word learning and learning of dense cross-modal categories. There is evidence that very young infants have difficulty in integrating input coming from different modalities, unless both modalities express the same amodal relation (e.g., when the same amodal relation [such as rhythm or rate] is presented cross-modally, cross-modal presentation is likely to facilitate processing of the amodal relation [see Lewkowicz, 2000; Lickliter & Bahrick, 2000, for reviews]). Initially the sensory systems are separated from one another, with multisensory integration being a product of development and learning. There is much recent neuroscience evidence pointing to slow postnatal maturation of multisensory neurons, coupled with slow maturation of functional corticotectal connections (see Wallace, 2004, for a review). Cross-modal integration is at the heart of the ability to learn cross-modal perceptual categories, which permeate early experience (e.g., dogs bark, cats meow, and humans speak).

Once the ability to integrate cross-modal information is somewhat functional, infants can start learning words, which requires binding auditory and visual input. However, given the immaturity of cross-modal processing, it is easier to learn words that denote perceptual categories that the child already knows. Furthermore, infants may spontaneously learn categories of items that are frequent in their environment and these categories would be the first to be labeled by parents. There is evidence (e.g., Nelson, 1973) that the most frequent type of words among the first 100 words produced by babies is a count noun, with most of these count nouns denoting perceptual categories of entities in the child's environment. Therefore, learning the first words could be a way of lexicalizing those perceptual categories that the child already learned. Lexicalization also opens the possibility of acquiring knowledge of unobservable properties about category members, as well as generalizing this knowledge. Unobservable information includes properties that one does not typically observe (e.g., that one's pet dog has a heart) as well as properties that cannot be observed in principle, but have to be inferred from the observed properties (e.g. ''that another person has thoughts and feelings''). Once acquired, these unobservable properties can be entered into the computation of similarity, thus enabling the development of more abstract superordinate categories.

Therefore, lexicalization is a critical step in the transition from perceptual groupings to concepts. The ability to process cross-modal input also enables children to use a combination of perceptual and linguistic cues in acquiring broad ontological distinctions (Jones & Smith, 2002; Samuelson & Smith, 1999; Yoshida & Smith, 2003).

The next important step is learning of dimensional words, denoting dimensional values (e.g., ''green'' or ''square''). Learning of these words coupled with further maturation of the PFC and the development of executive function may result in lexicalization of some stimulus dimensions (such as color, shape, or size). As argued by many researchers (Carey, 1982; Gasser & Smith, 1998), learning of dimensional words follows learning of count nouns. One explanation is that perceptual groupings, such as ''dog'' or ''cup,'' denoted by count nouns are dense—they are based on an intercorrelated set of features and feature dimensions. In contrast, dimensional groupings (e.g., ''red things'') are sparse. Therefore, the later, but not the former, requires selective attention, which appears later in development than the ability to learn perceptual groupings and to integrate cross-modal information.

Further development of the PFC coupled with learning of abstract words lays the foundation for the development of abstract concepts. However, unlike their concrete counterparts (such as ''dog'' or cup'') where category learning may precede word learning, there are reasons to believe that words denoting abstract concepts are learned prior to the concept itself (e.g., Vygotsky, 1964). For example, according to the MacArthur Lexical Development Norms (Dale & Fenson, 1996) a 30-month-old toddler may produce words, such as *love*, *time*, and *same*; however, it is unlikely that these children have concepts of LOVE, TIME, or EQUIVALENCE. Furthermore, because these abstract concepts refer to exceedingly sparse categories, it is likely that the acquisition of these categories requires supervision. The relative maturity of the PFC is of critical importance because learners need to focus on a small set of category-relevant features, while ignoring irrelevant features. The ability to lexicalize categories and the ability to acquire abstract concepts paves the way to acquisition of abstract mathematical and scientific concepts. However, some of these concepts are so sparse and put so much demand on selectivity that supervision alone may not be sufficient—and sophisticated explicit instruction is needed—for successful learning of these concepts (e.g., Kaminski, Sloutsky, & Heckler, 2008).

In sum, the proposal presented here attempts to connect conceptual development with the structure of input and the availability of the learning system necessary for processing of this input. This rough sketch, however, is just a first step in uncovering the great mystery of conceptual development—a progression from a newborn who has difficulty in perceiving the world to an adult who has the ability of changing the world.

## 6. Concluding comments

In this study, I considered the possibility of conceptual development progressing from simple perceptual grouping to highly abstract scientific concepts. I reviewed evidence suggesting that conceptual development is a product of an interaction of the structure of input,

the category learning system that processes this input, and maturational characteristics of the learner.

I also considered three steps that are critical for conceptual development. First, the development of the selection-based system of category learning that depends critically on the maturation of cortical regions subserving executive function. The second critical step is the ability to integrate cross-modal information. This ability is critical for word learning and lexicalization of spontaneously acquired perceptual groupings, as well as for forming broad ontological classes. And the third critical step, depending on the former two, is the ability to learn and use abstract categories. Unlike their concrete counterparts that can be acquired by perceptual means and lexicalized later, for learning of some abstract categories lexicalization might be a prerequisite.

The proposal presented here considers a complex developmental picture that depends on a combination of maturational and experience factors in conceptual development. Under this view, learning of perceptual categories, cross-modal integration, lexicalization, learning of conceptual properties, the ability to focus and shift attention, and the development of lexicalized concepts are logical steps in conceptual development. This proposal offers a theoretical alternative to the idea of innate knowledge structures specific to various knowledge domains. However, much research is needed to move from a rough sketch to detailed understanding of conceptual development.

## Note

1. For the moment, I will ignore a relatively small class of abstract concepts—‘‘electron’’ would be a good example—that start out as a lexical entry. However, I will return to this issue later in the study.

## Acknowledgments

Writing of this article was supported by grants from the NSF (BCS-0720135); the Institute of Education Sciences, U.S. Department of Education (R305B070407); and NIH (R01HD056105). The opinions expressed are those of the authors and do not represent views of the awarding organizations. I thank Catherine Best, Anna Fisher, Rubi Hammer, Susan Johnson, John Opfer, and Chris Robinson for helpful comments.

## References

Alvarado, M. C., & Bachevalier, J. (2000). Revisiting the maturation of medial temporal lobe memory functions in primates. *Learning & Memory*, *7*, 244–256.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.

Ashby, F. G., & Maddox, T. W. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.

Bar-Gad, I., Morris, G., & Bergman, H. (2003). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology*, *71*, 439–473.

Berg, E. A. (1948). A simple objective test for measuring flexibility and thinking. *Journal of General Psychology*, *39*, 15–22.

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. *Cognition*, *112*, 330–336.

Brown, R. G., & Marsden, C. D. (1988). Internal versus external cues and the control of attention in Parkinson's disease. *Brain*, *111*, 323–345.

Buffalo, E. A., Ramus, S. J., Clark, R. E., Teng, E., Squire, L. R., & Zola, S. M. (1999). Dissociation between the effects of damage to perirhinal cortex and area TE. *Learning & Memory*, *6*, 572–599.

Bunge, S. A., & Zelazo, P. D. (2006). Brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science*, *15*, 118–121.

Carey, S. (1982). Semantic development: State of the art. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 347–389). Cambridge, England: Cambridge University Press.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Carey, S., & Spelke, E. (1994). Domain specific knowledge and conceptual change. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 169–200). Cambridge, MA: Cambridge University Press.

Carey, S., & Spelke, E. (1996). Science and core knowledge. *Philosophy of Science*, *63*, 515–533.

Caviness, V. S., Kennedy, D. N., Richelme, C., Rademacher, J., & Filipek, P. A. (1996). The human brain age 7–11 years: A volumetric analysis based on magnetic resonance images. *Cerebral Cortex*, *6*, 726–736.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131–1157.

Chomsky, N. (1980). *Rules and representations*. Oxford, England: Blackwell.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via observation and via feedback. *Journal of Cognitive Neuroscience*, *19*, 249–265.

Cook, R. G., & Smith, J. D. (2006). Stages of abstraction and exemplar memorization in pigeon category learning. *Psychological Science*, *17*, 1059–1067.

Cools, A. R., van den Bercken, J. H. L., Horstink, M. W. I., van Spaendonck, K. P. M., & Berger, H. J. C. (1984). Cognitive and motor shifting aptitude disorder in Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry*, *47*, 443–453.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125–127.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). London, UK: Oxford University Press.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, *44*, 24–40.

van Domburg, P. H. M. E., & ten Donkelaar, H. J. (1991). *The human substantia nigra and ventral tegmental area*. Berlin: Springer-Verlag.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*, 340–347.

Fernandez-Ruiz, J., Wang, J., Aigner, T. G., & Mishkin, M. (2001). Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. *Proceedings of the National Academy of Sciences*, *98*, 4196–4201.

Fisher, A. V. (2007). Are developmental theories of learning paying attention to attention? *Cognition, Brain, and Behavior*, *11*, 635–646.

Fisher, A. V., & Sloutsky, V. M. (2005). When induction meets memory: Evidence for gradual transition from similarity-based to category-based induction. *Child Development*, *76*, 583–597.

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*, 382–397.

Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes*, *13*, 269–306.

Gelman, S. A. (1988). The development of induction within natural kind and artifact categories. *Cognitive Psychology*, *20*, 65–95.

Gelman, R. (1990). Structural constraints on cognitive development: Introduction to a special issue of Cognitive Science. *Cognitive Science*, *14*, 3–10.

Gelman, S. A., & Coley, J. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). New York: Cambridge University Press.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Giedd, J. N., Snell, J. W., Lange, N., Rajapakse, J. C., Casey, B. J., Kozuch, P. L., Vaituzis, A. C., Vauss, Y. C., Hamburger, S. D., Kaysen, D., & Rapoport, J. L. (1996a). Quantitative magnetic resonance imaging of human brain development: Ages 4–18. *Cerebral Cortex*, *6*, 551–560.

Giedd, J. N., Vaituzis, A. C., Hamburger, S. D., Lange, N., Rajapakse, J. C., Kaysen, D., Vauss, Y. C., & Rapoport, J. L. (1996b). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: Ages 4–18 years. *Journal of Comparative Neurology*, *366*, 223–230.

Goldman-Rakic, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development*, *58*, 601–622.

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language*, *21*, 125–155.

Gureckis, T. M., & Love, B. C. (2004). Common mechanisms in infant and adult category learning. *Infancy*, *5*, 173–198.

Hammer, R., & Diesendruck, G. (2005). The role of dimensional distinctiveness in children's and adults' artifact categorization. *Psychological Science*, *16*, 137–144.

Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: What makes the difference? *Cognition*, *112*, 105–119.

Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*, 319–340.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

Jernigan, T. L., Trauner, D. A., Hesselink, J. R., & Tallal, P. A. (1991). Maturation of the human cerebrum observed in vivo during adolescence. *Brain*, *114*, 2037–2049.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*, 219–232.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*, 499–516.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science*, *230*, 454–455.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science*, *6*, 449–476.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*, 52–72.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science*, *273*, 1399–1402.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory & Language*, *37*, 555–583.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K. (1993). Human category learning: Implications for back propagation models. *Connection Science*, *5*, 3–36.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*, 812–863.

Lewkowicz, D. J. (2000). Development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*, 281–308.

Lickliter, R., & Bahrick, L. E. (2000). The development of infant intersensory perception: Advantages of a comparative convergent-operations approach. *Psychological Bulletin*, *126*, 260–280.

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., & Hommer, D. W. (1999). Wisconsin Card Sorting Test performance following head injury: Dorsolateral fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology*, *21*, 2–16.

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 90–108.

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four- to eight-year-old children. *Neuropsychologia*, *36*, 273–293.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*, 17–46.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

Mandler, J. B., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, *23*, 263–298.

Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science*, *26*, 377–389.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: MIT Press.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental and Theoretical Artificial Intelligence*, *13*, 329–358.

Marks, W. (1991). Effects of encoding the perceptual features of pictures on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 566–577.

Napolitano, A. C., & Sloutsky, V. M. (2004). Is a picture worth a thousand words? The flexible nature of modality dominance in young children. *Child Development*, *75*, 1850–1870.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38*(1–2), 1–137.

Nomura, E. M., Maddox, W. T., Filoteo, J. V., Ing, A. D., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2007). Neural correlates of rule-based and information-integration visual category learning. *Cerebral Cortex*, *17*, 37–43.

Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews*, *32*, 279–291.

Nosofsky, R. M. (1986). Attention, similarity and the identification categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Opfer, J. E., & Bulloch, M. J. (2007). Causal relations drive young children's induction, naming, and categorization. *Cognition*, *105*, 206–217.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology*, *49*, 301–332.

Pfefferbaum, A., Mathalon, D. H., Sullivan, E. V., Rawles, J. M., Zipursky, R. B., & Lim, K. O. (1994). A quantitative magnetic resonance imaging study of changes in brain morphology from infancy to late adulthood. *Archives of Neurology*, *51*, 874–887.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

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*, 463–475.

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate-inanimate distinction. *Psychological Bulletin*, *127*, 209–228.

Rao, S. M., Bobholz, J. A., Hammeke, T. A., Rosen, A. C., Woodley, S. J., Cunningham, J. M., Cox, R. W., Stein, E. A., & Binder, J. R. (1997). Functional MRI evidence for subcortical participation in conceptual reasoning skills. *NeuroReport*, *8*, 1987–1993.

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*, 1387–1401.

Robinson, C. W., & Sloutsky, V. M. (2007a). Linguistic labels and categorization in infancy: Do labels facilitate or hinder? *Infancy*, *11*, 233–253.

Robinson, C. W., & Sloutsky, V. M. (2007b). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, *10*, 734–740.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. *Developmental Science*, *11*, 869–881.

Robinson, C. W., & Sloutsky, V. M. (2010). Development of cross-modal processing. *WIRES: Cognitive Science*, *1*, 1–7.

Rodman, H. R. (1994). Development of inferior temporal cortex in the monkey. *Cerebral Cortex*, *4*, 484–498.

Rodman, H. R., Skelly, J. P., & Gross, C. G. (1991). Stimulus selectivity and state dependence of activity in inferior temporal cortex of infant monkeys. *Proceedings of the National Academy of Sciences*, *88*, 7572–7575.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, *12*, 142–162.

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

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rueda, M., Fan, J., McCandliss, B. D., Halparin, J., Gruber, D., Lercari, L., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*, 27–52.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, *73*, 1–33.

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? Their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, *32*, 265–278.

Seger, C. A., & Cincotta, C. M. (2002). Striatal activation in concept learning. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 149–161.

Shannon, C. E., & Weaver, W. (1948). *The mathematical theory of communication*. Chicago: University of Illinois Press.

Shepp, B. E., & Swartz, K. B. (1976). Selective attention and the processing of integral and nonintegral dimensions: A developmental study. *Journal of Experimental Child Psychology*, *22*, 73–85.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*, 246–251.

Sloutsky, V. M., & Fisher, A. V. (2004a). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*, 166–188.

Sloutsky, V. M., & Fisher, A. V. (2004b). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*, *15*, 553–558.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*, 639–651.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*, 179–185.

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1: Linguistic labels and the development of similarity judgment. *Developmental Psychology*, *35*, 1478–1492.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. (2001). How much does a shared name make things similar? Linguistic labels, similarity and the development of inductive inference. *Child Development*, *72*, 1695–1709.

Sloutsky, V. M., & Napolitano, A. C. (2003). Is a picture worth a thousand words? Preference for auditory modality in young children. *Child Development*, *74*, 822–833.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in visual processing: From overshadowing to attentional tuning. *Cognitive Science*, *32*, 354–377.

Sloutsky, V. M., & Spino, M. A. (2004). Naive theory and transfer of learning: When less is more and more is less. *Psychonomic Bulletin and Review*, *11*, 528–535.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*, 125–144.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, *60*, 143–171.

Smith, J. D., & Kemier-Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General*, *113*, 137–159.

Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *137*, 390–401.

Soja, N., Carey, S., & Spelke, E. (1991). Ontological categories guide young children's inductions of word meanings: Object terms and substance terms. *Cognition*, *38*, 179–211.

Sowell, E. R., & Jernigan, T. L. (1999). Further MRI evidence of late brain maturation: Limbic volume increases and changing asymmetries during childhood and adolescence. *Developmental Neuropsychology*, *14*, 599–617.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Batth, R., Jernigan, T. L., & Toga, A. W. (1999). Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage*, *9*, 587–597.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, *2*, 859–861.

Spelke, E. S. (2000). Core knowledge. *American Psychologist*, *55*, 1233–1243.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Striedter, G. F. (2005). *Principles of brain evolution*. Sunderland, MA: Sinauer.

Teng, E., Stefanacci, L., Squire, L. R., & Zola, S. M. (2000). Contrasting effects on discrimination learning after hippocampal lesions and conjoint hippocampal-caudate lesions in monkeys. *Journal of Neuroscience*, *20*, 3853–3863.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory and Cognition*, *29*, 424–432.

Tipper, S. P., & Driver, J. (2000). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. In M. S. Gazzaniga (Ed.), *Cognitive neuroscience: A reader* (pp. 176–187). Malden, MA: Blackwell Publishing.

Twyan, A. D., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, *34*(7), 1315–1356.

Vygotsky, L. S. (1964). *Thought and language*. Cambridge, MA: MIT Press. (Original work published in 1934)

Wallace, M. T. (2004). The development of multisensory processes. *Cognitive Processing*, *5*, 69–83.

Wasserman, E. A., Young, M. E., & Cook, R. G. (2004). Variability discrimination in humans and animals: Implications for adaptive action. *American Psychologist*, *59*, 879–890.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12-13-month-old infants. *Cognitive Psychology*, *29*, 257–302.

Welder, A. N., & Graham, S. A. (2001). The influence of shape similarity and shared labels on infants' inductive inferences about nonobvious object properties. *Child Development*, *72*, 1653–1673.

Whitman, J. R., & Garner, W. R. (1962). Free recall learning of visual figures as a function of form of internal structure. *Journal of Experimental Psychology*, *64*, 558–564.

Wilson, C. (1995). *The contribution of cortical neurons to the firing pattern of striatal spiny neurons*. Cambridge, MA: Bradford.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*, 223–250.

Yamauchi, T., Kohn, N., & Yu, N.-Y. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, *35*, 852–863.

Yamauchi, T., Love, B. C., & Markman, A. B. (2002). Learning nonlinearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 585–593.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 776–795.

Yamauchi, T., & Yu, N.-Y. (2008). Category labels versus feature labels: Category labels polarize inferential predictions. *Memory & Cognition*, *36*, 544–553.

Yoshida, H., & Smith, L. B. (2003). Shifting ontological boundaries: How Japanese- and English speaking children generalize names for animals and artifacts. *Developmental Science*, *6*, 1–34.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.

Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research on Child Development*, *68*, vii–137.

Zentall, T. R., Wasserman, E. A., Lazareva, O. F., Thompson, R. K. R., & Rattermann, M. J. (2008). Concept learning in animals. *Comparative Cognition & Behavior Reviews*, *3*, 13–45.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1287–1314

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01130.x

# Knowledge as Process: Contextually Cued Attention and Early Word Learning

Linda B. Smith,[a] Eliana Colunga,[b] Hanako Yoshida[c]

[a]*Department of Psychological and Brain Sciences, Indiana University*
[b]*Department of Psychology and Neuroscience, University of Colorado*
[c]*Department of Psychology, University of Houston*

Received 24 February 2009; received in revised form 24 June 2010; accepted 24 June 2010

**Abstract**

Learning depends on attention. The processes that cue attention in the moment dynamically integrate learned regularities and immediate contextual cues. This paper reviews the extensive literature on cued attention and attentional learning in the adult literature and proposes that these fundamental processes are likely significant mechanisms of change in cognitive development. The value of this idea is illustrated using phenomena in children's novel word learning.

*Keywords:* Cued attention; Development; Noun learning

## 1. Introduction

In her introduction to the 1990 special issue of *Cognitive Science*, Rochel Gelman asked, ''How is it that our young attend to inputs that will support the development of concepts they share with their elders?'' Gelman posed the question in terms of attention, but the answers offered in that volume were not about attention. Instead, they were about innate knowledge structures, so-called first or core principles that guide learning in specific knowledge domains. In the years since 1990, inspired in part by the highly influential papers in that special issue, there have been many *demonstrations* of the remarkable knowledge that quite young children bring to bear on learning. But there have been very modest advances in understanding the *processes and mechanisms* through which that knowledge is realized and applied to aid learning. Accordingly, in this paper we return to Gelman's original question: How do children select the right information for learning?

Correspondence should be sent to Linda B. Smith, Department of Psychological and Brain Sciences, Indiana University, 1101 East 10th Street, Bloomington, IN 47405. E-mail: smith4@indiana.edu

We consider this question in the context of how children learn words, one of the topics also of central interest in the original special issue. We begin by reviewing well-documented and general mechanisms of attentional learning from the adult literature from the perspective of their relevance as mechanisms of cognitive development. We then ask—mostly without direct empirical evidence on the answer—what role these known mechanisms might play in early word learning. Our main goal is to encourage researchers to pursue these mechanisms as significant contributors to word learning and to cognitive development more generally. This—taking well-documented mechanisms from one area of research and asking whether they might apply in another, moving the field toward a more unified understanding of what might seem at first unrelated domains—should not be contentious. However, the literature on early word learning is highly contentious (Booth & Waxman, 2002; Cimpian & Markman, 2005; Smith & Samuelson, 2006) and not in a productive way, as the explanations pitted against each other are not empirically resolvable in a straightforward way.

One word-learning principle, mutual exclusivity, discussed in the original special issue may help illuminate the problem. The phenomenon, which is not at issue, is this: Given a known thing with a known name (e.g., a cup) and a novel thing and told a novel name (''Where is the rif?'') children take the novel word to refer to the novel object and not the known one. Markman (1990, p. 66) explained this behavior as follows: ''children constrain word meaning by assuming at first that words are mutually exclusive—that each object can have one and only one label.'' This description summarizes the phenomenon in terms of the macro-level construct of ''assumptions.'' This construct, of course, may be unpacked into a number of micro-level processes, including, as we will propose here, cue competitions. An account in terms of cue competitions and attention is not in opposition to an account in terms of assumptions because the two kinds of explanations are at fundamentally different levels of analysis that answer different questions. The mutual exclusivity assumption as proposed by Markman is a statement about an operating characteristic of the child's cognitive system that facilitates word learning. A proposal about cue competitions is a proposal about the more micro-level mechanisms that may give rise to that operating characteristic. An analogy helps: One possible account of why someone just ate a cookie is that they are hungry; another possible account is that they ate the cookie because of low levels of leptin and the release of ghrelin. The second account might be wrong, and the first might well be right; but the second is not in opposition to the first in any sensible way. Moreover, a choice between the two cannot be made according to which one better accounts for the macro-level behavior of eating the cookie. Instead, the relevance of leptin and ghrelin to cookie eating must be decided in terms of micro-level processes about which the macro-level construct of hunger makes no predictions.

In what follows, we consider Gelman's original question of how children might know to attend to the right properties for learning by considering contemporary evidence on attentional mechanisms and then asking whether these mechanisms might play a role in enabling children to attend to the right information for learning words. The evidence suggestive of a role for cued attention in word learning is primarily at a macro level. Accordingly, we next discuss the kind of micro-level studies needed to pursue these proposed mechanisms. We do not consider macro-level explanations of children's early word learning as competing

hypotheses to the proposals about cued attention. However, we conclude with a reconsideration of the contention in the early word learning and what is (and is not) at stake.

## 2. Cued attention

Attention is a construct that is so widely used in psychological theory that William James (1890) lamented ''everyone knows what attention is'' with the subtext that everyone may know but no one agrees. Within developmental psychology, ''attention'' is currently studied with respect to several different (but potentially deeply related, see Posner & Rothbart, 2007) phenomena, including sustained attention (e.g., Miller, Ables, King, & West, 2009; Richards, 2005, 2008), attentional switching and disengagement (e.g., Blaga & Colombo, 2006; Posner, Rothbart, Thomas-Thrapp, & Gerardi, 1998; Richards, 2008), executive control (e.g., Chatham, Frank, & Munakata, 2009; Diamond, 2006; Hanania & Smith, in press), and joint attention among social partners (e.g., Grossmann & Farroni, 2009; Hirotani, Stets, Striano, & Friederici, 2009). There are very few developmental studies specifically concerned with contextually cued attention (e.g., Goldberg, Maurer, & Lewis, 2001; Smith & Chatterjee, 2008; Wu & Kirkham, in press). Although we will briefly consider how cued-attention might be related to other forms of attention in children at conclusion, we focus on contextually cued attention in this paper precisely because so little is known about its development despite its apparent ubiquity in sensory, perceptual, and cognitive processing in adults. Briefly, the well-documented fact is this: Cues that have been probabilistically associated with some stimulus in the past enhance detection, processing, and learning about that stimulus (e.g., Brady & Chun, 2007; Chun & Jiang, 1998). In this section, we briefly summarize the very broad literature that supports these conclusions, noting how these processes capture regularities in the input, protect past learning, and guide future learning.

### *2.1. Predictive cues*

The attentional consequences of learned associations between cues and the stimuli they predict has been well known since the work of Mackintosh (1975) and Rescorla and Wagner (1972). Originally in the context of classical conditioning but more recently also in the broader domain of associative learning (see Chapman & Robbins, 1990; Kruschke & Blair, 2000; see also, Ramscar, Yarlett, Dye, Denny, & Thorpe, in press), these studies present learners with cues that predict specific outcomes. The results show that what is learned about the relation between those cues and outcomes depends on the cues present in the task, their relative salience, *and the learner's history of experiences with those cues in predicting outcomes.* Critically, these cued-attention effects are not about single cues associated with single attentional outcomes but rather are about the consortium of cues present in the moment and all of their predictive histories.

Three illustrative phenomena are overshadowing, blocking, and latent inhibition (Kamin, 1968; Lubow & Moore, 1959; Mackintosh, 1975; Rescorla & Wagner, 1972). The relevant task structures are shown in Table 1. Overshadowing refers to the situation in which two

Table 1
Cue interactions in associative learning

| Cue Interaction | First Learning Phase | Second Learning Phase | Learning Outcome |
|---|---|---|---|
| Overshadowing | **AB** → X | | A → X |
| Blocking | A → X | **AB** → X | A → X |
| Latent inhibition | B → | AB → X | A → X |
| Highlighting | AB → X | AC → Y | C → Y |
| Mutual exclusivity | A → X | AB → Y | B → Y |

*Note.* The letters A, B, C indicate cues and the letters X and Y indicate predicted outcome for the first phase of learning, and the subsequent second phase of learning. What is learned after the second phase is indicated in the third column. Bold letters indicate more salient cues (either through learning or intrinsic salience).

cues are presented together and jointly predict some outcome (e.g., in a category learning study, two symptoms might predict some disease, or in classical conditioning, a tone and a light might predict shock). The strength of association of each cue to the outcome depends on its *relative* salience. However, it is not simply that the most salient cue is learned better; rather, salience differences are exaggerated in that the more salient cue may ''overshadow'' the less salient cue with little or no learning at all about the less salient cue (Grossberg, 1982; Kamin, 1969; Kruschke, 2001; Mackintosh, 1976; Rescorla & Wagner, 1972). Blocking refers to the case in which the greater salience of one cue is not due to intrinsic salience but is due, instead, to past learning. If some cue regularly predicts some outcome and then, *subsequently*, a second cue is made redundant with the first and so also predicts that outcome, there is little or no learning about the second cue: Learning is blocked by the first predictive cue (Bott, Hoffman & Murphy, 2007; Kamin, 1968; Kruschke & Blair, 2000; Shanks, 1985), as if the first cue were more ''salient'' and thus overshadowing in this predictive context. Latent inhibition, like blocking, also makes the point that salience is a product of learning and predictive strength. But here the phenomenon is learned irrelevance: If a cue is first varied independently of the outcome so that it is not predictive at all but then it is made perfectly predictive, it will be hard to learn and ''overshadowed'' by other cues (e.g., Lubow, 1997; Lubow & Kaplan, 1997; Mackintosh & Turner, 1971).

These phenomena make three key points about the cues in cued attention: (a) cue strength is determined by predictive power, not mere co-occurrence; (b) cue strength depends on the ordered history of predictability, not batch statistics; and (c) individual cues interact, such that one cannot simply predict whether some cue will be learned by considering it alone, one must instead know its history and the history of the other cues in the learning environment. These phenomena are evident in many adult statistical learning tasks (e.g., Cheng & Holyoak, 1995; Kruschke, 2001; Kruschke & Blair, 2000; Kruschke & Johansen, 1999; Ramscar et al., in press) and as Ellis (2006) concluded, their ubiquity implies that human learning about predictive relations is bounded by basic mechansisms of cued attention (see Yu & Smith, in press; Yoshida, unpublished data).

Highlighting is a higher level phenomenon that may be understood as a product of blocking and overshadowing and it is a particularly robust phenomenon in adult associative learning (Kruschke, 1996, 2005; Ramscar et al., in press). It also demonstrates how cued

attention both protects past learning and guides new learning. Highlighting emerges when adults first learn one set of predictive cues to task-relevant information and then later are exposed to an overlapping set of *new and old* predictive cues that predict the relevance of *different* information. The task structure that leads to highlighting is also provided in Table 1. Learners are first exposed to a conjunctive cue (A + B) that predicts an outcome (X), and then are presented with a new conjunctive cue that contains one old component (A) plus a new one (C) that predicts a new outcome (Y). The key result concerns learning during the second phase: Learners associate the new cue with the new outcome more than they associate the old cue with the new outcome. For example, if the learner is first taught that RED and SQUARE predict category X and then is taught that RED and CIRCLE predict category Y, the learner does not learn about the relation of RED and category Y, but rather appears to selectively attend only to the novel cue, CIRCLE, and to learn that CIRCLE predicts category Y. In brief, novel cues are associated with novel outcomes. By one explanation, this derives from the rapid (and automatically driven) shift of attention away from the previously learned cue (RED and SQUARE) in the context of the new outcome, so that the new cue (CIRCLE) becomes attentionally highlighted and strongly associated with the new outcome (e.g., Kruschke, 1996, 2005). This attention-shifting account has also been supported by eye-tracking results (Kruschke, Kappenman, & Hetrick, 2005).

Highlighting, as well as blocking and overshadowing may be understood in terms of competitions among predictive cues, as if cues fight for associations with outcomes, so that once an association is established it protects itself by inhibiting the formation of new associations to the same predicted outcome. There are many reasons to think that these kinds of cue competitions should play a role in cognitive development, including the robustness of the phenomena in adults across a variety of task contexts. There is also one phenomenon in children's learning that shares an at least surface similarity to cue competition effects in associative learning: mutual exclusivity (Markman, 1989; see also Halberda, 2006); Hollich et al., 2000).

As illustrated in Table 1, the task structure that yields mutual exclusivity looks very much like that of blocking or highlighting: A first-learned association (the word ''cup'' to the referent cup) is followed by a subsequent learning task with a new cue (the novel word) and outcome (the novel object). Consistent with this proposal, several models (e.g., Mayor & Plunkett, 2010; Regier, 2005) have shown how mutual exclusivity might emerge in cue interactions in associative learning. Also consistent with this proposal are studies tracking moment-to-moment eye gaze direction in infants (Halberda, 2009); the attention shifting by infants in these studies resembles those of adults in highlighting experiments (Kruschke et al., 2005).

### 2.2. Lots of cues

Blocking and highlighting emerge in experimental studies in which researchers manipulate at most two to three cues and outcomes. The world, however, presents many probabilistically associated cues and outcomes yielding potentially complex patterns of predictability, with cues predicting other cues as well as potentially multiple outcomes. Importantly, such a

large ‘‘data set’’ of associations are also likely to have considerable latent structure, higher order regularities that might support deep and meaningful generalizations that go beyond specific cues and outcomes. Several connectionist models have sought to understand the structure of these higher order regularities (see, e.g., Colunga & Smith, 2005; Colunga, Smith, & Gasser, 2009; Kruschke, 1992, 2001; McClelland & Rogers, 2003). In one study, Colunga and Smith (2005) examined how the statistical regularities among perceptual properties of instances of basic level noun categories might create higher order partitions (object, substance) and also direct attention to relevant properties for categorizing artifacts (shape) and substances (material). To this end, they fed a connectionist net perceptual features (e.g., solidity, material, color, shape) that adults said characterized 300 noun categories commonly known by 2-year-olds. From this input, the network acquired generalized cued-attention effects that worked even when presented with novel entities; specifically, in making decisions about novel categories, the network weighted shape more in the context of solidity and weighted material in the context of nonsolidity. This result tells us that the corpus of early learned nouns and the correlated properties of the objects to which those nouns refer contain useable cue-outcome regularities of the kind that could train attention.

Several theoretical and empirical analyses suggest that systems of associations often present a form of *coherent covariation* (Rogers & McClelland, 2004) across cues and outcomes. The importance of coherent covariation to human learning has been demonstrated in experimental tasks showing that adults (and infants, Younger & Cohen, 1983) are more likely to attend to and learn about features that co-vary than those that merely co-occur (e.g., Kruschke & Blair, 2000; Medin & Schaffer, 1978; Medin, Altom, Edelson, & Freko, 1982). Other theorists have also pointed to what they call the *systematicity* (Billman & Heit, 1989) of associations; cues that probabilistically predict an outcome also often predict each other, and the systematicity across associations among cues, among outcomes, and among cues and outcomes matters in learning (e.g., Billman & Knutson, 1996; Goldstone, 1998).

Yoshida and Smith (2003b, 2005) (see also, Sloutsky & Fisher, 2008) proposed a cued-attention framework for thinking about the interactive effects of redundant cues that builds on the idea of interactive activation (see Billman & Knutson, 1996; Goldstone, 1998; Medin et al., 1982; O’Reilly, 2001). The proposal is illustrated in Fig. 1 using the simple case of

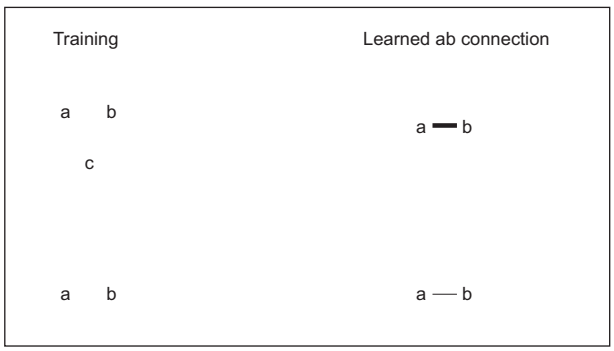


Fig. 1. Redundant correlations in the learning environment lead to stronger associations.

just three correlated cues: The correlation between cues **a** and **b** are learned either in the context of a third redundant cue, **c**, or without that third correlated cue. Experimental studies show that the learned connection between **a** and **b** is stronger if acquired in the context of **c**, which correlates with both **a** and **b**, than without that redundant correlation (Billman & Knutson, 1996; Yoshida & Smith, 2005). Moreover, the stronger associative link between **a** and **b** remains even when the redundant cue, **c**, is removed. Of course, in real-world learning, there may be many more than three correlated cues and much more complex patterns of correlation. Through mutually reinforcing correlations such as these, a system of many correlated cues may lead to what have sometimes been called ''gang effects'': Highly interconnected and dense patterns of associative links that give rise to patterns of internal activation that, as a result, are not tightly dependent on any one cue (e.g., Billman & Knutson, 1996; Goldstone, 1998; O'Reilly, 2001; Yoshida & Smith, 2003a,b). Instead, the interactive activation among co-varying cues can lead to snowball effects in which the joint activation of cues is stronger than the sum of the individually contributing associations.

One developmental result raised by Gelman (1990) in the original special issue, ''illusory projections,'' may be understood in terms of such gang effects (see Rogers & McClelland, 2004 for a more detailed account). Gelman reported that preschool children project properties onto an instance of a category that are not perceptually there; children say that they *saw* feet on an eyed yet ball-like and footless thing if they were told that the thing could move on its own. Fig. 2 provides an example of how this might work within the cued-attention framework. On the left is a hypothetical description of overlapping correlations among a set of perceptual properties: Things with eyes tend also to have feet, to move on their own, and so forth. Because of the overlapping correlations among all these properties, each will serve as a context cue that predicts and ''primes'' attention to the others (see Yoshida & Smith, 2003b) thereby potentially causing an individual to more readily detect or ''see'' a property in an ambiguous stimulus. Thus, on the right of Fig. 2 is a hypothetical description of the cluster of cues input to the system—eyed, furry, moving on its own, but footless. By hypothesis, fur, eyes, and movement will all increase attention to each other and, given the ambiguous stimulus, lead to the perception of feet.

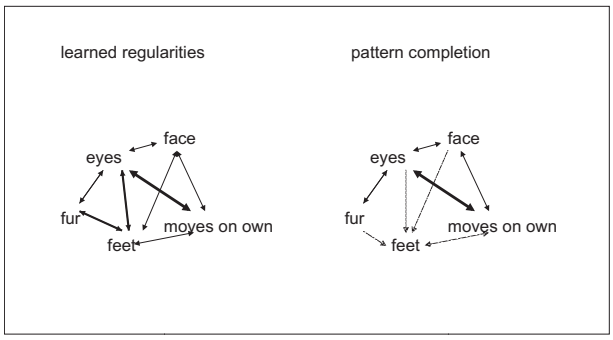


Fig. 2. An illustration of how redundant and overlapping correlations can lead to pattern completion and illusory projections.

### *2.3. Enhanced processing of predicted outcomes*

Cues predict outcomes, and in the present use of the term, an outcome is any stimulus event that is predicted (it could be another cue or the specifically task-relevant target information). Considerable and growing evidence suggests that cues that predict outcomes also enhance detection and processing of the predicted event. The relevant experiments often involve search or detection tasks in which participants are asked to detect some target—a shape or color—typically in a crowded field of competing distractors (Chun & Jiang, 1998; Jiang, Olson, & Chun, 2000). These studies have revealed strong effects of repeating arrays. Arrays that have been seen before show enhanced detection of the target, often with just one pre-exposure but increasing with repetition. This phenomenon is usually explained as a form of cued attention in which the array as a whole cues attention to a specific target at a specific location (Biederman, 1972; Boyce, Pollatsek, & Rayner, 1989; Lewicki, Hill, & Czyzewska, 1992; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

The experimental evidence also shows that in these tasks, contextual cues and their attentional effects emerge without awareness (Chun & Jiang, 1998; Jiang & Chun, 2001, 2003; Jiang & Leung, 2005; Lewicki, Hill, & Czyzewska, 1997; Lewicki et al., 1992; Olson & Chun, 2002; Shanks, Channon, Wilkinson, & Curran, 2006) and result both in rapid shifting of attention to a location (e.g., Chun and Turke-Browne, 2007; Clohessy, Posner, & Rothbart, 2001; Jiang & Chun, 2001; Summerfield et al., 2006) and also to the enhanced processing of particular stimulus features (Bichot & Rossi, 2005; Kruschke, 1996; Maunsell & Treue, 2006; Rossi & Paradiso, 1995). The growing neuroscience evidence on cued attention also indicates that contextually cued enhancements of stimulus processing are pervasive across early sensory processing and higher level perceptual and cognitive systems (Beck & Kastner, 2009; Gilbert, Ito, Kapadia, & Westheimer, 2000; Pessoa, Kastner, & Ungerleider, 2003). For example, neurophysiological studies have shown that cue-target associations enhance baseline firing rates (Chelazzi et al., 1993, 1998; Kastner et al., 1999) at early and middle levels in the visual system and alter neuronal tuning of target properties (Spitzer et al., 1988; Treue & Martinez Trujillo, 1999; Williford & Maunsell, 2006; Yeshurun & Carrasco, 1998). Context cued enhancements of processing have also been shown to play a role in decision making and in integrating sensory information across systems (Bichot et al., 1996; Boynton, 2009; Gold & Shadlen, 2007; Reynolds & Heeger, 2009).

Contemporary theories of these effects, often based on analyses of neural patterns of excitation, share a great deal with more classical explanations of cued attention (e.g., Mackintosh, 1975; Rescorla & Wagner, 1972) that emphasize prediction (or preactivation, e.g., Summerfield et al., 2006) and competition (e.g., Desimone & Duncan, 1995; Duncan, 1996). For example, the biased-competition theory of selective attention (see Beck & Kastner, 2009; Desimone & Duncan, 1995; Duncan, 1996) begins with the starting assumption that competition characterizes representational processes at the sensory, motor, cortical, and subcortical levels. In general, activation of a representation of some property, event, or object is at the expense of other complementary representations. Selection, or attention, then, occurs by biasing (e.g., priming) some representations in favor of others or by inhibiting competing repetitions. The presence of contextual cues previously associated with some

representation (at any of these levels) is thus thought to bias the competition. Multiple activated cues compete, and thus also interfere with each other, with stronger cues inhibiting weaker ones in a manner common to lateral inhibition models (Ludwig, Gilchrist, & McSorley, 2005; Walley & Weiden, 1973; de Zubicaray & McMahon, 2009). On these grounds, some have suggested that attention and selection are fundamentally a form of lateral inhibition in which the degree of activation of one representation inhibits that of nearby competitors (see Beck & Kastner, 2009; Duncan, 1996).

Here, then, is what we know about cued attention: Adults readily and unconsciously learn cues that predict other sensory events (outcomes) that are relevant in tasks. These predictive cues interact—both by supporting activation of the predicted event when they are correlated and also through competition that depends in fine-grained ways on the relative strengths of these cues and their history of prediction. These interactions are such that (a) early learned predictive relations tend to be preserved; (b) attention is systematically shifted to novel cues in the context of novel outcomes; and (c) the coherent covariation in large systems of cues and outcomes can capture latent structure that organizes attention in meaningful ways. Finally, predicted sensory events are processed faster and tuned more sharply than unpredicted events. All this suggests that cued attention is a basic mechanism that is likely to play a contributing role in many knowledge domains.

## 3. Cued attention as developmental process

Cued attention is also a mechanism of change and one that seems capable of driving considerable change in the cognitive systems. This is because cued attention is a single mechanism that aggregates knowledge, that is a repository of knowledge, and that guides learning, thereby driving the acquisition of new knowledge. In brief, attentional learning is a self-organizing process that builds on itself, becoming more directed, more knowledge driven, and potentially more domain specific as a consequence of its own activity. Fig. 3 builds on a prior proposal by Colunga and Smith (2008) about how cued attention gathers, integrates, and applies information over nested time scales. The three separate boxes on the left illustrate attentional processes at three time scales. The large box represents long-term associations among the many cues and outcomes in the learning environment. The middle box indicates the task context and the in-task dynamics of prediction and preactivation of cues and outcomes (see Samuelson & Smith, 2000a, 2000b; Samuelson, Schutte, & Horst, in press). Finally, there are the processes of interaction and competition driven by the current task and stimuli. As illustrated on the right side of the figure, these are nested processes: In-the-moment attention depends on task context as *modulated by* long-term associations; in-the-moment attention also adds to those long-term associations.

These nested interactions mean that attention can be biased in different ways in different contexts. Context cues that co-occur with (and define) specific tasks will come with repeated experience to shift attention to the task-relevant information. Fig. 4 represents this idea in terms of a series of context-sensitive salience maps, with the relative salience of regions in the perceptual field indicated by the relative darkness of the locations. The idea is this:

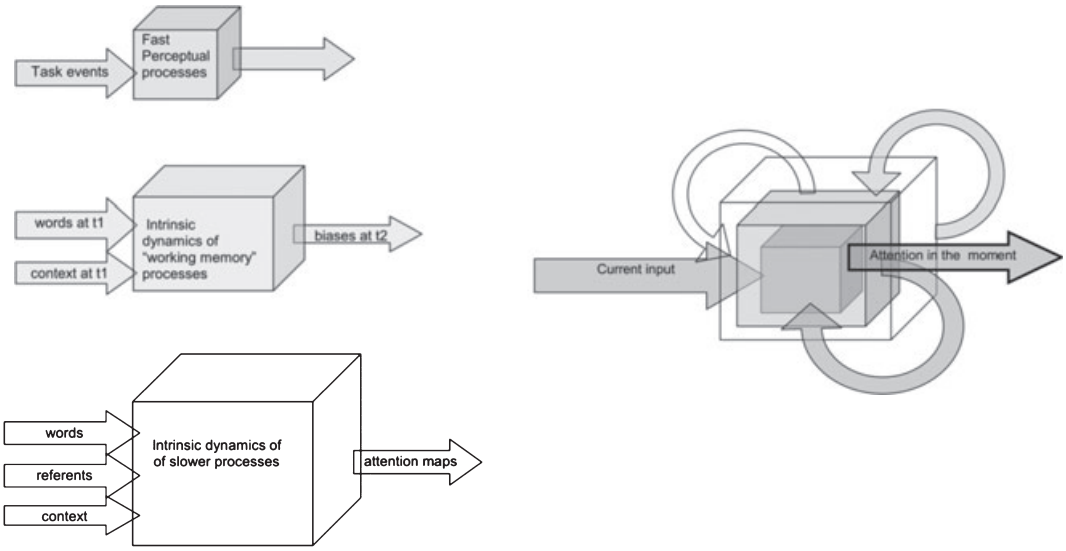


Fig. 3. The nested time scales of attentional learning. The three separate boxes on the left illustrate attentional processes at three time scales. The large box represents long-term associations among the many cues and outcomes in the learning environment. The middle box indicates the task context and the in-task dynamics of prediction and preactivation of cues and outcomes. The small box indicates processes of interaction and competition to a momentary stimulus. As illustrated on the right side of the figure, these are nested processes: In-the-moment attention depends on task context as *modulated by* long-term associations; in-the-moment attention also adds to those long-term associations.

Because the history of associated cues increases and decreases the relative activation of features and task targets in the perceptual field, the salience map *will* change with changes in contextual cues. This means potentially dramatic shifts in the detection, selection, and processing of stimulus events in different domains of expertise—word learning, quantity judgments, or spatial reasoning. These domains have associated contextual cues that—given sufficient experience—may structure the salience maps in consequentially different ways. Being smart in a domain may reflect (in part) predicting what information is relevant in that domain. For learners who have sufficient experiences in different domains, attention will nimbly dance about from context to context, enabling those learners to attend to just the right sort of information for that domain. In sum, what we know about cued attention and attentional learning suggests that these mechanisms will play a role across all domains of cognition and their self-changing nature will create domain-specific competencies. We consider next one domain in which these mechanisms may be at work.

## 4. Children's novel word generalizations

Many well-controlled studies show that young children need to hear only a *single* object name to systematically generalize that name to new instances in ways that seem right to

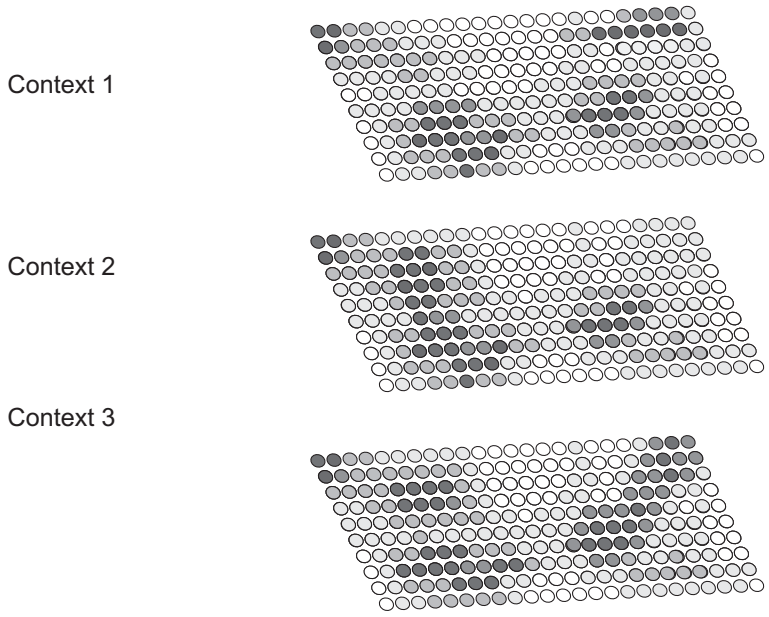


Fig. 4. Contextually changing salience maps. Relative salience is indicated by the relative darkness of the locations. By hypothesis, these change with changes in context cues that predict sensory events in the field.

adults (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989; Smith, 1995; Waxman & Markow, 1995). Moreover, children generalize names for different kinds of things by different kinds of similarities, shifting attention to the right properties for each kind of category. Thus, for the task of learning common nouns, young children have solved Gelman's problem: They know what properties to attend to so as to form categories ''shared with their elders.''

In the most common form of the novel word generalization task, the child is shown a single novel entity, told its name (e.g., *This is the toma*) and then asked what other things have the same name (e.g., *Where is the toma here?)* Many experiments have examined three kinds of entities (examples are shown in Fig. 5) and found three different patterns of generalization. Given objects with features typical of animates (e.g., eyes or legs), children extend the name narrowly to things that are similar in multiple properties. Given a solid inanimate artifact-like thing, children extend the name broadly to all things that match in shape. Given a nonsolid substance, children extend the name by material. These are highly reliable and replicable results—obtained by many researchers—and in their broad outline characteristic of children learning a variety of languages (e.g., Booth & Waxman, 2002; Gathercole & Min, 1997; Imai & Gentner, 1997; Jones & Smith, 2002; Jones, Smith, & Landau, 1991; Kobayashi, 1998; Landau, Smith, & Jones, 1988, 1998; Markman, 1989; Soja, Carey, & Spelke, 1991; Yoshida & Smith, 2001; see also Gelman & Coley, 1991; Keil, 1994).

Cued attention may a play a role in these generalizations in the following way: Artifacts, animals, and substances present different features (angularity, eyes, nonsolidity) and these

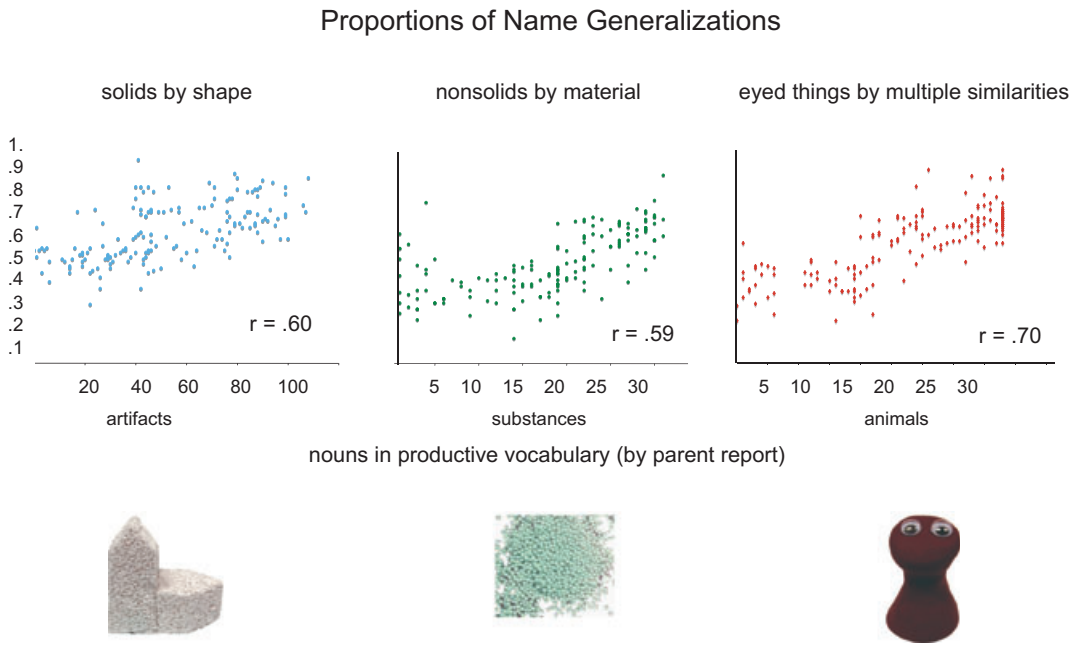


Fig. 5. Scatterplots of individual children's performances in the Novel Noun Generalization task as a function of the number of different kinds of nouns in their productive vocabularies. Individual data are from: Smith et al., 1992; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002; Yoshida & Smith, 2003b; Jones & Smith, 1998, 2002; and Jones & Smith, unpublished data.

features co-occur with different words (quantifiers, verbs, adjectives, etc.). The features and co-occurring words are thus potential context cues that could shift attention in systematic ways to the relevant properties for the particular kind of category—to multiple properties for animals to shape for artifacts, and to material for substances. And, indeed, the literature is filled with experimental demonstrations of these effects (Booth & Waxman, 2002; Colunga, 2006; Colunga & Smith, 2004; Gathercole, Cramer, Somerville, & Jansen op de Haar, 1995; Jones & Smith, 1998; McPherson, 1991; Samuelson, Horst, Schutte, & Dobbertin, 2008; Soja, 1994; Ward, Becker, Hass, & Vela, 1991; Yoshida & Smith, 2005; Yoshida, Swanson, Drake, & Gudel, 2001).

These perceptual- and linguistic-context effects on children's novel noun generalizations also increase with age and language learning, just as they should if children are learning relevant contextual cues (Jones et al., 1991; Landau et al., 1988; Samuelson, 2002; Samuelson & Smith, 1999, 2000a; Smith, 1995; Soja et al., 1991). Fig. 5 presents a summary of the developmental pattern. The figure shows scatterplots of individual children's novel noun generalizations for solid artifactual things, things with eyes, and nonsolid things from a variety of different experiments (with many different unique stimuli and task structures) conducted in our laboratories over the years. Each individual child's generalizations by the category-relevant property (shape for solid things, multiple similarities for eyed things, material for nonsolid things) is shown as a function of the number of artifact, animal, and

substance names in the individual child's vocabulary. The figures show that the systematicity of these novel noun generalizations increases with early noun learning. Thus, in the broad view, the developmental pattern of children's novel noun generalizations fits what might be expected if the underlying mechanism was cued attention: Children learn cues for task-relevant properties and after sufficient experience, those cues come to shift attention in task-appropriate ways.

## 5. Cross-linguistic differences

By hypothesis, the learning environment presents clusters of perceptual cues (e.g., eyes, legs, mouths, body shapes in the case of animates) and clusters of linguistic cues (e.g., the words ''wants,'' ''is happy,'' ''mad,'' and ''hungry'') that are associated with each other and that predict the relevant similarities for categorizing different kinds. If perceptual and linguistic cues are both part of the same attentional cuing system they should interact, and given coherent covariation, should reinforce each other. The ''natural'' experiment that provides evidence is the comparison of children learning different languages.

### *5.1. Systematicity*

The novel word generalizations of children learning English and Japanese have been examined in a number of studies (Imai & Gentner, 1997; Yoshida & Smith, 2001, 2003a, 2005). Both languages provide many linguistic cues that correlate with artifacts (and attention to shape), animals (and attention to multiple similarities), and substances (and attention to material). However, English arguably provides more *systematic* and coherently co-varying cues distinguishing objects and substances, whereas Japanese arguably provides more coherently co-varying cues distinguishing animates and inanimates (see Yoshida & Smith, 2003b, for a discussion).

In particular, the count-mass distinction in English partitions all nouns into discrete countable entities (objects and animals) or masses (substances) and the various linguistic markers of this distinction (determiners, plural) are strongly correlated with the perceptual cues (and particularly solidity) that predict categorization by shape. Moreover, the predictive relations between linguistic cues and perceptual cues are particularly strong in the 300 English nouns that children normatively learn by 2½ years (see Colunga & Smith, 2005; Samuelson & Smith, 1999; Smith, Colunga, & Yoshida, 2003). By hypothesis, these linguistic cues should augment attention to shape for solid artifactual things and to material for nonsolid substances. Because these linguistic cues lump object and animal categories together, they might also weaken the predictability of the distinction between artifact categories as shape-based categories and animal categories as organized by multiple properties.

Japanese, unlike English, makes no systematic distinction between count and mass nouns (and has no English-like plural). Therefore, there is less redundancy in the cue-category correlations with respect to object and substance categories. However, Japanese offers more systematic cues with respect to animates and inanimates than does English. As just one

example, every time a Japanese speaker refers to the location of an entity, in frames as ubiquitous as *There is a____,* they must mark the entity as animate or inanimate *(ga koko ni **iru*** vs. *__ga koko ni **aru**,* respectively). In this, as well as other ways, Japanese, relative to English, adds extra and systematic cues that correlate with perceptual cues predictive of animal versus nonanimal category organizations.

Under the cued-attention framework, the coherent variation of linguistic and perceptual cues within the two languages should create measurable cross-linguistic differences in the noun generalizations of children in the artificial word-learning task. For children learning both languages, solidity predicts attention to shape, nonsolidity predicts attention to material, and features such as eyes and feet predict attention to multiple similarities. But for children learning English, the solidity–nonsolidity cues co-vary and are supported by linguistic cues. For children learning Japanese, eyed–noneyed cues correlate with pervasive linguistic contrasts. Thus, there should be stronger, earlier, and sharper distinctions in novel noun generalizations for solid versus nonsolids for English-speaking children than for Japanese-speaking children and stronger, earlier, and sharper distinctions between eyed and noneyed stimuli for children learning Japanese. Experimental studies have documented these differences (Imai & Gentner, 1997; Yoshida & Smith, 2003b).

### *5.2. Gang effects*

Yoshida and Smith (2005) provided an experimental test of the cued-attention account of these cross-linguistic differences in a 4-week training experiment that taught monolingual Japanese children redundant linguistic cues analogous to count-mass cues in English. The experiment used a $2 \times 2$ design: Linguistic cues versus no linguistic cues correlated with category organization and solidity during training, and the presence or absence of those linguistic cues at test. The key results are these: Children who were trained with correlated linguistic and perceptual cues outperformed those who were not so trained, attending to the shape of solid things and the material of nonsolid thing and did so even when they were tested with totally novel entities *and even when the trained linguistic cue was not present at test.* That is, learning the links between solidity and nonsolidity and shape and material *in the context of redundant linguistic cues* made the perceptual cue-outcome associations stronger.

### *5.3. Illusory projections*

In a related study, Yoshida and Smith (2003a) presented English and Japanese children with novel entities like that in Fig. 6. The protrusions are ambiguous; depending on context, both Japanese- and English-speaking children could see them as legs or as wires. However, when presented in a neutral context, Japanese-speaking children were more likely to see them as leg-like and to generalize a novel name for the exemplar by multiple similarities; English-speaking children were more likely to see them as wires and to generalize a novel name for the exemplar by shape. In brief, Japanese-speaking children showed an enhanced sensitivity to subtle cues vaguely suggestive of animacy, a sensitivity that may be created in

Fig. 6. An ambiguous object: The protrusions may be seen as wires or as limbs.

a history of redundant linguistic-perceptual cues that reinforce attention to each other and in so doing prime the relevant representations for animal-like features.

These cross-linguistic differences indicate that perceptual and linguistic cues interact in a single system that organizes attention in novel word generalization tasks. The findings also suggest that cued attention might be one mechanism behind Whorfian effects in which speakers of different languages are found to be particularly sensitive to and attentive to relations and properties that are lexicalized in their language (see Boroditsky, 2001; Bowerman, 1996; Choi & Bowerman, 1991; Gathercole & Min, 1997; Gentner & Boroditsky, 2001; Levinson, 2003).

## 6. Competition and prediction

### *6.1. One attentional pathway*

One early study by Smith, Jones, and Landau (1992) using the novel-noun generalization task directly examined cue-competitions by pitting cues against each other. The study examined sentence frames associated with naming objects (*This is a riff*) and frames associated with labeling properties (*This is a riff one*) asking under what contexts the count noun frame would result in increased attention to shape and in what contexts the adjective frame would result in increased attention to color. Children know more nouns and know them earlier than they do adjectives, so the noun frame might be expected to be a stronger attentional cue than the adjective frame. Particularly relevant to possible underlying mechanisms of cued-attention, Smith et al. examined the effects of noun and adjective frames when the colors of the labeled things were dull (olive green), intrinsically salient (glittery silver-gold), or very salient (glittery silver-gold under a spotlight).

The results provide strong support for a competition among cues that interacts with other stimulus-driven (exogenous) pulls on attention. Given dull colors, children generalized novel words presented in both the noun and the adjective frames by shape. Given glittery colors without a spotlight, children generalized novel words in a noun frame by shape but in

an adjective frame by color. Apparently the association between the noun frame and shape is strong enough to overcome competition from glitter, but the pull from glitter helps the weaker adjective frame in guiding attention to the property. Finally, when given glittery colors under a spotlight, children often generalized the word—even in the count noun frame—by color. Children's performance strongly suggests a winner-take-all competition for attention in which learned cues and other pulls on attention directly interact.

These results also illustrate an important point about attention: It forces a decision and there is but one pathway and one decision. The single pathway means that attention must *integrate* information—from past learning, from task contexts, from immediate input—into a single attentional response. Because attention integrates multiple sources of information in the moment, it will also be flexible and adaptive to the idioscyncracies of specific tasks. The results also underscore the idea that words, whatever else they are, are also cues for attention and play a role in guiding in-the-moment attention.

### *6.2. Cue strength*

Colunga et al. (2009) offered an analysis of how the structure of the Spanish and English count-mass system may yield different cue strengths and as a consequence different patterns of cue interactions. The relevant difference between the two languages can be illustrated by thinking about how speakers can talk about a block of wood. An English speaker talking about a wooden block might say ''a block'' if he or she is talking about its shape or ''some wood'' if he or she is talking about its substance. Further, an English speaker cannot say ''some block'' or ''a wood'' because ''block'' is a count noun and ''wood'' is a mass noun. In contrast, a Spanish speaker, talking about the wooden block, could say ''un bloque'' (a block) or ''una madera'' (a wood) when talking about the coherent bounded and shaped thing but would say ''algo de madera'' (some wood) when talking about the substance irrespective of its shape. That is, count-mass quantifiers are used more flexibly across nouns in Spanish than English to signal different task-relevant construals of the entity (Iannucci, 1952). Note that English has some nouns that work like ''madera'' in Spanish: ''a muffin'' predicts the context relevancy of muffin shape, but ''some muffin'' predicts the context relevancy of muffin substance. But such nouns are not common in English, whereas in Spanish, in principle, all nouns work this way, and in everyday speech many more nouns are used in both count and mass frames than in English (Gathercole & Min, 1997; Gathercole, Thomas, & Evans, 2000; Iannucci 1952). In brief, in Spanish, count-noun syntax is more predictive of attention to shape or material than the noun (*madera*) itself or the solidity of the object. In English, syntax, the noun, and solidity are (not perfectly but) more strongly correlated with each other, and equally predictive of whether the relevant dimension is shape or material.

In light of these differences, Colunga et al. hypothesized that Spanish count-mass syntax should overshadow other cues—solidity, the noun—in children's novel noun generalizations. In the experiment, monolingual Spanish-speaking children and monolingual English-speaking children were presented with a novel solid entity named with either mass or count nouns, e.g., ''a dugo'' or ''some dugo'' and then tested with the two syntactic frames. The syntactic frame had a stronger effect on the Spanish-speaking children's noun

generalizations than on the English-speaking children's noun generalizations. In other words, Spanish-speaking children learn to attend to mass-count syntax as a way to disambiguate nouns that sometimes refer to shape and sometimes refer to material, and in the context of mass syntax they attend to material even when the items are solid. These results fit the idea that children learn predictive cues and that stronger cues inhibit weaker ones. They also emphasize how individual cues reside in a system of cues and that it is the system of interacting cues that determine attention.

### *6.3. Protecting past learning*

Cue competition privileges old learning as stronger cues inhibit new ones and force attention elsewhere. Yoshida and Hanania (2007) provided an analysis of adjective learning that illustrates how this aspect of cued attention could play a positive role in word learning. The motivating question for their analysis was how, in the context of both a novel adjective and a known noun (e.g., in the context of ''a *stoof* elephant''), attention could be properly shifted to a property (such as texture) rather than the shape of the thing. If the word ''elephant'' is a strong cue for attention to elephant shape, how does the child manage to learn novel adjectives? The answer offered by Yoshida and Hanania is much like the mutual-exclusivity account (Markman, 1989) but is in terms of a cued-attention mechanism rather than children's knowledge about how different kinds of words link to different kinds of meanings.

Their experiments were based on a prior study by Mintz and Gleitman (2002) that showed that the explicit mention of the noun (e.g., the *stoof* elephant) helps children learn the novel adjective relative to conditions in which the noun is not explicitly spoken (e.g., a *stoof* one). Yoshida and Hanania proposed that the role of the noun could be understood as a kind of attentional highlighting via competition. In brief, their cued-attention explanation of the explicit mention of the noun is this: Children usually experience the word *elephant* in the context of elephant-shaped things with typical elephant textures (e.g., ROUGH). In these novel adjective-learning experiments, the named object is an elephant-shaped elephant with a totally novel texture (e.g., with tiny holes punched throughout). Thus, in the context of *stoof elephant*, the known cue-outcome (*elephant*-elephant shape) shifts attention to the novel cue—novel outcome (*stoof*-holey texture). In the context of *stoof one*, there is no conjunctive cue containing both a known and a novel cue and thus less competition, which in this case means less directed attention to the novel property. Yoshida and Hanania provided indirect support for this account by showing that the mere conjunction of words (in an unordered list format, e.g., *elephant, stoof, red)* rather than in a sentence format in which *stoof* modifies the known noun *elephant*) was sufficient for the novel-word to novel-property mapping and by showing that the attentional shift away from shape to the right property, depends on the number of competitive cues (*stoof* in the context *red* and *elephant* leads to stronger effects than *stoof* in the context of elephant alone).

Competitive processes are ubiquitous in the sensory and cognitive system (see Beck & Kastner, 2009). They are at the core of current theories of lexical access and on-line sentence processing (e.g., Bowers, Davis, & Hanley, 2005; Davis & Lupker, 2006). It seems

likely that they play an important role in lexical development as well (Halberda, 2009; Hollich, Hirsh-Pasek, Tucker, & Golinkoff, 2000; Horst, Scott, & Pollard, in press) and, as suggested by Yoshida & Hanania, 2007; Yoshida and Hanania, unpublished data), competitive attentional processes may be particularly important in early development because by protecting already learned associations, they guide learning about novel cues and outcomes, and in this way may leverage little bits of learning into strong forces that effectively speed up learning.

The evidence on children's novel noun generalizations show that context—both linguistic and perceptual—cues children to form categories on the basis of different properties and to select one object over another as the referent of some word. The contextual cueing effects that characterize these early word generalizations share a number of (at least surface) similarities to cued attention—the role of coherent covariation, competition with other pulls on attention, attention shifting to protect past learning. However, none of these early word-learning experiments unambiguously show that these are *attentional* effects. They cannot because the measures of performance are at a macro level—generalizing names to things. To show that mechanisms of attentional learning play a role in these phenomena, we need measures at finer levels of resolution.

## 7. Going micro

Our main thesis is that cued attention is a potentially powerful developmental mechanism and that early word learning—and particularly the phenomena associated with children's smart novel noun generalizations—would be a fertile domain in which to investigate this idea. However, the granularity of the phenomena studied under cued attention—rapid attentional shifts, rapid detection, enhanced processing, and discrimination—and those studied in most novel word-learning experiments—generalization of an object name to a new instance—are not comparable. What is needed are finer-grained measures of component behaviors—orienting, disengagement, detection, and discrimination—in the context of words and correlated object properties. One might ask, for example, whether a count noun sentence frame primes detection of an odd shape but not an odd texture in a visual search task or whether a count noun sentence frame enhances shape discriminations over texture discriminations.

Recent studies of early word learning using moment-to-moment tracking of eye gaze suggest the new insights to emerge from analyzing attention and word learning at a finer temporal scale. For example, recent studies using this methodology suggest that older (24 month old) and younger (18 month old) word learners may both map words to referents, but that older learners more rapidly process the word and object and more rapidly shift attention to the right object (Fernald, Zangl, Portillo, & Marchman, 2008). Other studies manipulating the sound properties of the words have shown in the time course of looking at two potential referents, that the time course of competition and its resolution depend on the similarity of cues (Swingley & Aslin, 2007). Other recent studies showed that time course of looking to the named referent closely tracked the co-occurrence probabilities of words, referents, and

distracters (Fernald, Thorpe, & Marchman, 2010; Vouloumanos & Werker, 2009; Yu & Smith, in press). These finer-grained methods might also be used to directly test the idea that learned clusters of coherently varying cues organize orienting to, detecting, and processing of predicted properties in ways that nimbly guide attention to the right components of a scene for word learning.

We also need studies on cued attention outside of the domain of early word learning, studies that investigate the processes and mechanisms in adult literature. At present there are very few studies (Kirkham, Slemmer, & Johnson, 2002; Reid, Striano, Kaufman, & Johnson, 2004; Richards, 2005; Sloutsky & Robinson, 2008; Wu & Kirkham, in press) and no active area of concentrated research effort. The studies that do exist suggest that cued attention will not work exactly the same way in children as in adults. At present, we know very little about attentional learning and cued attention in young children. One distinction in the study of attentional processes that may be particularly relevant is that between exogenous and endogenous cueing (see Colombo, 2001; Goldberg et al., 2001; Smith & Chatterjee, 2008; Snyder & Munakata, 2008, 2010). Exogenous attention refers to the quick capture of attention by salient stimuli, such as the flashing of lights. Endogenous attention is attention directed by associated cues. Many studies of endogenous versus exogenous cuing use the covert-orienting paradigm (Jonides, 1980; Posner & Raichle, 1994): A cue directs attention to the target information either exogenously (a flashing light) or endogenously (by a previous association) and either correctly or incorrectly. There is considerable development in the dynamics of both exogenous and endogenous cueing and in disengaging attention when miscued (see Colombo, 2001; Goldberg et al., 2001. However, advances in endogenous cuing appear to lag behind exogenous cuing particularly in the overriding of misleading cues and in switching attention given a new cue (Cepeda, Kramer, & de Sather, 2001; Snyder & Munakata, 2010). By one proposal the major developmental changes lie in the processes that resolve cue competition (Snyder & Munakata, 2008, 2010). Two recent studies of preschoolers in attention-switching tasks suggest that experimentally training clusters of different cues for targets (Sloutsky & Fisher, 2008) or training more abstract cues (Snyder & Munakata, 2010) enhances switching, a result that may link cued attention, word learning, and executive control of attention.

In brief, internally directed attention that smartly moves to the right information for learning is likely to have its own compelling developmental story, with successful attention dependent on one's history with a particular set of cues and overlapping mechanisms of activation, preactivation, cue interactions, and competition. We need systematic and fine-grained studies of the development of cued attention from infancy through the early word-learning period and into childhood.

## 8. Knowledge as process

The systematicity with which young children learn the concepts they share with their elders, and they systematicity with which they generalize a newly learned name in different ways for different kinds, clearly demonstrates that they have knowledge. That is not in

doubt; the contentious question is the nature of the knowledge. Macro-level accounts of this knowledge have been in terms of propositions and core principles about kinds of categories and kinds of word meanings, knowledge of the sort proposed in the 1990 special issue (see also Booth & Waxman, 2002; Dewar & Xu, 2009; Diesendruck & Bloom, 2003; Hall, 1996; Soja, 1992). These explanations are based on the traditional view of cognition as sharply divided into knowledge and process and competence and performance. Thus, in terms of this traditional framework, the two levels of knowledge and process illustrated on the left of Fig. 7 were separable and one could sensibly ask of children's performance in any word-learning task whether their performance reflected knowledge about how words map to meaning (competence) or whether it reflected some performance variable. One example of this are the various debates as to whether the shape bias emerges because of conceptual knowledge about how nouns map to categories (e.g., Booth & Waxman, 2002; Cimpian & Markman, 2005) or whether it emerges because of attentional learning (Smith & Samuelson, 2006). This traditional explanation would seem to suggest that one could pit conceptual explanations against explanations in terms of (only) real-time processing. This framing of the question—concepts versus attention—is problematic on its own terms. In any real-time

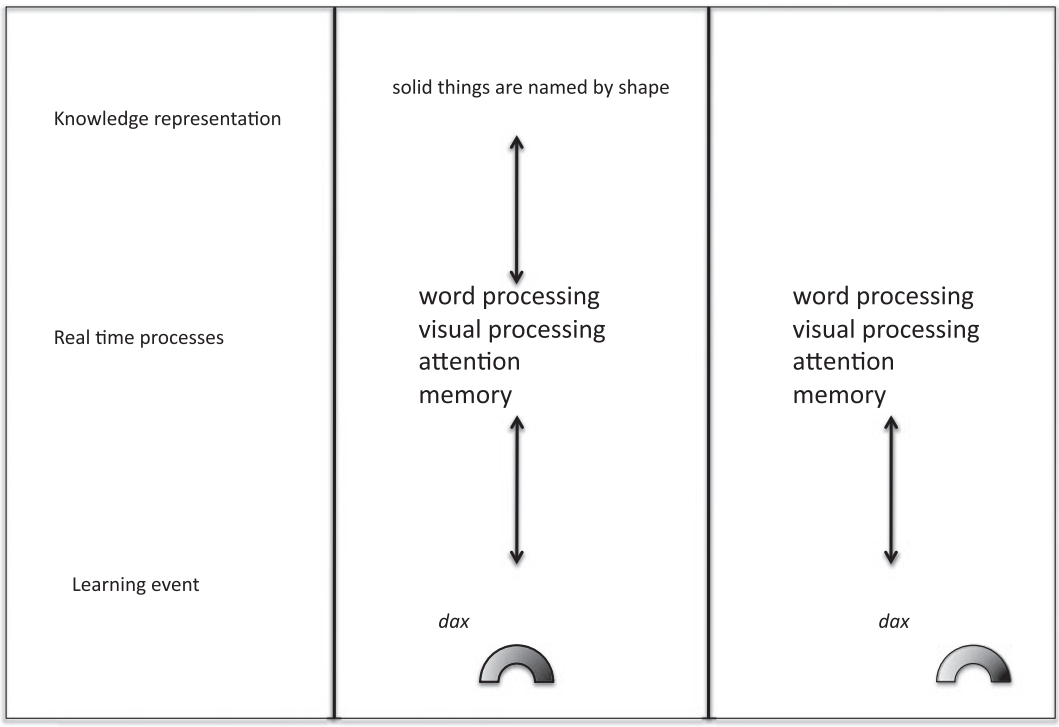


Fig. 7. An illustration of two views of cognition: On the left, knowledge representations are connected to real-time task performance as separable aspects of cognition. On the right, processes (which through their operating characteristics and history instantiate knowledge) generate performance in real-time tasks. Within this framework, beliefs, concepts, and propositional knowledge are higher level theoretical descriptions of the knowledge embedded in these processes.

task, conceptual knowledge requires supporting processes as well—the perceptual, attention, memory, and decision processes. Whatever can be explained by the account on the right (the process-only account) can *necessarily* be explained by the account on the left: Whatever can be explained by A (process) can *always* also be explained by A+B (process plus propositions).

The framing of the contending explanations in terms of concepts versus processes such as attention is also more profoundly misguided. Modern-day understanding of neural processes makes clear that knowledge has no existence outside of process and that the relevant processes often encompass many different systems and time-scales (see Barsalou, 2009). What we call knowledge is an abstraction over many underlying processes. Although these higher level abstractions may summarize meaningful regularities in the operating characteristics of the system as a whole, we also need to go underneath them—and understand the finer-grained processes from which they are made. ''Word learning rules'' and ''concepts'' may be to attentional processes as ''hunger'' is to hormones: Not a separate mechanism but a portmanteau abstraction that includes attentional learning (as well as other processes). Clearly, one goal of science is to unpack these carry-all abstractions. Cued attention is a particularly intriguing potential mechanism from this perspective because it is one that melds competence and performance, knowledge and process, perception and conception. It does so because cued attention is a process, operating in real time, strongly influenced by the momentary input, with effects at the sensory and perceptual level but also driven by the rich data structure of predictive relations between cues and outcomes in a lifetime of experiences. As such, attention is a process that aggregates, acquires, and applies knowledge.

## Acknowledgments

The research summarized in this paper was supported by NIMH grant R01MH60200, and NICHD grants R01HD 28675 to Linda Smith and NICHD 1R01HD058620-01 to Hanako Yoshida.

## References

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of Royal Society B*, *364*, 1281–1289.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*, 1154–1165.

Bichot, N. P., & Rossi, A. F. (2005). Parallel and serial neural mechanisms for visual search in Macaque Area V4. *Science*, *308*(5721), 529–534.

Bichot, N. P., Schall, J., & Thompson, K. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, *381*, 697–699.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*(4043), 77–80.

Billman, D., & Heit, E. (1989). Observational learning from internal feedback: A simulation of an adaptive learning method. *Cognitive Science*, *12*, 587–625.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 458–475.

Blaga, O. M., & Colombo, J. (2006). Visual processing and infant ocular latencies in the overlap paradigm. *Developmental Psychology*, *42*(6), 1069–1076.

Booth, A. E., & Waxman, S. R. (2002). Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition*, *84*, B11–B22.

Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22.

Bott, L., Hoffman, A., & Murphy, G. (2007). Blocking in category learning. *Journal of Experimental Psychology: General*, *136*, 685–699.

Bowerman, M. (1996). Cognitive versus linguistic determinants. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 145–186). Cambridge, England: Cambridge University Press.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*, B45–B54.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556–566.

Boynton, G. M. (2009). A framework for describing the effects ofattention on visual responses. *Vision Research*, *49*, 1129–1143.

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology*, *33*, 798–815.

Cepeda, N. J., Kramer, A. F., & de Sather, J. (2001). Changes in executive control across the life span: Examination of task switching performance. *Developmental Psychology*, *37*, 715–730.

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, *18*(5), 537–545.

Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *106*(14), 5529–5533.

Chelazzi, L. et al. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*, 345–347.

Chelazzi, L. et al. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*, 2918–2940.

Cheng, P. W., & Holyoak, K. J. (1995). Complex adaptive systems as intuitive statisticians: Causality, contin-qency, and prediction. In H. L. Roitblat & J. Meyer (Eds.), *Comparative approaches to Cognitive Science* (pp. 271–302 . Cambridge, MA: The MIT Press.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*, 83–121.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current opinion in Neurobiology*, *17*, 177–184.

Cimpian, A., & Markman, E. (2005). The absence of a shape bias in children's word learning. *Developmental Psychology*, *41*, 1003–1019.

Clohessy, A. B., Posner, M. I., & Rothbart, M. K. (2001). Development of the functional visual field. *Acta Psychologica*, *106*, 51–68.

Colombo, J. (2001). The development of visual attention in infancy. In S. T. Fiske, D. L. Schacter, & C. Zahn-Waxler (Eds.), *Annual review of psychology* (pp. 337–367). Palo Alto, CA: Annual Reviews.

Colunga, E. (2006). The effect of priming on preschooler's extensions of novel words: How far can ''dumb'' processes go? *Proceedings of the 30th Annual Boston University Conference on Language Development*, *23*, 96–106.

Colunga, E., & Smith, L. B. (2004). Dumb mechanisms make smart concepts. *Proceedings of the Annual Conference of the Cognitive Science Society*, *26*, 239–244.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*(2), 347–382.

Colunga, E., & Smith, L. B. (2008). Flexibility and variability: Essential to human cognition and the study of human cognition. *New Ideas in Psychology*, *26*, 174–192.

Colunga, E., Smith, L. B., & Gasser, M. (2009). Correlation versus prediction in children's word learning: Cross-linguistic evidence and simulations. *Language and Cognition*, *1*(2), 197–217.

Davis, C. J., & Lupker, S. J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 668–687.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Reviews of Neuroscience*, *18*, 193–222.

Dewar, K., & Xu, F. (2009) Do early nouns refer to kinds or distinct shapes? Evidence from 10-month-old infants. *Psychological Science*, *20*, 252–257.

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). New York: Oxford University Press.

Diesendruck, G., & Bloom, P. (2003). How specific is the shape bias? *Child Development*, *74*, 168–178.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI* (pp. 549–578). Cambridge, MA: The MIT Press.

Ellis, N. C. (2006). Selective attention and transfer phenomena in LZ acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27* (2), 164–194.

Fernald, A., Thorpe, K., & Marchman, V. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology*, *60*, 190–217.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. Fernández, & H. Clahsen (Eds.), *Language processing in children* (pp. 97–135). Amsterdam, The Netherlands: Benjamins.

Gathercole, V., Cramer, L., Somerville, S., & Jansen op de Haar, M. (1995). Ontological categories and function: Acquisition of new names. *Cognitive Development*, *10*, 225–251.

Gathercole, V., & Min, H. (1997). Word meaning biases or language-specific effects? Evidence from English, Spanish and Korean. *First Language*, *17*(49 Pt 1), 31–56.

Gathercole, V., Thomas, E. M., & Evans, D. (2000). What's in a noun? Welsh-, English-, and Spanish-speaking children see it differently. *First Language*, *20*, 55–90.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and animate-inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.

Gelman, S. A., & Coley, J. D. (1991). Language and categorization: The acquisition of natural kind terms. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought: Interrelations in development* (pp. 146–196). New York: Cambridge University Press.

Gentner, D., & Boroditsky, L. (2001). Individuation, relational relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge, England: Cambridge University Press.

Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Research*, *40*, 1217–1226.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review Neuroscience*, *30*, 535–574.

Goldberg, M., Maurer, D., & Lewis, T. (2001). Developmental changes in attention: The effects of endogenous cueing and of distractors. *Developmental Science*, *4*, 209–219.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

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.

Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory. *Psychological Review*, *89*(5), 529–572.

Grossmann, T., & Farroni, T. (2009). Decoding social signals in the infant brain: A look at eye gaze perception. In M. de Haan & M. R. Gunnar (Eds.), *Handbook of developmental social neuroscience* (pp. 87–106). New York: Guilford Press.

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.

Hall, D. G. (1996). Naming solids and non-solids: Children's default construals. *Cognitive Development*, *11*, 229–264.

Hanania, R., & Smith, L. B. (in press). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*.

Hirotani, M., Stets, M., Striano, T., & Friederici, A. D. (2009). Joint attention helps infants learn new words: Event-related potential evidence. *NeuroReport: For Rapid Communication of Neuroscience Research*, *20*(6), 600–605.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R., Brand, R. J., Brown, E., Chung, H. L., Hennon, E., & Rocroi, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the society for Research in child Development*, *65*(3), v–123.

Horst, J., Scott, E., & Pollard, J. (in press). The role of competition in word learning via referent selection. *Developmental Science*.

Iannucci, J. E. (1952). *Lexical number in Spanish nouns with reference to their English equivalents*, Philadelphia: University of Pennsylvania.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

James, W. (1980). *The Principles of psychology*. Chicago: University of Chicago Press.

Jiang, Y., & Chun, M. M. (2001). Asymmetric object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(4), 895–918.

Jiang, Y., & Chun, M. M. (2003). Contextual cueing: Reciprocal influences between attention and implicit learning. In L. Jiménez (Ed.), *Attention and implicit learning* (pp. 277–296). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Jiang, Y., & Leung, A. W. (2005). Implicit learning of ignored visual context. *Psychonomic Bulletin & Review*, *12*(1), 100–106.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 683–702.

Jones, S. S., & Smith, L. B. (1998). How children name objects with shoes. *Cognitive Development*, *13*, 323–334.

Jones, S. S., & Smith, L. B. (2002). How children know the relevant properties for generalizing object names. *Developmental Science*, *5*, 219–232.

Jones, S. S., Smith, L. B., & Landau, K. B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*, 499–516.

Jonides, J. (1980). Towards a model of the mind's eye's movement. *Canadian Journal of Psychology*, *34*, 103–112.

Kamin, L. J. (1968). 'Attention-like' processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9–33). Coral Gagbles, FL: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment* (pp. 279–296). New York: Appleton-Century-Crofts.

Kastner, S. et al. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.

Keil, F. (1994). The birth and nurturance of concepts by domains: The origins of concepts of living things. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 234–254). New York: Cambridge University Press.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.

Kobayashi, H. (1998). How 2-year-old children learn novel part names of unfamiliar objects. *Cognition*, *68*, B41–B51.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

Kruschke, J. K. (1996). Base rates in category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 3–26.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, *45*(6), 812–863.

Kruschke, J. K. (2005). Learning involves attention. In G. Houghton (Ed.), *Connectionist models in cognitive psychology, Ch. 4* (pp. 113–140). Hove, East Sussex, UK: Psychology Press.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, *7*(4), 636–645.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognitive*, *25*(5), 1083–1119.

Kruschke, J. K., Kappenman, E. S., & Hetrick, W. P. (2005). Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 830–845.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Landau, K. B., Smith, L. B., & Jones, S. (1998). Object perception and object naming in early development. *Trends in Cognitive Science*, *2*, 19–24.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge, England: Cambridge University Press.

Lewicki, P., Hill, T., & Czyzewska, M. (1992). Nonconscious acquisition of knowledge. *American Psycholgoist*, *47*(6), 796–801.

Lewicki, P., Hill, T., & Czyzewska, M. (1997). Hidden covariation detection: A fundamental and ubiquitous phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 221–228.

Lubow, R. E. (1997). Latent inhibition as a measure of learned inattention: Some problems and solutions. *Behavioural Brain Research Special Issue: Psychobiology of Learned Inattention*, *88*(1), 75–83.

Lubow, R. E., & Kaplan, O. (1997). Visual search as a function of type of prior experience with target and distractor. *Journal of Experimental Psychology: Human Perception and performance*, *23*(1), 14–14.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology*, *52*, 415–419.

Ludwig, C. J. H., Gilchrist, I. D., & McSorley, E. (2005). The remote distractor effect in saccade programming: Channel interactions and lateral inhibition. *Vision Research*, *45*(9), 1177–1190.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*, 276–298.

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning & Behavior*, *4*, 186–192.

Mackintosh, N. J., & Turner, C. (1971). Blocking as a function of novelty of CS and predictability of UCS. *Quarterly Journal of Experimental Psychology*, *23*, 359–366.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction. MIT Series in learning, development, and conceptual change*. Cambridge, MA: MIT Press.

Markman, E. M. (1990). Constraints children place on word meaning. *Cognitive Science*, *14*, 57–77.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neuroscience*, *29*(6), 317–322.

Mayor, J., & Plunkett, K. (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. *Psychological Review*, *117*(1), 1–31.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Review Neuroscience*, *4*(4), 310–322.

McPherson, L. (1991). A little goes a long way: Evidence for a perceptual basis of learning for the noun categories COUNT and MASS. *Journal of Child Language*, *18*, 315–338.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, & Cognitive*, *8*(1), 37–50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238.

Miller, J. L., Ables, E. M., King, A. P., & West, M. J. (2009). Different patterns of contingent stimulation differentially affect attention span in prelinguistic infants. *Infant Behavior & Development*, *32*(3), 254–261.

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.

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibitory competition and Hebbian learning. *Neural Computation*, *13*, 1199–1242.

Pessoa, L., Kastner, S., & Ungerleider, L. G. (2003). Neuroimaging studies of attention: From modulation of sensory processing to top-down control. *The Journal of Neuroscience*, *23*(10), 3990–3998.

Posner, M., & Raichle, M. (1994). *Images of mind*. New York: W. H. Freeman and Company.

Posner, M., & Rothbart, M. K. (2007). Research on attention as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23.

Posner, M. I., Rothbart, M. K., Thomas-Thrapp, L., & Gerardi, G. (1998). The development of orienting to locations and objects. In R. D. Wright (Ed.), *Visual attention* (pp. 269–288). New York: Oxford University Press.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (in press). Feature-label-order effects and their implications for symbolic learning. *Cognitive Science*.

Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science: A Multidisciplinary Journal*, *29*(6), 819–865.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport: For Rapid Communication of Neuroscience Research*, *15*(16), 2553–2555.

Rescorla, R. A., & Wagner, A. R. (1972) *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. In A. H. Black & W. F. Prokasy (Eds.), *Classical Conditioning II* (pp. 64–99). New York: Appleton-Century-Crofts.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*, 168–185.

Richards, J. E. (2005). Localizing cortical sources of event-related potentials in infants' covert orienting. *Developmental Science*, *8*(3), 255–278.

Richards, J. E. (2008). Attention in young infants: A developmental psychophysiological perspective. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (2nd ed.) pp. 479–497). Cambridge, MA: MIT Press.

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

Rossi, A. F., & Paradiso, M. A. (1995). Feature-specific effects of selective visual attention. *Vision Research*, *35*(5), 621–634.

Samuelson, L. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds. *Developmental Psychology*, *38*, 1016–1037.

Samuelson, L. K., Horst, J. S., Schutte, A. R., & Dobbertin, B. (2008). Rigid thinking about deformables: Do children sometimes overgeneralize the shape bias? *Journal of Child Language*, *35*, 559–589.

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 generalizations. *Cognition*, *110*, 322–345.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure, and syntax correspond? *Cognition*, *73*(1), 1–33.

Samuelson, L. K., & Smith, L. B. (2000a). Children's attention to rigid and deformable shape in naming and no naming tasks. *Child Development*, *71*(6), 1555–1570.

Samuelson, L. K., & Smith, L. B. (2000b). Grounding development in cognitive processes. *Child Development*, *71*, 98–106.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66.

Shanks, D. R. (1985). Forward and backward blocking in human contigency judgement. *Quarterly Journal of Experimental Psychology*, *37B*, 1–21.

Shanks, D. R., Channon, S., Wilkinson, L., & Curran, H. V. (2006). Disruption of sequential priming in organic and pharmacalogical amnesia: A role for the medial temporal lobes in implicit contextual learning. *Neuropsychopharmacology*, *31*(8), 1768–1776.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, auotmatic attending, and a general theory. *Psychological Review*, *84*(2), 127–190.

Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*, 639–651.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science: A Multidisciplinary Journal*, *32*(2), 342–365.

Smith, L. B. (1995). Self-organizing processes in learning to learn words: Development is not induction. In C. Nelson (Ed..), *Basic and applied perspectives on learning, cognition, and development, Vol. 28* (pp. 1–32). Marwah, NJ: Lawrence Erlbaum Associates.

Smith, S. E., & Chatterjee, A. (2008). Visuospatial attention in children. *Neurological Review*, *65*, 1284–1288.

Smith, L. B., Colunga, E., & Yoshida, H. (2003). Making an ontology: Cross-linguistic evidence. In D. Rakison & L. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 275–302). London: Oxford University Press.

Smith, L. B., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial. Life*, *11*, 13–39.

Smith, L. B., Jones, S. S., & Landau, B. (1992). Count nouns, adjectives, and perceptual properties in children's novel word interpretation. *Developmental Psychology*, *28*, 273–286.

Smith, L. B., Jones, 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.

Smith, L. B., & Samuelson, L. (2006). An attentional learning account of the shape bias: Reply to Cimpian and Markman (2005) and Booth, Waxman, and Huang (2005). *Developmental Psychology*, *42*, 1339–1343.

Snyder, H. R., & Munakata, Y. (2008). So many options, so little time: The roles of association and competition in underdetermined responding. *Psychological Bulletin & Review*, *15*, 1083–1088.

Snyder, H. R., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition*.

Soja, N. N. (1992). Inferences about the meaning of nouns; the relationship between perception and syntax. *Cognitive Development*, *7*, 29–45.

Soja, N. N. (1994). Evidence for a distinct kind of noun. *Cognition*, *51*(3), 267–284.

Soja, N. N., Carey, S., & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meanings: Object terms and substance terms. *Cognition*, *38*(2), 179–211.

Spitzer, H. et al. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311–1314.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*, 99–132.

Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*, 575–579.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*, 1611–1617.

Walley, R. E., & Weiden, T. D. (1973). Lateral inhibition and cognitive masking: A neuropsychological theory of attention. *Psychological Review*, *80*(4), 284–302.

Ward, T. B., Becker, A. H., Hass, S. D., & Vela, E. (1991). Attribute availability and the shape bias in children's category generalization. *Cognitive Development*, *6*(2), 143–167.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories. *Cognitive Psychology*, *29*(3), 257–302.

Williford, T., & Maunsell, J. H. (2006). Effects of spatial attention on contrast response functions in macaque area V4. *Journal of Neurophysiology*, *96*, 40–54.

Wu, R., & Kirkham, N. (in press). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*, 72–75.

Yoshida, H., & Hanania, R. (2007). Attentional highlighting as a mechanism behind early word learning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual meeting of the Cognitive Science Society* (pp. 719–724. Austin, TX: Cognitive Science Society.

Yoshida, H., & Smith, L. B. (2001). Early noun lexicons in English and Japanese. *Cognition*, *82*, 63–74.

Yoshida, H., & Smith, L. B. (2003a). Shifting ontological boundaries: How Japanese- and English- speaking children generalize names for animals and artifact. *Developmental Science*, *6*(1), 1–34.

Yoshida, H., & Smith, L. B. (2003b). Correlation, concepts and cross-linguistic differences. *Developmental Science*, *6*(1), 30–34.

Yoshida, H., & Smith, L. B. (2005). Linguistic cues enhance the learning of perceptual cues. *Psychological Science*, *16*(2), 90–95.

Yoshida, H., Swanson, J., Drake, C., & Gudel, L. (2001). Japanese- and English-speaking children's use of linguistic cues to animacy in category formation. Biennial meeting of the Society for Research on Child Development, Minneapolis, MN.

Younger, B. A., & Cohen, L. B. (1983). Infant perception of correlations among attributes. *Child Development*, *54*, 858–867.

Yu, C., & Smith, L. B. (in press). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science*.

de Zubicaray, G. I., & McMahon, K. L. (2009). Auditory context effects in picture naming investigated with event-related fMRI. *Cognitive, Affective & Behavioral Neuroscience*, *9*(3), 260–269.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1315–1356

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01081.x

# Five Reasons to Doubt the Existence of a Geometric Module

Alexandra D. Twyman, Nora S. Newcombe

*Temple University*

Received 26 August 2008; received in revised form 14 August 2009; accepted 15 August 2009

**Abstract**

It is frequently claimed that the human mind is organized in a modular fashion, a hypothesis linked historically, though not inevitably, to the claim that many aspects of the human mind are innately specified. A specific instance of this line of thought is the proposal of an innately specified geometric module for human reorientation. From a massive modularity position, the reorientation module would be one of a large number that organized the mind. From the core knowledge position, the reorientation module is one of five innate and encapsulated modules that can later be supplemented by use of human language. In this paper, we marshall five lines of evidence that cast doubt on the geometric module hypothesis, unfolded in a series of reasons: (1) Language does not play a necessary role in the integration of feature and geometric cues, although it can be helpful. (2) A model of reorientation requires flexibility to explain variable phenomena. (3) Experience matters over short and long periods. (4) Features are used for true reorientation. (5) The nature of geometric information is not as yet clearly specified. In the final section, we review recent theoretical approaches to the known reorientation phenomena.

*Keywords:* Modularity; Adaptive combination; Spatial reorientation; Development; Geometric module

## 1. Introduction

Love and marriage, horse and carriage, modularity and nativism—the last pair of words lacks the ring of the first two pairs, but the relation in each case is the same. The concepts are different, and each may exist separately and independently, yet they seem for the most part to get along naturally and easily. One expects to see them together. There can be emergent modularity (Karmiloff-Smith, 1992) but that is the marked case, just as a marriage of

Correspondence should be sent to Alexandra D. Twyman, Psychology Department, Temple University, 1701 N. 13th Street, Philadelphia, PA 19122-6085. E-mail: atwyman@temple.edu

convenience is a marked version of marriage (to pursue the analogy). Modularity without modification is generally thought to be inborn (Fodor, 1983; but see Fodor, 2000 for a different view). Similarly, although there can be versions of nativism that are not domain specific (Elman et al., 1996), domain-general learning ability is not what is usually meant by nativism. As generally discussed, nativism is what Elman et al. (1996) call *representational nativism*, and hence is domain specific. And then, once native endowments have domain-specific content, they must have neural instantiations, and those instantiations often seem to involve specialized areas. Voilà—something we would call a module. In fact, it has even been argued that one cannot have an evolutionarily informed cognitive psychology that does not involve modules, because natural selection must have a target on which to act (Cosmides & Tooby, 1992). Against this intellectual backdrop, modules have proliferated—the theory of mind module, the cheater detection module, the face processing module, and so forth. Indeed, we have seen claims that the human mind is ''massively modular'' (e.g., Carruthers, 2006). This notion has permeated the popular press. For example, consider the following passage from *Newsweek*, ''Behaviors that conferred a fitness advantage during the era when modern humans were evolving are the result of hundreds of genetically based cognitive 'modules' programmed in the brain,'' or ''evolutionary psychologists claim that human behavior is constrained by mental modules that calcified in the Stone Age'' (Begley, 2009).

The massive modularity position is not the only modularity proposal. Other theorists argue that there are a small number of modules that are the foundation of cognition. One such position is the core knowledge position advocated by Spelke and Kinzler (2007). According to this view, there are five modules that comprise core knowledge: object, action, number, geometry, and social partner representation. In a similar fashion to the Fodorian view of modularity, these modules are domain specific, innately endowed, and shared across species. However, these modules do not persist across the life span from this perspective. According to this point of view, language is the mechanism that moves infants from an innate modular representation to integrated cognition as adults. This core knowledge position is gaining popularity both in the academic and public domains. For example, Vallortigara, Sovrano, and Chiandetti (2009) advocate the core knowledge position, including innate endowment, as a result of their rearing experiments with chicks, writing that ''the similarities in cognitive capacities seen near the start of life gives reason to take seriously the hypotheses that core systems have a long evolutionary history and are largely preserved across the many evolutionary events that distinguish chicks from humans'' (p. 24). The notion of a limited number of core modules is gaining popularity with the general public as well, although perhaps in the personality not the cognitive domain: ''Each of us is born with our own individual level of six big traits: intelligence, openness to new things, contentiousness, agreeableness, emotional stability and extraversion. These modules are built into humans and other animals (apparently squid can be shy)'' (Brooks, 2009).

However, many of the proposed modules, especially those proposed by massive modularity theorists, do not conform to the strict definition proposed by Fodor (1983). Few seem to be encapsulated (i.e., unable to accept relevant information that is not the kind the module is built to process). Most are central to cognition, whereas Fodor thought that modules would involve primarily perceptual phenomena, and that higher-order cognition would prove not

to be modular (and hence, difficult or impossible to study). The word *module* has come to be used in a way that has many connotations but few agreed-upon core characteristics. Indeed, a recent discussion of the meaning of modularity proposed to strip the concept of most of its interesting attributes, including encapsulation, automaticity, neural specialization, and innateness, and yet ended up arguing that such stripping merely clarified the concept, leaving it still valuable because it leads to a focus on function (Barrett & Kurzban, 2006). Given this lack of agreed-upon definition, the modularity position becomes analogous to the Hydra, the many-headed monster that Heracles found difficult to combat because there were too many heads to take on simultaneously, and, worse, because other heads grew while he addressed a specific one.

In this article, we address only one head of the Hydra of modularity: the proposal of a geometric module (Cheng, 1986; Gallistel, 1990; Hermer & Spelke, 1994, 1996). The geometric module has welcome properties from the point of view of engaging in clear debate about modularity. It is well specified, it concerns an interesting aspect of human cognition, and it is increasingly well studied in a variety of species and over development using a variety of techniques. In addition, it has been augmented with an interesting account of how the innately specified module is penetrated in development by the acquisition of specific linguistic terms (Shusterman & Spelke, 2005). Developmental change is always hard for nativists to explain when it is clearly evident yet cannot be dismissed as parameter triggering. In this version of modularity nativism, change occurs because of the supplementary role of another module, the language module. This position is a hybrid of nativism and Vygotskyan or Whorfian thinking. Strongly antinativist theorists sometimes cite the hypothesis approvingly, seemingly without realizing its nativist roots (Levinson, 2003).

Versions of the modularity-plus-language account vary in the strength. Some can be characterized as a strong view of modular cognition (Spelke & Kinzler, 2007). Other versions of the modularity-plus-language positions are gaining momentum. One example is the Momentary Interaction hypothesis, advocated by Landau and Lakusta (2009). In this version, reorientation is still accomplished primarily by a geometric module; the properties of reorientation ''fall within the criterion that Fodor proposed for modular systems, in particular, domain-specificity, localization, ontogenetic invariance, and characteristic breakdown patterns'' (p. 3). The Momentary Interaction hypothesis differs from modularity-plus-language, however, in its view of language. From the modularity-plus-language view, language radically alters the cognitive representation of the human mind. While language still plays a role in the Momentary Interaction hypotheses, it is viewed as a less powerful tool. Within this account, language serves as an attention tool that is a ''flexible and powerful enhancement of spatial representations through language, but not through radical restructuring'' (p. 3).

Core knowledge and modularity-plus-language provide an interesting and tantalizing perspective on the ontogeny of knowledge. Nevertheless, there is good reason to doubt the validity of this approach. This article reviews reasons for doubt. We begin by briefly defining the geometric module and the core research that initially defined it. We proceed to discuss five reasons to question its existence (or perhaps more precisely, four reasons and a counterargument to a recent claim made by modularity proponents). These five reasons, and

Table 1
Five reasons to doubt the existence of a geometric module

| Five Reasons |
|---|
| 1. Language does not play an essential role in the integration of feature and geometric cues |
| (a) Nonhuman animals are able to use geometric and feature cues |
| (b) Adults' feature use is not uniquely dependent on language |
| (c) 18-month-old children can integrate geometric and feature cues in large spaces |
| 2. A model of reorientation requires flexibility to explain variable phenomena |
| (a) The relative use of geometric and feature information depends on room size |
| (b) Flexibility to predict when overshadowing and blocking will or will not occur |
| 3. Experience matters over short and long periods |
| (a) Short-term training experiments demonstrate plasticity |
| (b) Rearing experiments demonstrate plasticity |
| 4. Features are used for reorientation: Evidence against a recent two-step model |
| (a) Reorientation in an octagon |
| (b) Features are used as landmarks for indirect orientation |
| 5. Redefining the analysis of geometric information |
| (a) Not all kinds of geometry are used early in development |
| (b) Use of scalar and nonscalar cues by toddlers |
| (c) Use of scalar and nonscalar cues by mice |

their related subpoints, are summarized in Table 1. This review builds on the literature review by Cheng and Newcombe (2005), but it includes recent papers written since that review and is also more critical and selective in nature. It should be read in conjunction with Cheng (2008), an article in which the original proposer of the geometric module also questions the status of the hypothesis. In the last section, we discuss alternatives to modularity theory, as there are now several proposed theories of the relevant phenomena.

## 2. The geometric module proposal

Mobile creatures need to be able to navigate efficiently through their environment. Occasionally, we lose track of our position in the world, for example, after tumbling down a hill or emerging from a subway system. Before we are able to continue with the task at hand, for example, finding our way home or to work, we need to re-establish knowledge of our position in the spatial world, in a process known as reorientation. The scientific study of this domain started with the seminal work of Ken Cheng (1986). The experiments in his paper showed a rather odd pattern of behavior in disoriented rats. The search space was a rectangular arena full of multimodal features (including distinct odors, lights, patterned panels, or colored walls). In a working memory task, the rat found a food reward and then was removed after eating only a portion of the treat and immediately transferred to an identical enclosure. The crucial observation was where the rat searched for the remaining food. Surprisingly, the rat only returned to the correct location about half of the time. The other half of the time, the rat went to the diagonally opposite position, which might, for example, smell of peppermint when the rewarded place had smelled of licorice. Apparently, the only information being used from the

search arena was geometric information about its shape. As shown in Fig. 1, by combining geometric and sense information, the rat can narrow search for food to the corners with the long wall to the right and a short wall to the left (or vice versa). Although feature information would double the frequency of reward, in working memory tasks (when the correct corner changes from trial to trial) rats reoriented using geometric information to the exclusion of feature cues. However, when the reinforced corner stayed the same across training sessions, in a reference memory task, the rats learned to use feature information.

Cheng (1986) proposed the idea of a geometric module to explain this phenomenon. Reorientation is accomplished first on the basis of metric information and then nongeometric information can be pasted onto this metric frame. This conclusion was subsequently extended to a different task, a different stimulus array and a reference memory situation. In a water maze task, Benhamou and Poucet (1998) demonstrated that rats are able to extract geometric information from an array of discrete and unique landmarks. The landmarks were placed in a circular pool and arranged as the vertices of either an equilateral or isosceles triangle. Even when the geometric information of the isosceles triangle had to be extracted from the separated points in the array, it was still much easier for the rats to learn the location of the hidden platform from geometric than from featural information. Thus, it seems that rats rely primarily on geometric information for reorientation and use feature information only secondarily, if at all.

Gallistel (1990) proposed an evolutionary account for the primacy of geometric information. He noted that many features change from season to season, such as the color of the foliage, or even day to day, such as the clarity of a river, but that geometric information remains much more constant, stable, and reliable, so that it may be evolutionarily adaptive to reorient based on the invariant properties of the environment. Another point to consider is that geometric ambiguities may not arise often in the complex natural world. Thus, even though in experiments animals may be rewarded only half the time when they reorient using only geometric information, in a naturalistic setting the payoff for using geometric information is probably much higher.

Subsequent to this work with rats, researchers asked what human capabilities are and how they appear in ontogeny. Evidence appeared that there seemed to be a developmental

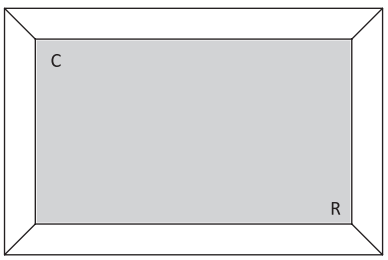


Fig. 1. A schematic of a typical rectangular enclosure for the reorientation studies. The organism finds a food reward at the correct corner, C. On the basis of geometric information, the diagonally opposite corner, R, is equivalent to the correct corner. As an example, the short wall is to the creature's left and the long wall to the right at both the correct and rotationally equivalent corner.

progression for reorientation, from exclusive use of geometric cues early in life, to a more flexible system later in development. At 18–24 months of age, children do not spontaneously use feature information for reorientation in small rectangular spaces, but readily use geometric information (Hermer & Spelke, 1994, 1996). Similarly, disoriented 12- to 18-month-old infants cannot use distinctive landmarks to localize a peekaboo event (Lew, Foster, & Bremner, 2006). For children at or above the age of 6 years as well as adults, feature and geometric information are flexibly combined (Hermer-Vazquez, Moffet, & Munkholm, 2001).

The modularity-plus-language hypothesis proposed that people are able to penetrate the geometric module for reorientation that we share with other animals with the production of spatial language (Shusterman & Spelke, 2005). The hypothesis was supported by relations between the age at using features and the productive capacity to use the terms ''left'' and ''right'' (Hermer-Vazquez et al., 2001), by findings that adults prevented from using language to encode the space by a concurrent linguistic task failed to use features (Hermer-Vazquez, Spelke, & Katsnelson, 1999), and by demonstrations that linguistic reminders and training elicited earlier use of features (Shusterman & Spelke, 2005). Additionally, spatial linguistic cues of ''left'' or ''right'' aid 4-year-old children's performance on a left-right color location memory task (Dessalegn & Landau, 2008) Taken together, the findings from rats, children, and adults seem to present a formidable case for the existence of an encapsulated module that runs automatically, appears early, and is shared across mobile species, with developmental change in humans due only to the intervention of a specifically human capability, namely language.

Fueled by the interest that the geometric module generated, many investigators have examined reorientation over the past 20 years. Although we call the geometric module into question, we would like to point out that the hypothesis has provided a solid starting place for studies of reorientation and still pushes the field to refine its empirical paradigms and theoretical positions. Over the last decade, considerable amounts of counter-evidence have appeared, collectively casting a substantial shadow on this modular picture of development and of the architecture of spatial adaptation. Nevertheless, the data serve to constrain the search for alternative conceptualizations of spatial functioning and its development, and recently, several models have been proposed that contend to explain the relevant phenomena.

In overview, we discuss five reasons to doubt the existence of a geometric module. First, we address the language part of the modularity-plus-language position, discussing reasons to doubt that language plays a unique role in the integration of feature and geometric information. There is evidence that nonhuman animals flexibly use geometric and feature cues for reorientation, that human adults' reorientation ability is not dependent solely on language, and that toddlers are also able to flexibly use geometric and feature information in larger spaces before they possess the relevant language. Second, we review variability in the geometric module phenomena, including the fact that geometric and feature cue use depend on environment size, and that overshadowing, blocking, and facilitation effects have all been observed. These examples show that any successful theory of reorientation must be flexible enough to explain these fluctuations, a difficult challenge for a modularity or core knowledge position. Third, the core knowledge approach postulates that the reorientation

system is innate, and thus downplays the effects of experience on the behavior of reorienting organisms. In contrast to this hypothesis, we demonstrate that experience, both in short-term training experiments and over the long term in rearing experiments, has an important influence on the orientation performance of participants. Fourth, we turn to a recent two-step model of reorientation. Here, advocates of the core knowledge position argue that geometric information is used alone for true reorientation, although subsequently, features can be used associatively to pinpoint a goal location. In contrast, we review evidence that features can be used for true reorientation, both in the presence and absence of geometric information. Finally, we discuss what types of geometric information can be used across development and across species for reorientation. It has become apparent that not all types of geometry are used for reorientation, and that a more specific definition of geometric information is needed. A summary of these five reasons can be found in Table 1.

## 3. Reason 1: Language does not play an essential role in the integration of feature and geometric cues

Initially, there seemed to be quite a bit of evidence to support the hypothesis that language plays a fundamental role in the integration of geometric and feature information. When Hermer and Spelke (1994, 1996) first adapted the reorientation paradigm for use with people, children below the age of 6 years did not use feature information in small spaces. In contrast, older children and adults are able to flexibly combine geometric and feature information. Human language (and in particular the productive use of the spatial relational terms ''left'' and ''right'') was proposed to puncture the geometric module that we share with other species (Shusterman & Spelke, 2005). Additionally, language training interventions facilitated children's use of features. However, there are three kinds of evidence that cast doubt on the hypothesis that language plays an essential role in the integration of geometric and feature cues.

### *3.1. Nonhuman animals use features for reorientation*

On the modularity-plus-language view, nonlinguistic species should clearly have difficulty using feature cues for reorientation. Since the initial work with rats and young children, a wide range of vertebrate species have been studied, including chickens, pigeons, monkeys, and fish, in both appetitive and escape reorientation paradigms, and there are many demonstrations of use of feature cues (see Cheng & Newcombe, 2005 for a review). Recently, an invertebrate representative, the ant, has also been demonstrated to use feature as well as geometric cues (Wystrach & Beugnon, 2009).

There are important distinctions among feature cues, however, that need to be examined to put these data in context. Features may be direct markers of the target location (sometimes called beacons), or they may be indirect markers of a goal location (sometimes called landmarks). In many of the experiments cited above, the experimental paradigm or data analyses did not distinguish between using the feature cue as a beacon or as a landmark. For

example, in a rectangle with discrete feature panels, the feature cue could be used either as a beacon or as a landmark (Fig. 2A). Participants learn that a reward is hidden in a particular corner, near the black panel in this example. After participants are trained to perform well on the task, follow-up tests can be conducted. One of these types of tests, the distal test in Fig. 2C, is particularly useful for distinguishing between features used as landmarks versus beacons. After training, the target feature panel and the rotationally equivalent panel are removed. Therefore, for the participant to be able to return to the correct corner, the distal feature cues must be used, and thus successful search indicates that features can been used as landmarks for reorientation. However, if the black panel is retained, successful search can be based on a beacon strategy. This distinction may seem to be a minor point; however, it is quite important. When there are markers of a specific location (i.e., a beacon), the participant may or may not be using an associative process to guide search, that is unrelated to the reorientation process. In contrast, indirect feature use is strong evidence that features are used to guide the reorientation process, rather than just search strategies.

There has been mixed evidence of use of features as landmarks as well as beacons in geometric module research with nonhuman animals. Pigeons are successfully able to return to the target corner on distal panel tests (Kelly, Spetch, & Heth, 1998). Thus, features can be used as landmarks (rather than just beacons), and this process does not critically depend on language. In contrast, rats, mice and chicks divide search between the two geometrically equivalent corners (Cheng, 1986; Twyman, Newcombe, & Gould, 2009; Vallortigara, Zanforlin, & Pasti, 1990). Therefore, for rats, mice, and chicks, it is only clear that features can be used as beacons. It is also possible that features could be used as landmarks with these species, with altered experimental parameters. As the other feature panels were not explicitly required for the task, it is possible that subjects ignored the other feature panels and did not encode them into their spatial representation. Thus, manipulations that draw

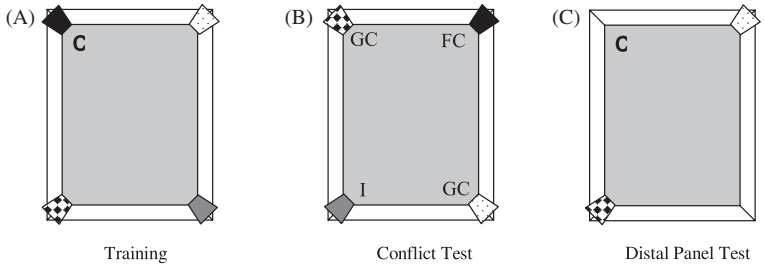


Fig. 2. Typical tests for reorientation paradigms. After training with a feature (A), various follow-up tests are conducted to examine how features and geometry are used. In the conflict test (B), an affine transformation is achieved by rotating each panel one position, clockwise in this example. Now, the correct feature corner (FC) is no longer in a geometrically correct position. The animal may choose to visit either a featurally correct corner or one of the two geometrically correct corners (GC). The animals usually avoid the incorrect corner (I) as it is neither featurally nor geometrically correct. Finally, it is possible that participants could have used either the local features (the panel at the target location—as a beacon) or the distal features (the panels at the other corners—as indirect landmarks) to reorient. This is tested by removing the feature panels at the two geometrically equivalent corners and then observing whether the animals are able to focus their search at the previously trained corner by using the remaining feature information indirectly, as we can see in panel (C).

attention to the nonreinforced feature panels (perhaps by expanding the search size, and therefore making the features more distal) may demonstrate that even for rats, mice, and chicks, features can be used for reorientation as both beacons and landmarks. In sum, while the use of features as landmarks (rather than just beacons) remains an open question for some species, it has been clearly demonstrated that pigeons can use features as landmarks, and this process does not critically depend on language.

An additional argument for the geometric module-plus-language position argues that geometric cues should be more readily used for reorientation than feature cues (either beacons or landmarks). Evidence from the animal literature demonstrates that there are instances where feature information can rival the strength and utility of geometric information, as revealed through conflict tests (Fig. 2B). In conflict experiments, feature and geometric cues are put in opposition by rotating the feature cue, so that the correct feature corner is in a geometrically incorrect position (Fig. 2B). Some animals divide search between the two geometrically equivalent corners and the one featurally correct corner. This is true for mountain chickadees, pigeons, and redtail splitfin fish (Gray, Bloomfield, Ferrey, Spetch, & Sturdy, 2005; Kelly et al., 1998; Sovrano, Bisazza, & Vallortigara, 2007). Moreover, some animals choose the featural corner over a geometrically correct one. For chicks, this was true only when the features directly marked the correct location (as a beacon, as discussed above) (Vallortigara et al., 1990). Rats follow the feature to a geometrically incorrect corner, which is perhaps surprising as the strongest evidence for a geometric module has been found for this particular species (Wall, Botly, Black, & Shettleworth, 2004). Thus, there appear to be many instances where feature information can rival the strength and utility of geometric information.[1]

Why is there a contrast with the original Cheng findings—that rats seemed to exclude nongeometric information—while many other nonhuman animals (and perhaps even rats) can use feature information for reorientation? It was originally proposed that the geometric module operates only in working memory tasks (as the defining pattern of rotational errors was only found when the target corner changed from trial to trial) in Cheng's (1986) work. All other tests of reorientation in nonrat species have been done with a reference memory paradigm. There are several problems with this argument, however. First, Benhamou and Poucet (1998) found the encapsulation effect in a reference memory task with rats. Second, the work with human children has used a reference memory paradigm too, so the contact between the animal and the developmental literatures is tenuous if the importance of working memory is stressed. Third, it is hard to see how reorientation in the natural world is less important when stable (reference memory) rather than variable locations (working memory) must be retained. That is, a module that operates in working memory but not reference memory would seem to be comparatively insignificant from an adaptive point of view. Nevertheless, the existing findings can be largely (if not completely) reconciled if one accepts that encapsulation is only evident in working memory paradigms.

Advocates of the modularity position also argue that the extensive training required to test reorientation in animals supports learning of a different nature than possible with a linguistic toolkit. For instance, Hermer-Vasquez et al. (2001) wrote that ''no result has yet conclusively shown that nonhuman [animals] show the spontaneous, flexible indirect landmark use found with human adults' reorientation'' (p. 267). Pigeons have been

demonstrated to use landmarks indirectly as reorientation cues (Kelly et al., 1998). However, training is still required for pigeons to perform that task, and thus the reorientation ability of extensively trained animals may be different from the spontaneous reorientation ability of people. Thus, we turn to work with humans.

### *3.2. Adults' feature use is not uniquely dependent on language*

One of the most striking findings supporting the modularity-plus-language position is Hermer-Vazquez et al.'s (1999) finding that human adults required to do a verbal shadowing task (but not a control task that seemed equally attention-demanding) fail to use a feature as large as a colored wall to guide search. Without access to language, they seem to revert to the core knowledge mode of functioning, similar to that of children and nonhuman animals. However, it is possible that the verbal shadowing task used by Hermer-Vazquez et al. (1999) might disrupt the ability to use featural landmarks not (or not only) by interfering with a linguistic encoding process within a geometric module. The nonverbal rhythm-clapping task used by Hermer-Vazquez et al. (1999) is ill-suited to control for this possibility because it involves primarily cerebellar regions of the brain (Woodruff-Pak, Papka, & Ivry, 1996) and would not be expected to engage spatial coding systems. A nonverbal *spatial* task might interfere with the integration of geometric and featural information in the reorientation task.

In confirmation of this idea, Ratliff and Newcombe (2008a) found that participants required to verbally shadow while doing the reorientation task were less likely to use features to reorient, just as Hermer-Vazquez et al. (1999) had reported, although not reduced to chance levels. A nonverbal spatial task also produced a similar dampening effect on the use of features. Hupbach, Hardt, Nadel, and Bohbot (2007) also report that a spatial as well as a verbal task impairs spatial reorientation, and they found that verbal shadowing did *not* disrupt use of features in a square environment. Taken together, these studies strongly suggest that language is not crucial in combining information from geometric and nongeometric sources in order to reorient, although it may sometimes be helpful. Instead, the combination of featural and geometric information may depend on cognitive and neural mechanisms that are involved in both spatial and verbal attention and memory (Newcombe, 2005).

### *3.3. Young children can flexibly integrate feature and geometric cues in larger spaces*

Although more complete discussion of the effect of room size will follow in reason 2, a preview of this reason is essential to reviewing the role of language in feature use. The initial studies of human toddlers were conducted in very small environments. When the scale of the space is increased, even by as little as to double the dimensions of each wall, children as young as 18 months are able to flexibly integrate geometric and feature cues. Children of this age are not yet able to produce the spatial terms ''left'' and ''right,'' which has been claimed to predict success at using the feature wall for reorientation (Shusterman & Spelke, 2005). Thus, there is reason to doubt that language penetrates the geometric module for children.

### 3.4. Summary

In summary, there is substantial evidence to doubt that language is crucially important for the integration of geometric and feature information in humans. First, nonhuman animals have been demonstrated to flexibly combine geometric and feature information, sometimes preferring feature information, and pigeons (at least) can use landmarks as indirect reorientation cues. Second, for human adults, although language can be a tool for reorientation, it is not a necessary mechanism for successful reorientation. Finally, children as young as 18 months old, before they are able to produce relevant spatial language, are able to flexibly integrate geometric and feature cues under certain conditions. Thus, there is reason to doubt that language is required to penetrate the proposed geometric module for reorientation.

## 4. Reason 2: Models of reorientation require flexibility to explain variable phenomena

Since the early experiments of Cheng (1986) and Hermer and Spelke (1994, 1996), many studies have shown that some of the crucial phenomena depend on the parameters of the experiment. These facts are difficult to explain from a core knowledge or modularity position. In the text that follows, we will review the two central phenomena: room size effects, and the presence and absence of overshadowing, blocking, and potentiation effects.

### 4.1. The relative use of geometric and feature information depends on room size

Features are more likely, and geometric cues are less likely, to be used for reorientation as the size of the enclosure increases. Importantly, the room size effect is common across children, adults, and nonhuman species. Thus, the room size effect may depend more on the salience of the cues, rather than development or cross-species difference per se.

#### 4.1.1. The room size effect for children

Hermer and Spelke's (1994, 1996) original experiments made an odd choice of experimental environment, given that the overall aim of research in spatial orientation is to understand how we manage to navigate in natural environments, which are generally fairly large. Specifically, they used a rectangular enclosure that was only 4 feet by 6 feet in size. In a space with walls of length in the same proportional relationship (8 by 12 feet), but with four times the square footage, Learmonth, Newcombe, and Huttenlocher (2001) found that children of 18–24 months succeeded in using a variety of features to reorient, including features that only indirectly mark the correct corner. Learmonth, Nadel, and Newcombe (2002) confirmed the central role of scale, manipulating room size in an experiment with preschool children. More recently, Smith et al. (2008) also showed the importance of scale, manipulating array size in a naturalistic environment. Children between the ages of 3 to 7 years were able to successfully reorient with feature cues, although performance increased with age. Additionally, the size effect was replicated in this natural environment, as children were more accurate using feature information in the larger arrays than the smaller arrays.

However, it should be noted that the Smith et al. experiment was conducted outside, and hence the study was not able to control for potential confounding cues such as the position of the sun. Nevertheless, the study represents an important extension to studying how orientation might be guided in a naturalistic setting. Even the ''larger'' room used in the Learmonth research is not very large, in terms of the real world. Thus, for children, feature cues appear to only be neglected in small spaces.

#### *4.1.2. The room size effect for adults*

Does the room size effect that was demonstrated for children hold for adults? Ratliff and Newcombe (2008b) provide evidence that room size influences human adults' use of geometric and feature cues. Adults participated in four trials in either a small or a larger rectangular room with a feature panel. After the training trials, the adults were administered two conflict trials in the same sized rectangle. Thus, adults who were trained in the small room were administered the conflict trials in the small room and adults who were trained in the large room participated in the conflict trials in the large room. Adults who were in the small room group selected a geometrically correct corner, while adults in the larger room group decided to go with the featurally correct corner. Thus, it appears that for humans, children and adults alike, the size of the room is a vital determinant of when geometric and feature cues are used for reorientation.

#### *4.1.3. The room size effect for nonhuman animals*

The role of enclosure size has also been confirmed in experiments with nonhuman animals, specifically chicks, fish, and rats. Chicks trained to find food at a distinctive panel in either a small or a large rectangle showed higher use of feature information when trained in the larger one (Chiandetti, Regolin, Sovrano, & Vallortigara, 2007). Room size also influenced decisions in conflict situations. In the large room, the chicks followed the feature, while in the small room, the chicks divided their search between the two geometrically equivalent corners (Sovrano & Vallortigara, 2006). Redtail splitfin fish showed similar patterns (Sovrano et al., 2007). Similarly for rats, increasing the size of the enclosure resulted in increased attention to the feature wall, and decreasing the enclosure size increased attraction to the geometric cue (Maes, Fontanari, & Regolin, 2009). Thus, there seems to be something about the experimental parameters that increases geometry use in smaller spaces and augments feature use in larger spaces.

#### *4.1.4. Possible reasons for the effect of room size*

Why does the enclosure size matter? There are several possible explanations for the room size effect. First, it is possible that the two sources of information are not combined in a small space because the small space restricts movement. Restrained rats do not seem to learn the same information about a maze that freely moving rats do (Foster, Castro, & McNaughton, 1989), and children perform more accurately when they are actively moving, rather than passively moved (Acredolo, 1978; Acredolo & Evans, 1980; McComas & Dulberg, 1997). Additionally, geometric information may be more dominant in small enclosures because the lengths of the walls, and aspect ratios, are more easily observed (Sovrano

& Vallortigara, 2006). Second, distal rather than proximal features provide more precise information about location as movement occurs (Gallistel, 1990; Nadel & Hupbach, 2006; Vlasak, 2006). In the large room, the blue wall is (often) farther away and could be a sufficiently distal cue to be weighted more heavily by the reorientation system. It is possible that this effect may reflect the underlying properties of hippocampal place cells. Place cells are particularly attuned to distal features, rather than proximal cues (Cressant, Muller, & Poucet, 1997), and this may be why feature use increases with increasing apparatus size. Additionally, as the size of the testing space has now been shown to affect the location of hippocampal firing in exploring rats (Kjelstrup et al., 2008), greater attention to spatial scale in studies of spatial functioning is clearly warranted across a wide range of species.

Learmonth, Newcombe, Sheridan, and Jones (2008) report a series of experiments designed to examine what processes might support the room size effect. Specifically, the role of restricted movement and landmark distance was examined for young children between the ages of 3 and 6 years. The basic principle of these experiments was to restrict children's activity to an area the size of the small space used in the Hermer and Spelke (1994, 1996) experiments, within an area the size of the larger space used in the experiments by Learmonth et al. (2001, 2002). Although the children's motion was restricted to the small space, they had visual access to a larger space, containing a distal feature, and with its characteristic aspect ratio. Table 2 summarizes the findings regarding ages at which features are first used to reorient as a function of whether children could move freely in the space, whether the colored wall was distally located in a larger enclosure, and a third factor that turns out to be important—whether the target for which children searched after disorientation was adjacent to the feature. Comparing across studies in which two of these factors were constant allows us to draw inferences regarding whether the third factor affects the age at which successful use of features is first observed. From the contrast between rows 2 and 4 in Table 2, we can infer that restriction of movement has a powerful effect on children's ability to use the landmark to reorient, as children that are able to move freely are able to use the feature at 18 months of age, where children with restricted motion cannot use the feature until 4 years of age. Similarly, the contrast between the Hermer and Spelke (1994, 1996) studies and Experiments 2 and 3 in Learmonth et al. (2008) (rows 1 and 4 in Table 2) shows that whether the colored wall was distal from the child in a larger room has an important effect on children's ability to use the landmark to reorient, namely children can use

Table 2
Age of success in rectangular spaces as a product of variations in the task demands

| Experiment | Colored Wall Distal? | Action Possible? | Target Proximal to Colored Wall? | Age of Success |
|---|---|---|---|---|
| Hermer & Spelke | No | No | Yes | 6 years |
| Learmonth et al. | Yes | Yes | Yes | 18 months |
| Experiment 1 | Yes | No | No | 6 years |
| Experiments 2 and 3 | Yes | No | Yes | 4 years |
| Experiment 5 | Yes | Yes then No | Yes | 3 years |

*Note*. From Learmonth et al. (2008, p. 423). Reprinted with permission.

distal features earlier (4 years) than proximal features (6 years). Finally, the comparison between Experiment 1 and Experiments 2 and 3 in Learmonth et al. (2008) (rows 3 and 4 in Table 2) shows that whether the target was located adjacent to that wall or one of the white walls in the larger enclosure is important, in that correct searches with a distal feature appeared earlier in development when the target was directly adjacent to the feature wall (4 years) than at an all-white corner (6 years). Thus, the ability to move around and the presence of distal (rather than proximal) features are two properties of larger spaces that enable children to use features at younger ages in development.

#### *4.1.5. Summary of room size effects*

Across a wide range of species, the size of the search space matters for reorientation. For all ages of people, and across species, geometry is used more predominantly in small spaces, and feature information in larger spaces. This room size has empirically been demonstrated to hinge on at least two important principles: Movement enhances spatial navigation, and distal landmarks are more likely to be used for reorientation. For the modularity position to be able to explain the room size effects, it would have to propose that reorientation is only modular in small spaces. It would then be a challenge to imagine why it would be advantageous to have a modular system in small spaces, when navigation through the environment occurs on a much larger scale.

### *4.2. Fluctuating findings on overshadowing, blocking, and facilitation*

In addition to the fluctuating integration of feature and geometric cues based on room size, research has found a variable presence versus absence of overshadowing, blocking, and facilitation effects that is difficult to explain on a modularity view.

#### *4.2.1. Principles of associative learning*

First, let us review a few of the principles of associative learning. In overshadowing, when predictive cues are presented together, less may be learned about each cue than if they had been presented independently. For example, if a bright light and a faint tone predict a food reward, the more salient bright light *overshadows* the faint tone. As a result, the light controls behavior while the tone does not control behavior, although the faint tone would have been learned if presented on its own. In spatial tasks, beacons (landmarks at the goal location) can overshadow more distal landmarks, as has been demonstrated for rats (Diez-Chamizo, Sterio, & Mackintosh, 1985) as well as pigeons and humans (Spetch, 1995). In blocking, if a particular cue is learned first, and then is subsequently paired with a second cue, the second cue is not associated with the reward. For example, the participant first learns that one cue, such as a light, is predictive of a reward. Then the light is presented with a second cue, such as a tone. Even though both the light and the tone are equally predictive of reinforcement, the past learning history of the organism *blocks* the learning of the new cue. The principle of blocking also seems to apply to the spatial domain since blocking has been demonstrated in landmark learning with oriented rats (Biegler & Morris, 1999; Roberts & Pearce, 1999; Rodrigo, Chamizo, McLaren, & Mackintosh, 1997).

*4.2.2. Associative learning and the reorientation paradigm*

Do the principles of overshadowing and blocking hold for reorientation? Initially there were reports of a failure to find the traditional associative learning effects for geometric cues. Feature cues did not seem to block the learning of geometry. Wall et al. (2004) trained rats to find food in a square arena where one black feature panel indicated the food location. Once the rat learned to find the food, the black panel was presented within a rectangular search arena. Blocking would predict that the rats would not encode the geometry of the space because they already know that the black panel predicts food location. In contrast, geometry-only test trials revealed that the rats were able to use the geometric information for reorientation.[2] Additionally, feature cues did not seem to overshadow the learning of geometry (Cheng, 1986). These findings were taken as strong evidence for the modularity position. As the cues were not interacting with each other, it would seem that geometric information is processed at least separately, if not exclusively, from feature information.

However, an absence of overshadowing or blocking effects is not always observed in studies using feature and geometric cues. For rats, both overshadowing and blocking have been demonstrated. Rats were required to find a submerged platform in the corner of a rectangular arena with a feature wall. If the feature wall had previously been trained, then the feature blocked the learning of the geometry. When the feature wall was presented together with the geometric information, then the feature information overshadowed the geometric information (Pearce, Graham, Good, Jones, & McGregor, 2006). Additionally, in a recent set of experiments, a failure to find blocking of geometry by a beacon that was suspended over the platform was found when 12 blocking sessions were administered, although if the blocking sessions were doubled to 24 sessions, then the beacon cue blocked the geometry of the enclosure (Horne & Pearce, 2009). For mountain chickadees, when the blue feature wall was adjacent to the food reward, the geometry-only probe trials revealed that the chickadees had not encoded the overall shape of the arena. Thus, feature information overshadowed geometric information for this species in some training situations (Gray et al., 2005, but see Batty, Bloomfield, Spetch, and Sturdy, 2009, for black-capped chickadees).

It has recently been found that features can even augment (rather than block or overshadow or have no effect on) the acquisition of geometry cues, in a phenomenon called potentiation or facilitation. Rats searching for a hidden platform in an enclosure shaped like a kite did better on a geometry-only test if trained with both feature and geometric information than if trained only with geometry (Graham, Good, McGregor, & Pearce, 2006). Hence, there is evidence that the two types of orientation cues interact with each other, and this is difficult to explain from a modular position.

*4.2.3. How should traditional association effects, or a lack thereof, be interpreted?*

How should association effects be used? Are they gold standards that can support or refute a modular system? If so, then clearly the results are quite mixed. Or are they variable phenomena that should be explainable and predicted by a theoretical position? We discuss this point further in the last section, as at least one current model of the findings was developed specifically to account for variable effects.

However, let's now move up in our level of analysis. There has been a long debate in the broader field of spatial cognition concerning whether mobile organisms navigate with response learning or place learning (Hull, 1943; Tolman, 1948). A large part of this debate has centered on the success or failure to find associative learning. On one side, the response learning tradition has found instances of associative effects (Hamilton & Sutherland, 1999). On the other side, place learning theorists, or cognitive map theorists, have failed to find blocking and overshadowing, as predicted by their theoretical position (Hardt, Hupbach, & Nadel, 2009). The cognitive map approach proposes an explanation for the existence of blocking and overshadowing effects, centered on the role of exploration. When participants are allowed opportunities to explore, then cognitive map theorists maintain that the organism is spontaneously, independent of reward, creating and updating a cognitive map. Thus, when blocking and overshadowing effects are found, this theory proposes that participants have not been given an opportunity to explore, and thus a cognitive map was never formed during the experiment and the participant manages as best he/she can through associative learning.

The important point here is that, in this debate, a failure to find blocking and overshadowing is taken as support of a cognitive map—a unified representation. In contrast, in the reorientation debate, a failure to find blocking and overshadowing is taken as support of a modular system. Thus—should blocking and overshadowing be used as a gold standard? We would like to argue, in light of what we have just reviewed, that it would be a more fruitful approach to acknowledge that the most productive line is to determine the conditions that lead to the variety of findings. This variation is one of the facts in this literature that a successful theory should be able to explain, and modularity theory clearly does not have the required flexibility. At the same time, as we explain later in the article, some of the theories that do not assume modularity (e.g., Miller & Shettleworth, 2007; Newcombe & Ratliff, 2007) can successfully handle this flexibility.

### *4.3. Summary*

Since the early experiments of Cheng (1986) and Hermer and Spelke (1994, 1996), many studies have been conducted. In the first section, it was demonstrated that the size of the reorientation enclosure drove the relative use of geometric (in small spaces) and feature (in larger spaces) cues. This phenomenon was found across development and across species. Larger spaces allow for movement that makes it easier to form an integrated representation of the space, and distal features are more likely to be integrated than proximal features. The second section reviewed the presence or absence of blocking, overshadowing, and potentiation effects. A failure to find these associative effects initially supported modularity, as it demonstrated that feature and geometric cues were not interacting with each other. However, future research demonstrated that under certain conditions, blocking, overshadowing, and even potentiation effects have all been found for spatial reorientation. The core knowledge or modularity positions do not have the flexibility to explain either kind of fluctuating phenomena. Next, we turn to the role of experience in reorientation behavior.

## 5. Reason 3: Experience matters over short and long durations

The modularity position predicts that reliance on geometry alone should be difficult or impossible to modify (except by the intervention of language) and in fact, language training does work to help children use features (Shusterman & Spelke, 2005). However, training effects not dependent on language have been found for adults, young children, and pigeons. In addition, rearing conditions seem to be important. Here we review data that indicate that experience matters both in short-term training studies and for long-term rearing studies.

### *5.1. Training experiments*

Training experiments have been conducted with quite a range of participants, including adults, children, and pigeons. In all of these studies, short-term experience influenced reorientation. Ratliff and Newcombe (2008b) provide evidence that training affects human adults' use of geometric and feature cues. Adults participated in four training trials in either a small or a larger rectangular room with a feature panel. After the training trials, the adults were administered two conflict trials in the opposite sized room (for example, adults who had been trained in the small room were tested in the large room). For adults trained in the larger room and then tested in the small room, the first choice was often to the featurally correct corner. In addition, for adults who practiced the task in the small room, where geometry is more salient than features, and then were tested in the larger room, the conflict choice was also to the featurally correct corner. Thus, there is an asymmetric relationship between the trainability of feature and geometric information for adults. If training were to have an equal effect on the conflict trials, then one would have predicted that practice using geometry in the small space would have transferred to the larger space, but this did not occur. It is possible that the participants did not trust the enclosure information after a change in scale. On these conflict tests, the feature remains unchanged from training to testing. Even though the ratio of the long to short walls was held constant across expansion or contraction, participants did report noticing the change in enclosure size and thus may have relied more heavily on the feature information. Regardless, adults' responses were influenced by the short-term experience.

Training effects have also been found with young children (Twyman, Friedman, & Spetch, 2007). Four and five-year-old children are not normally able to integrate feature and geometric information in small, $4 \times 6$ foot enclosed spaces. However, after a small number of training trials, between 4 and 12 trials, these young children are able to flexibly integrate feature and geometric information in small enclosures. This training is equally successful when practice with the feature cue is administered in presence (rectangle) or absence (equilateral triangle) of unique geometric cues. Similar findings are reported by Learmonth et al. (2008, Experiment 5) who used four trials of training in the larger room, and then found feature use in the small room. Thus, children's prior experience influences the relative use of feature and geometric information. For the core knowledge position (Kinzler & Spelke, 2007; Spelke & Kinzler, 2007), each of the five systems is characterized by signature limitations or cognitive errors. It is difficult to imagine that the limited number of training trials,

as few as four, would be sufficient to overcome the characteristic limitation as outlined from the core knowledge position, which are claimed to persist into adulthood, albeit in attenuated and more flexible forms.

Pigeons' choices are also affected by their training regimen, and these effects are clearly not dependent on language. When pigeons are trained with rectangular geometry alone, and then given subsequent trials in which features are added and features and geometry are put into conflict, they divide their search between the geometric and feature corners. When both geometry and feature cues were present from the start of training, the pigeons followed the feature panel in conflict trials (Kelly et al., 1998). These findings suggest that geometry is used in conflict situations only when it has had the advantage of initial training alone, whereas features are used even without that advantage. Therefore, short-term reweighting of the relative use of feature and geometric information for reorientation has been demonstrated for adults, children, and pigeons.

### *5.2. Rearing conditions count*

If short-term exposures in the laboratory influence use of geometry versus features, it seems reasonable to suppose that an organism's natural environment will also influence its behavior, perhaps especially the characteristics of the environment to which an immature organism is exposed. Many of the species that have been tested for their use of geometry have been raised in the geometrically regular environments of laboratories or houses. Would use of geometry be as prevalent when organisms have been exposed to environments with few regular geometric enclosures? This question was raised by Cheng and Newcombe (2005), and there are now several studies that attempt to answer it, with birds, fish, and mammals.

Gray et al. (2005) examined the use of feature information in wild-caught chickadees from forested mountain areas rich in feature information, but with little salient geometric information. When the rewarded corner was directly adjacent to the feature corner, the chickadees did not encode the overall shape of the enclosure. When the rewarded corner was at an all-white corner, chickadees were able to use the geometric information. On conflict tests, chickadees that had been trained to go to the feature adjacent to the food focused their search on almost every trial at the featurally correct but geometrically incorrect corner. In contrast, the chickadees who were trained to find food across from the feature wall divided their searches evenly between the featurally and geometrically correct corners. Therefore, the use of geometric information was not a dominant strategy for mountain chickadees. Recently, the same research group has addressed part of this question by comparing laboratory reared black-capped chickadees, wild-caught black-capped chickadees, and wild-caught mountain chickadees (Batty et al., 2009). For the black-capped chickadees, there were no differences between the two groups. The wild-caught mountain chickadees relied less on geometric information than either rearing groups (hand-reared or wild-caught) of black-capped chickadees.

While these results are tantalizing, it remains uncertain if the differences were because of the species or the rearing environment, as wild-caught and laboratory reared mountain

chickadees were not compared. In a controlled laboratory environment, using a different species, Brown, Spetch, and Hurd (2007) altered the rearing environment of fish (*Convict cichlids*). Half of the fry were raised in uniform white circular tanks lacking unique geometric information. The others were raised in white rectangular tanks where geometric information was salient. After 4 months in these environments, half of the fish in each rearing group were trained in either an all-white geometry condition or a rectangle with a blue feature wall adjacent to the correct corner. The circular-reared fish learned the feature training task faster than the geometrically reared fish. When feature and geometric information was placed in conflict, the circular-reared fish selected the featurally correct corner, while the rectangle-reared fish selected a geometrically correct corner. This supports the idea that early experiences can augment feature use for reorientation, contrary to modularity theory, and appears to reweight the hierarchy of orientation cues, in support of adaptive combination theory.

Vallortigara et al. (2009) have critiqued one aspect of the fish rearing study. They propose that as the fish were reared for the long period of time in groups of fish, that the experimenters may have ''directly exposed the experimental fish to geometrical and featural information as visible on conspecifics' bodies, and in particular favored using the individual conspecifics' location as cues for spatial orientation and navigation'' (p. 22). However, there are a few reasons why this is an unlikely concern. Fish in both conditions were reared in the same types of groups, and therefore the differences in behavior are unlikely to arise out of the rearing dynamics. Second, although there are markings on the fish, which can contribute to normal visual system development, it is not clear how markings on the body of the conspecifics might contribute to reorientation performance, because the fish are moving in the enclosures, and thus are not stable reference points either for orientation or navigation.

The critics of the fish study have conducted their own rearing studies with chicks. Chiandetti and Vallortigara (2008) raised all male chicks for 3 days in either circular or rectangular enclosures. Over the next 3 days, chicks were trained to find food in a rectangular apparatus that was either uniformly white or had a unique feature panel at each corner. During training, both groups of chicks made geometric errors and they required the same number of trials to learn the task. Additionally, when the feature panels were removed, chicks spontaneously encoded the geometry of the enclosure irrespective of rearing condition. Thus, for chicks, the early rearing environment does not seem to influence reorientation, in contrast to the more flexible system of fish. However, the rearing studies with fish and chicks differ in several ways. The fish were raised in distinctive environments for a much longer period of time than the chicks. Additionally, fish have an extended juvenile period, which may support cognitive flexibility.

Recently, Twyman et al. (2009) conducted a rearing study with a mammalian species—the mouse. Mice were housed in either circular environments (which were featurally enriched) or rectangular environments (which were geometrically enriched). Young mice that were housed in the circular environment were faster to learn to use features during training. In contrast, young mice that were housed in the rectangular environment were more accurate using geometric information when it was not explicitly required for the task (i.e., when the feature panels were removed from the training rectangle). Thus, for young mice, the rearing environment alters the use of feature and geometric cues for spatial reorientation. As an extra

component of this study, the plasticity of adult and juvenile mice was compared. Interestingly, adult mice retained some plasticity. The rectangular housed adult mice retained the advantage using geometric cues. In contrast, the circular housed adult mice did not outperform their rectangular adult mice counterparts on tests of feature cue use. Thus, for mice, experience plays an important factor, particularly during the juvenile period, but also for mature participants. Thus, initial studies indicate that mice, convict cichlids, and mountain chickadees seem to display larger rearing effects than domestic chicks or black-capped chickadees.

### *5.3. Summary*

Experience matters. In the first section, we reviewed evidence that short-term training experience alters the relative use of feature and geometric information for human adults, children, and pigeons. Each of the experiments is important for different reasons, from the modularity or core knowledge positions. First, since pigeons were studied, it is unlikely that language training is the catalyst for feature use. However, the rebuttal of the modularity position is that the extensive training required for nonhuman animals is not on par with the fluid reorientation system of humans. Thus, the short-term training experiments demonstrate that with quite limited exposure, adults and children's reorientation strategies are altered by experience. In the second half of this point, it was demonstrated that experience also matter over long-term rearing studies, for mountain chickadees, fish, and mice. These experiments are problematic for the innate endowment positions of modularity and core knowledge theory. Next, we turn to a recent reformulation of the geometric module hypothesis.

## 6. Reason 4: Features are used for true reorientation

We have offered three reasons so far to doubt the existence of a geometric module, by refuting the unique role of language, by demonstrating that modularity theory is too rigid to be able to explain variable phenomena, and by refuting the innate endowment claim through demonstration that experience alters reorientation. There is a recent rebuttal, however, that attempts to rescue the geometric modularity proposal by advocating a two-step account in which feature use is merely associative. We review evidence that contradicts this claim and that demonstrates that features can be used for true reorientation.

### *6.1. The two-step model*

Lee, Shusterman, and Spelke (2006) suggest that there are two separable systems of spatial processing and that only the geometric system is used for reorientation per se. Features can be used, but not for reorientation. They can only be used as beacons (as direct markers of a goal location). To support this argument, they disoriented 4-year-old children in an all-white circular space containing three hiding containers arranged as an equilateral triangle. One of the containers had a distinctive color and shape. Although children could find objects hidden in the distinctive container, they failed to use it to choose between the two other

identical containers. Based on this finding, Lee et al. argue that ''search behavior following disorientation depends on two distinct processes: a modular reorientation process…and an associative process that directly links landmarks to locations'' (p. 581).

It might be argued that prior data already contradict the two-stage associative account. Specifically, recall that Learmonth et al. (2001) showed that children's search for an object hidden in an all-white corner of a rectangle with one blue wall was as good as their search for an object hidden in the blue-and-white corner; it may seem initially that the all-white corner provides no associative cue for such performance. However, that characterization is not correct in the two-step account. In a rectangular room, there is only *one* all-white corner that is geometrically correct, so ''all whiteness'' marks the corner as distinct from the geometrically correct alternative as much as the ''blue and whiteness'' marks the other geometrically congruent corner as correct. In fact, ''all white'' is used as one of the pieces of encoded information in a recent associative model of the reorientation task (Miller & Shettleworth, 2007). Lee et al. would clearly argue that reorientation of the kind at stake in the geometric module debate implies that people can use a feature to choose correctly among *more than one* all-white corners with the same geometric characteristics. Testing this hypothesis would require the use of an enclosure with more than four sides.

### *6.2. Evidence against the two-step model*

One reason that Lee et al. may have failed to find that 4-year-olds use features to reorient may be the fact that the feature they used was extremely proximal to the layout (in fact, was part of it) and that the feature was obviously moveable. As we have seen, distal landmarks are known to be more useful than proximal ones for spatial functioning in general and reorientation in particular (Learmonth et al., 2008; Nadel & Hupbach, 2006). In addition, moveable landmarks are less likely to be used to guide spatial search than landmarks that are larger and apparently unlikely to move (Gouteux, Thinus-Blanc, & Vauclair, 2001; Presson & Montello, 1988). Thus, the two-step model of reorientation—with geometry guiding true reorientation, and then features as direct markers of a goal location (i.e., a beacon)—seemed possible.

Newcombe, Ratliff, Shallcross, and Twyman (2009) addressed the issue of whether larger and more distal features can be used for true reorientation, using two approaches. The first step was to use an enclosed octagonal search space with alternating short and long all-white walls. The use of the octagon is interesting for a few reasons. Not only is the geometry more complex than that generally used (obtuse angles, and more potential hiding locations), but it is radial symmetric and therefore lacks a single principle axis of space. Recently, Cheng and Gallistel (2005) have proposed that geometry is used for reorientation by encoding the principle axis of space, and then maintaining the correct left-right position along this line.

In Newcombe et al.'s first experiment, 2- and 3-year-old children were able to select a geometrically correct corner in an octagon 70% of the time, significantly greater than chance of 50% (as there are four geometrically correct, and four geometrically incorrect corners). Thus, young children were able to use geometry for reorientation in spaces lacking a single principle axis of space, and therefore reorientation does not seem to solely depend on

the encoding of the principal axis of the shape of the search area. In a second experiment, a feature wall was added to the octagon search space. For this experiment, 3- and 5-year-old children were studied, spanning the age of the children (4 years) in the Lee et al. (2006) study. By adding a feature wall, there is now one geometrically correct corner directly adjacent to the feature wall—and this location can be solved with a beacon strategy. However, there are three other geometrically correct corners that are all white. At these corners, if children are able to use the feature wall, then they must process the feature wall as a landmark, an indirect use of the feature, to successfully reorient. The Lee et al. two-step account predicts that children would first reorient based on the overall geometry of the space, narrowing the search down to the four geometrically equivalent corners. Next, the two-step account predicts that children will successfully search at the beacon target corner, but crucially not at the landmark search corners. With the feature wall present, children were still able to successfully reorient with the geometry of the space, which is explainable by both the two-step and a unified account of reorientation. The crucial comparison is for search at the beacon (adjacent to red) and the landmark (three all-white, geometrically equivalent corners). In the beacon condition, all children performed above chance, which is predicted by the two-step account. In the landmark conditions, all children performed above chance, which is unexplainable with the two-step account to reorientation. Additionally, the 5-year-old children were better (68%) than the 3-year-old children (35%) using the feature as a beacon. This finding is quite difficult for the two-step account. Not only are young children using features as landmarks, but it seems that the beacon system is emerging later in development than the landmark system.

Now that indirect feature use has been demonstrated in the presence of geometric information, the reorientation ability of children was also examined in the absence of geometry (Newcombe et al., 2009; Experiment 3). This experiment was interesting for two reasons. The first is to ask whether the presence of useful geometric information is required for children to be able to use features to reorient. In other words, is geometry required as a catalyst, or can the reorientation system be activated exclusively with feature cues? Second, the triangle array is the closest comparison to the experimental design of Lee et al. (2006). To maintain as tight of a comparison as possible with Lee et al.'s study, 4-year-old children participated in this experiment. In the first portion of the experiment, children were asked to search within an equilateral triangle search array. In the Lee et al. study, the landmark was a uniquely shaped and colored hiding location. In the Newcombe et al. study, all of the hiding locations were identical, and the landmark was displayed on the perimeter of the search space (a circular space made out of a uniformly white curtain). Lee et al. found that children were able to use the feature cue as a beacon, but not as a landmark. In contrast, when the feature was positioned on the wall, the feature was used as a landmark in Newcombe et al.'s study. However, it is possible that the 4-year-old children were able to infer geometry between the three hiding locations and the feature curtain hanging on the circular wall. To rule out this possibility, the feature served as one point of the equilateral triangle, and then children we asked to search between two containers (composing the rest of the triangle) equidistant from the feature wall. The two-step associative account predicts that children should search equally often at each of the hiding locations. However, 4-year-old children

were able to focus search on the correct hiding location. Thus, indirect feature use is successful in both the presence and absence of geometric information, refuting the two-step modular account.

### *6.3. Summary*

A recent revised modular, two-step account has been proposed for reorientation (Lee et al., 2006). In the first step, disoriented participants reorient with the geometry of the space. In the next step, participants can use features only as beacons to home in on a goal location, but crucially, features cannot be used as landmarks for orientation. In contrast to this hypothesis, we presented data that with stable features, children are able to use features as landmarks. Additionally, the presence of geometric information is not required to activate the reorientation system. Hence, early reorientation is not modular, at least not in the sense of Fodor (1983), in contradiction to the arguments of Lee et al. (2006) and the core knowledge position (Spelke, 2008).

## 7. Reason 5: What exactly is the nature of geometric information?

In reason 5, we question what is meant by geometric information. It is implied in the term *geometric module* that any type of geometric cue should be able to support reorientation. However, we will first demonstrate that not all geometric cues are created equal. Next, we will turn to the specificity of the geometric module. Modularity and core knowledge theorists have claimed that the geometric module is dedicated to the reorientation task, and geometry has certainly been demonstrated to be important for reorientation. However, the modularity position must also demonstrate that the geometry findings cannot be explained by a more general cognitive skill, as this would imply that the system is not specifically dedicated for reorientation.

### *7.1. Not all kinds of geometry are used early in development*

Traditionally, the first type of geometric information to be studied was relative length—alternating short and long walls. However, there are conditions under which relative length can be difficult to use. In a recent study, 4-year-old children were asked to reorient with variations of the height and continuity of geometric information (Lee & Spelke, 2008). In one condition, children were asked to reorient with a rectangular array of four large, stable columns. Children did not use the geometric information that was suggested by the rectangular shape. The experimenters next outlined a rectangle on the floor with tape. Thus, the shape was clear and uniform, but did not have elevation. Again, for this condition, the children did not use the geometric information. Finally, when elevation was added, with either 12-inch or 35-inch-tall walls, children were able to use geometric information for reorientation, with no difference in accuracy depending on wall height. It may be surprising that the taller walls did not increase geometric responding, as the taller walls ought to be

more salient than the low walls. However, what might be important is the presence of any elevation at all. Elevation is required to activate boundary vector cell firing (Solstad, Boccara, Kropff, Moser, & Moser, 2008), and this may have a role to play in the reorientation task. The role of elevation is still ambiguous, however, as is whether geometry can be imputed from separated landmarks (an ambiguity noted by Cheng & Newcombe, 2005). Lew, Gibbons, Murphy, and Bremmer (2009) found that 2-year-old children could reorient using the geometry of the search space for both enclosed spaces and using the imputed geometry from an array of landmarks. In this experiment, the reorientation performance of toddlers was not better for the enclosed spaces compared to geometric conditions that were defined by uniform landmark arrays.

The Lew et al. (2009) experiment further provides evidence against the geometric module hypothesis, by showing how the phenomena may be limited to situations unlikely to occur in natural ecology. Toddlers were asked to reorient in regular rectangular or isosceles triangle conditions and, as would be expected, performed above chance in these conditions. Importantly, however, toddlers were next examined in *irregular* quadrilateral and *irregular* triangular environments. The irregular environments are interesting because they contain unique geometric information, including unique corner angles and relative wall length differences, but disrupt the symmetry of the space. Toddlers' choices fell to chance levels in the irregular conditions. Thus, even when corner angle and unique lengths are present, there are some conditions when toddlers fail to reorient using geometric information. This finding is problematic for the geometric module hypothesis, because the geometry that is found in the natural environment is much more likely to resemble the irregular configuration (where toddlers fail to use geometry) than the symmetric search spaces that have been traditionally used to study this process (Lew et al., 2009). There are clearly variations in the likelihood of use of various kinds of relative length cues for reorientation. Because the natural environment does not contain unambiguous enclosures whose geometry is defined by continuous elevations, the generality of the geometric module approach seems uncertain.

Geometry also includes more than simply the length and relative positions of lines and extended surfaces—angles are geometric. The original studies of the geometric module mainly used rectangles in which all walls met at 90 degree angles. Hupbach and Nadel (2005) asked whether children could use the angular information in a rhombus to recover from disorientation. As shown in Fig. 3, a rhombus has four sides of equal length, with two


Fig. 3. The rhombus enclosure. Each of the sides of the enclosure is equal in length. Therefore, the only unique geometric information available in these enclosures is that of corner angle. There are two acute (A) and two obtuse (O) angles in each corner. When there is no additional information, the best that the disoriented child can do is to divide search between the two corners with identical angles, as in the first rhombus. However, when a feature wall is added, as in the second rhombus, then it is possible for participants to use the feature wall to disambiguate the two equal angle corners ($O_F$ from O or $A_F$ from A). From Hupbach and Nadel's (2005) study, children start using the angle information as well as the feature information at 4 years of age.

equal obtuse angles and two equal acute angles. Reorientation analogous to that achieved in rectangular rooms would involve children concentrating search on the two angles that correspond to the corner in which they saw something hidden. However, although children as young as 18 months use wall-length information successfully (Hermer & Spelke, 1996), children did not succeed in using angular information until the age of 4 years. And, by the time they were using this kind of geometric information, they were also using a feature to choose successfully between the two corners. It is hard to see how an ability can be characterized as ''geometric'' if it does not include information about angle. Next we turn to the claim of specificity.

### *7.2. Use of scalar and nonscalar cues by toddlers*

Recently, Huttenlocher and Lourenco (2007) have questioned the geometric module on the grounds of specificity. It has been demonstrated that geometry can be used for reorientation when it is available. However, if there is a dedicated geometry system that is dedicated only for reorientation, then it must also be demonstrated that reorientation fails in the absence of geometric information. As we review below, toddlers can succeed in square environments that lack unique geometric information. As an alternative, Huttenlocher and Lourenco (2007) proposed that the reorientation behavior can be explained by a more general ability to discriminate and compare scalar (or relative) cues. The lengths of walls define a continuum of size, a scalar comparison. By contrast, colored versus white walls define contrasting categories and are nonscalar. It might be that scalar cues are easier to use for reorientation than nonscalar ones. This contrast would be more general than (and different from) the contrast between geometric information and features. To test this idea, Huttenlocher and Lourenco (2007) tested 18- to 24-month-old children in square enclosures. The toddlers were shown the hiding location of a toy and then were disoriented before being allowed to search for the toy. Since the square provides no unique geometric information, the modularity position predicts a failure of reorientation. However, with small and large polka dot patterns, as shown in Fig. 4, the toddlers were able to successfully pick the target

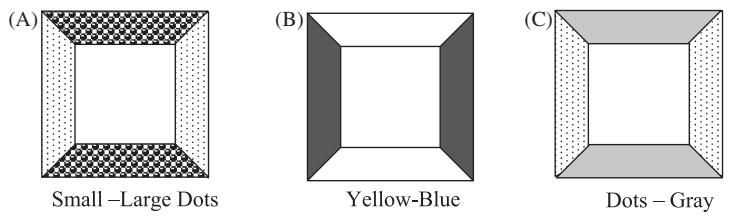


Fig. 4. Enclosures to test the use of scalar (A) and nonscalar cues (B,C) in the absence of geometric information. As we can see in this figure, the diagonal corners are identical to each other. For example, in panel (A), the small dots are to the left and the large dots to the right in both the top left and the bottom right corners. Thus, choices to either of these corners are scored as correct, and hence performance is compared to 50% chance. In panel (B), the walls appear to be black and white. In the actual experiment, the walls were either blue and red for children, or blue and yellow for mice (as mice cannot perceive the color red).

corner about 70% of the time, significantly above chance levels of 50%. When the adjacent walls were defined by nonscalar cues, either alternating blue and red walls or walls with dots alternating with gray walls, the toddlers were unable to reorient at above 50% chance levels. As there are four potential hiding locations, one might expect chance to be 25%. However, there are two equivalent corners that preserve the relationship between dot size and left-right position (e.g., large dots on the left). Thus, the two equivalent corners are scored as correct and search performance is compared to 50% chance.

There are two possible explanations for why toddlers were able to reorient in the scalar cue condition and were unsuccessful in the nonscalar cue conditions. Ordinal relationships may be more readily mapped onto spatial position than the other two cues. Alternatively, the small-large comparison has multiple co-occurring cues such as size, spatial density, and the number of dots per wall while the other conditions have only one cue. Lourenco, Addy, and Huttenlocher (2009) explored the latter possibility by comparing reorientation with a single dimension to that with a compounded dimension. When adjacent walls differed in luminance, reorientation was slightly but reliably above chance (59%). When adjacent walls differed in a single way, namely solely orientation of lines, reorientation was somewhat lower and not significantly above chance (55%). When the two were combined, reorientation was more successful than with either cue alone, and toddlers were able to find the toy on the first search about 70% of the time. However, although the luminance effect is small, the importance of scalar cues is underlined by findings from an additional composite-cue condition. While composite cues help, at least one of the cues must be relative in order to get a boost in performance. When two categorical cues were combined (red Xs and blue Os), reorientation was still at chance.

Categorical cues may not be absolutely impossible to use, but just more difficult. Working with 18- to 24-month-old children in square enclosures, Nardini, Atkinson, and Burgess (2008) found that children were able to reorient using a nonrelative cue, alternating blue and white walls. However, performance was far from perfect, at 61% accuracy. Toddlers may have performed slightly better in this experiment than in Huttenlocher and Lourenco's for several reasons. First, the size of the enclosure was nearly three times larger. From Learmonth et al.'s (2001, 2002) work, we know that features are easier for children, and adults and nonhuman species for that matter, to use in larger spaces. Additionally, toddlers participated in up to eight trials (compared to four) and we know that practice augments feature use (Learmonth et al., 2008; Twyman et al., 2007). In fact, if only the first four trials are analyzed for Nardini et al.'s experiment, then reorientation is only marginally significant with a two-tailed test. Interestingly, the distinction between right and left sense appears to already be developing in the 18–24-month age range. In one of the conditions of Nardini et al.'s experiment, the opposing walls were covered with animals. In the symmetric condition, the animals were arranged in a mirror image fashion so that the toddler would have to combine the feature information with the left-right distinction in order to focus search on the correct corner. For example, the target corner may have been adjacent to a lion. However, there would be a lion in each of the four corners, so the toddler was required to remember that it was the lion on the left that marked the goal location. In the asymmetric condition, the animals were identical at the two diagonal corners, but different on the

opposite diagonal. So the target corner may be adjacent to the lion on the left, but now there is a flamingo on the right so the toddler no longer needs to make the left-right distinction for reorientation. It is interesting to note that the toddlers searched above chance when the left-right distinction was required for successful performance, but removing the sense requirement boosted the toddlers' performance to 73%, close to the performance seen with small and large dots, when the left-right judgment was not required of children in the reorientation task.

### *7.3. Use of scalar and nonscalar cues by mice*

Mice as well as human children show a greater ability to use scalar (as opposed to nonscalar) information to reorient, showing that the advantage of scalar information is not due to symbolic or linguistic ability. In addition, the data from mice make clear that nonscalar information *can* be used to reorient, albeit with somewhat greater difficulty than scalar information, as was hinted at in the data from human children. These conclusions come from a study based on Huttenlocher and Lourenco's (2007) study, in which C57/BL/6 mice were trained to find a food reward in the same conditions (Twyman, Newcombe, & Gould, 2009). Additionally, the extra trials that can be collected from mice compared to toddlers allowed a closer look at the ability to use nonscalar information for reorientation. The mice were able to reorient using scalar information (12 trials) much faster than when offered either the nonscalar color (38 trials) or dots-gray (33 trials) comparisons. This difference in acquisition time for the different types of feature is difficult for the modularity position to explain as there is no difference in the geometric information available across groups.[3] Furthermore, this experiment reconciles the seeming discrepancy between Huttenlocher and Lourenco (2007) and Nardini et al. (2008) by demonstrating that nonscalar cues can in fact be used for reorientation, provided there is sufficient time for learning.

The common thread across all of these experiments with toddlers in square enclosures is that in some cases they are able to successfully reorient in the absence of geometric information, contrary to the predictions of a geometric module. Scalar information is preferred. The finding that nonscalar information can be used in a square enclosure, but is harder to use than scalar information, is confirmed in studies with mice. Local view theory, which entails matching a snapshot of the to-be-remembered location with the current view, would be able to explain Nardini et al.'s (2008) results. However, proponents of modularity theory are not likely to subscribe to local view theory, in the manner it is being discussed in the literature, as this position is explicitly nonmodular (Cheung, Sturzl, Zeil, & Cheng, 2008; Sturzl, Cheung, Cheng, & Zeil, 2008).

### *7.4. How does the vertical axis fit in?*

In just about all of the experiments reviewed, reorientation has been examined on a horizontal surface. For people, the majority of our navigation occurs on the horizontal plane, although there are notable exceptions when traveling up or downhill. However, other species, such as marine species or animals that can fly, may spend much more time navigating

in both the horizontal and vertical plane. This vertical aspect of navigation has recently been examined with the reorientation paradigm (Nardi & Bingman, 2009). In this task, pigeons were asked to reorient in an isosceles trapezoid arena. When the ground was flat, pigeons were able to learn the task. When the ground was sloped, pigeons were faster to learn the task and more accurate. How should slope be classified in terms of a reorientation cue? The two main classes of cues that have been studied with the reorientation task are geometric and feature cues. However, slope does not appear to fit easily into the classification of a geometric or a feature cue. Thus, information along the vertical axis, namely slope, appears to be a salient reorientation cue that may warrant its own category of cue type.

### *7.5. Summary*

The data reviewed in this section make the point that the dichotomy between use of geometry to reorient (obligatory and early) versus use of features (variable, late, and dependent on language) is overly stark. First, relative length is only used with continuous enclosures defined by raised barriers, which poses problems for how useful the geometric module would be in natural ecology. Second, one kind of geometry, angular information, is not used until fairly late in development. Once it is used, features are used as well. Third, features can be used to reorient in square rooms, especially with scalar information but likely with nonscalar information as well. Fourth, a salient cue in the natural world—slope—is important in reorientation and seems to constitute an additional class of information that does not fit neatly into the geometry versus feature dichotomy. Overall, in reason 5, we question what is meant by geometric information.

## 8. Summary of the five reasons to doubt the geometric module

Here, we have outlined five reasons and evidence to doubt the existence of a dedicated geometric module for reorientation. There were many reasons to doubt that language played a unique role in the integration of feature and geometric information, based on evidence that nonhuman animals flexibly used geometric and feature cues for reorientation, and that human's reorientation ability was not dependent solely on language. Next, variable phenomena were reviewed and it was demonstrated that modularity theory and core knowledge positions do not have enough flexibility in their theoretical accounts of reorientation to be able to explain these variable phenomena. For example, geometric and feature cue use depended on environment size, and the presence or absence of overshadowing, blocking, and potentiation effects were not explained by modularity theory. Additionally, both modularity and core knowledge postulate that the reorientation system is innate, and thus experience should not influence the behavior of reorienting organisms. We demonstrated that experience, through short-term training experiments and over the long-term with rearing experiments, had an influence on the orientation performance of participants. Next, a recent two-step modular model of reorientation was outlined. Advocates of the core knowledge position argue that geometric information is first used, and solely used, for reorientation. Subsequently, features

are used associatively to pinpoint a goal location, but crucially features cannot be used for reorientation. In contrast, we reviewed evidence that features can be used for true reorientation, both in the presence and absence of geometric information. Finally, we discussed what types of geometric information can be used across development, and across species for reorientation. It has become apparent, that not all types of geometry are used for reorientation, and a more specific definition of geometric information may be required. For all of these reasoned, discussed more thoroughly above, there are many limitations to modularity and core knowledge theory. In the next section, we will review more recent theories that attempt to explain the reorientation phenomena, as summarized in Table 3.

## 9. Alternatives to modularity

The first part of this paper offered five reasons to doubt the existence of a geometric module. In the next section, we change our focus to review recent alternative theoretical models. We evaluate each of them using the score card of whether it could account for the phenomena shown in Table 3, which are quite well established and in Table 4, phenomena that are less clear and need further exploration. We emphasize the criterion of whether the model

Table 3
Phenomena to be explained by any model of reorientation

| Phenomena |
|---|
| 1. Reorientation using relative length is easier than reorientation using angle size. |
| 2. Reorientation relies more on features and less on geometry as enclosure sizes become larger. |
| 3. Features are more likely to be used as children get older, but the improvement is continuous in larger rooms whereas, in smaller rooms, features are not used spontaneously until 6 years of age. |
| 4. Feature use is enhanced by language training. |
| 5. Feature use is enhanced by prior experience with features in a variety of situations. |
| 6. Feature use is attenuated by both language interference and spatial interference. |
| 7. Scalar information is easier to use for reorientation than nonscalar information. |
| 8. Overshadowing and blocking are sometimes but not always observed with featural and geometric information—and potentiation is even possible. |
| 9. Distal feature cues are used at a younger age than proximal feature cues. |
| 10. Movement enhances the integration of feature and geometric cues. |

Table 4
Phenomena that require clarification and further experimentation

| Unclear Phenomena |
|---|
| 1. Is geometry more likely to predominate over features in a working memory task as opposed to a reference memory task? |
| 2. Is geometry harder to use when it must be imputed from separated points rather than being instantiated by continuous surfaces? |
| 3. Why are direct features sometimes (but not always) easier to use than indirect features in search after disorientation? |

could account for development in general, as well as for the specific developmental facts, such as the use of length before the use of angle or the use of features in large but not small spaces from an early age.

### *9.1. Adaptive combination*

Adaptive combination was introduced by Newcombe and Huttenlocher (2006) as a theory whereby multiple sources of spatial information are integrated into a nonmodular and unified representation. It is one version of a number of models of spatial functioning that postulate the weighted integration of a variety of relevant sources of information to support spatial functioning (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007). Information that is high in salience, reliability, familiarity, and certainty, and low in variability, is given priority over other sources of information. Unlike modularity, the adaptive combination model suggests that this information is continually being modified by the creatures' experience. Cues that lead to adaptive behavior are elevated in the probability of their use, by increasing the relative weights of that information source, and cues that led to maladaptive behavior are decreased in weight. In terms of development, interaction with and feedback from the environment allows the evolution of relative weights for the potential sources of information that are increasingly well adapted.

Adaptive combination offers an explanation of many of the facts outlined in Table 3. Points 2, 5, 6, and 7 were predicted by adaptive combination theory and then empirically tested. Point 2 was predicted by adaptive combination theory since as the size of the enclosure increases, there are several reasons to expect that the use of features would increase: First, the size of the feature wall is larger in the larger spaces, and therefore is more salient; second, movements around the feature wall create less variability in the large space than the smaller enclosures. Points 5 and 6 are related to each other as both involve either enhancing feature use through training and experience or decreasing feature use through interference tasks. Adaptive combination predicts that practice either directly with the features (or through language training—Point 4) will increase feature use as this experience increases the cue weightings of features relative to geometry. Because there is no reason to think that features are only encoded verbally, adaptive combination predicts that interference tasks with either verbal or spatial encoding will be detrimental to performance, although not fatal as there are back-up systems of encoding that will still enable partially successful performance. Finally, in terms of point 7, scalar information is potentially easier to use than nonscalar information for two reasons. The less interesting explanation is that scalar information, as it was studied, contained more potentially useful cues (size, number, density) than the nonscalar comparison. The more interesting possibility is that scalar cues may be more readily mapped onto spatial position than nonscalar cues.

Point 1 can also be explained by adaptive combination theory, although arguably in a post hoc fashion. Relative length may be easier for children to use than corner angle because of differences in memory demands between the two types of cues. When facing a corner, especially in small rooms, the small and large walls intersect at the corner right in front of the child. This facilitates a comparison of relative length. For corner angle in contrast, the child

who is facing the correct corner must rotate to compare the current corner angle to the other corners where memory demands may increase the difficulty of using corner angle over wall length as a reorientation cue.

Adaptive combination is an overarching theoretical perspective that argues for an active and adaptive use of relevant information to support spatial reorientation. Recently, other research teams have proposed more specific models. These models can be seen as compatible with the general principles of adaptive combination theory, and they have tested some of the specific reorientation phenomena listed in Table 3. We turn now to those models.

### *9.2. Operant model using Rescorla-Wagner principles*

Miller and Shettleworth (2007) proposed a model of reorientation based on Rescorla and Wagner's (1972) principles of association. However, the Rescorla-Wagner model applies to classical conditioning, so Miller and Shettleworth revised the model to apply to an operant situation because, after disorientation, the animal chooses the corner to approach and therefore selects for itself the stimuli experienced during the experiment. Thus, the model includes a measure of the probability of encountering each corner, based on the associative strengths of all of the cues at each corner.

This model had as its central goal to provide a unified account of the blocking, overshadowing, and potentiation effects discussed earlier in the article. One of the key concepts of the model is feature enhancement. If the organism learns the geometry of the environment, then this cue leads to reward half of the time. However, a key assumption of the model is that the creature quickly learns to select the correct corner on the basis of feature information. Since the correct geometric information is paired with the reward quite frequently, the feature is aiding the creature to learn about the geometry of the space. Then, because the organism has overvalued the contingency of the geometric cue to greater than 50%, the organism starts making rotational errors and eventually the organism learns the actual contingency of reward for geometric information.

As for the more general adaptive combination theory, this approach is a nonmodular model. The transitory nature of feature enhancement can explain the mixed results with overshadowing and blocking effects depending on what point in training the test trials were administered. This operant model is able to explain many of the reorientation findings, including the original Cheng (1986) experiments and the blocking, overshadowing, and facilitation effects of Pearce and colleagues. Additionally, a recent revision of Miller and Shettleworth's (2007) model is able to account for the room size effects for children, fish, chicks, and pigeons.

Recently, there has been a critique of this model (Dawson, Kelly, Spetch, & Dupuis, 2008). Although the authors agree with the fundamental premise that reorientation is an operant learning task, they point out that Miller and Shettleworth's formula for probabilities is based on associative strengths, which can be positive or negative. Thus, the mathematical equation sometimes produces impossible probabilities of less than 0% or greater than 100%. As a solution, the authors provide an alternative engine, a perceptron, to drive the mathematical side of the model while keeping the operant vision for reorientation. A perceptron is

an artificial neural network that has inputs that encode stimuli, outputs that respond to stimuli, and flexible weighted connections between the inputs and outputs. Thus, the perceptron is still based upon associative weights, but the probabilities remain between 0% and 100% and the operant nature of the task is preserved. The Dawson et al. article arguably represents a friendly amendment to the Miller and Shettleworth approach. Miller and Shettleworth (2008) took the opportunity to reply to Dawson et al. They agreed that there was a flaw in one of the calculations and have modified the equation to eliminate aberrant probabilities. This modification has again demonstrated ''how what appeared to be exceptional kind of cue interactions in geometry learning experiments can arise from an unexceptional competition for learning among geometric and other cues'' (p. 422).

How then does the model fare in explaining the phenomena listed in Table 3? One issue is that, although the Miller (2009) model tackled age effects, it does so simply by adjusting parameters in the model to create age differences. There is no independent motivation for why such parameters might be age graded. The model also has yet to address phenomenon 1 (angle size is harder to use than wall length), some aspects of phenomenon 5 (training and malleability), phenomenon 6 (interference effects), or phenomenon 7 (scalar information is easier to use than nonscalar). Phenomena 1 and 7 could be tackled by adjusting parameters, but as with the treatment of age, it could be argued that such adjustments are ad hoc. Some aspects of Phenomenon 5 are very naturally explained by the model, which after all is an operant model, but it is not clear that the model could cover the more abstract generalization of the training studies of Twyman et al. (2007). However, this model is an excellent start, and it will be interesting to see the model refined to be able to explain more of the phenomena listed in Tables 3 and 4.

### *9.3. Local view theory*

Because of accumulating evidence against modularity, in particular the findings of Pearce and colleagues, Cheng (2008) is now quite skeptical that there is a geometric module. In its place, Cheng predicts that either a version of Miller and Shettleworth's (2007) operant model or local view theory will take its place. Local view theory has grown out of the research on insect navigation, which is largely accomplished by matching a stored retinal image to the current image. When applied to the reorientation task, local view theory postulates that rotational errors arise out of the image-matching process (Cheung et al., 2008; Sturzl et al., 2008). To explain the original Cheng (1986) finding that rotational errors are made by rats in rectangular enclosures, this nonmodular account suggests that the organism stores an image of the target location. Once disoriented and released, the organism looks around and then moves in the direction that minimizes the discrepancy between the stored and the current image. The process is repeated until the organism arrives at the end point. It is a nonmodular account because geometric cues are not given a privileged status. Local view theory circumvents the issue of what should be counted as a feature and what should be considered geometry by assigning equal status to all possible cues. From this perspective, geometric and feature information are stored together in the target corner image.

How do rotational errors arise if all of the pertinent information is available to home in on the correct corner? The authors propose that the target image is segmented into information at the edges of the enclosure and internal information of the walls. Furthermore, the saliencies of the edges and internal information may be equal, or one may be stronger than the other. The model predicts that when the edges are more salient, then the agent will make more rotational errors. However, if the internal information is enhanced, then the agent will go towards the correct corner.

To test this theory, the authors simulated the reorientation paradigm using virtual reality simulations and a robot that stored the target image and then moved along the image difference function (minimizing the difference between the current and stored panoramic image) to determine view whether based matching would result in rotational errors. When the rectangular arena was all black, either with or without feature panels, the robot ended up in either the correct or the rotationally equivalent corner. A similar result was found with three black walls and one white feature wall. Furthermore, the authors were able to demonstrate that increasing the salience of the internal information reduced rotational errors and vice versa. Thus, the authors demonstrated that the rotational errors from Cheng's original experiment do not require a modular account of reorientation.

After demonstrating that rotational errors can arise in view-based matching in rectangular spaces, the authors tackle kite-shaped spaces (Graham et al., 2006; Pearce, Graham, Good, Jones, & McGregor, 2004). They selected these experiments as they deemed them particularly problematic for modularity theory. In one of these experiments, the rats were searching for the platform in a kite-shaped enclosure. When the wall color changed from trial to trial, it should have been very easy for the rats to depend on geometric information. In contrast, this was a very hard task for the rats to learn; in fact, a facilitation effect was found since making the feature wall stable enhanced learning of the geometric information. When the robot was put to the test, local view modeling performed well, with a few minor exceptions.

The hypothesis of an image-matching mechanism that produces rotational errors is gaining momentum. In a similar vein to the robot modeling just discussed, a different group of authors has modeled some of the reorientation data with computational models of rats (Sheynikhovich, Chavarriaga, Strösslin, Arleo, & Gerstner, 2009). A computational neural model was developed that had both an egocentric stimulus-response strategy as well as an allocentric place-based navigation strategy. One interesting point of this model was that the allocentric strategy arose out of the combined input of visual snapshots, path integration, place cells, and grid cells. When this simulated model was tested in different conditions, the authors also found that rotationally equivalent errors arose in the allocentric conditions, even though the underlying representation was nonmodular. Thus, for both the robotic as well as the computational neural model, it is at least possible, in theory, for rotational errors to arise out of a unified cognitive representation as an artifact of an image-matching process.

A recent study with children is one of the first attempts to test an image-matching approach with children (Nardini, Thomas, Knowland, Braddick, & Atkinson, 2009). In this study, children were disoriented and then were asked to retrieve a hidden toy either from a position that facilitated image matching (i.e., could be encoded in a viewpoint-dependent

manner such as left of the feature wall) or from a novel position that was viewpoint independent, and thus prohibited image matching. Four-year-old children were successful only in the viewpoint-dependent condition. At 5 years of age, children appear to be transitioning from search with a viewpoint-dependent strategy, to a viewpoint-independent strategy. By 6 years of age, children are able to reorient using a viewpoint-independent strategy. Thus, below the age of 6, there is some support that children may be using an image-matching approach to reorientation. However, by 6 years of age, children are able to successfully reorient from novel viewpoints, and thus, it appears that there is an alternative strategy that older children and adults can use for reorientation that seems difficult for local view theory to explain.

There are many attractions to local view theory, including its simplicity, specificity, and testability. As the authors mention, it has yet to be determined how the reference image is acquired in the real world, by specific animals or by people of different ages. In addition, perhaps because the image acquisition problem has not been tackled, we do not know how the model would account for age-related differences in feature use, as well as for the room size effect, training and malleability effects, or differences between scalar and nonscalar information (see points 2, 3, and 5 in Table 1). Perhaps most troubling is the fact that there is evidence that directly challenges the central premise of this approach, that is, that organisms reorient using local views without encoding the overall shape of enclosures. Huttenlocher and Vasilyeva (2003) found that toddlers formed a representation of enclosures shaped like an isosceles triangle that was abstract enough to permit recognition of a corner as the ''same'' despite large variations in the triangle's appearance that resulted because the children might be either inside or outside the enclosure after disorientation. They also found that children typically went straight to the correct corner from a variety of initial facing points, without needing to survey the whole enclosure or large parts of it, as would seem to be predicted by the local-view approach.

### *9.4. Different neural substrates?*

Although not yet directly relevant to reorientation, a recent set of studies by Doeller, Burgess, and colleagues provide an elegant approach to studying spatial cognition. They have examined the behavioral and brain bases of landmark and boundary information in goal location tasks when adult males are semioriented (Doeller & Burgess, 2008; Doeller, King, & Burgess, 2008). In an object memory task, participants were introduced to a virtual reality environment where distal cues (mountains, clouds, and a sun) could be used as an orientation cue, but not as a distance cue to the object's location. Additionally, there was a circular boundary defined by a uniform stone wall, and a local landmark for the object, such as a traffic pylon. Participants were asked to navigate through a sequence of objects while remembering where they found them. Some of the objects were stable relative to the landmark (the pylon) and others stable relative to the boundary (the stone wall). Then participants were serially shown the objects and asked to place each where it should go. They received feedback during the training part of the experiment. To prevent egocentric responding, after an item was found, the screen went blank and participants reappeared at a

new location along the boundary, facing inwards. Thus, participants are prevented from following the same set of body responses to replace the target. However, it is unclear whether participants were fully disoriented in the same manner as the reorientation paradigm.

Behaviorally, the data suggested that the landmark and boundary information were learned in parallel. However, the principles of learning appeared to be fundamentally different for each type of information. The landmark learning followed associative learning principles, as demonstrated by blocking, overshadowing, and learned irrelevance learning. Boundary learning did not show any of these effects and was proposed to be learned incidentally.

To follow up on the behavioral data, an fMRI study was conducted to determine whether there were differences in how the brain processed boundary and landmark cues. The boundary cues activated the right posterior hippocampus while landmark cues activated the right dorsal striatum. The study also suggested how the brain combines these types of information. After parallel processing in independent systems, if only one or the other region predicts behavior, there is no additional activation. However, if the sources of information are in conflict or are both required for adaptive behavior, then the ventromedial prefrontal cortex mediates the combination of information from each system. Thus, in a place-finding task, there appears to be different neural instantiations of each cue type. This type of experimental approach, when applied to the reorientation paradigm, could be a fruitful line of inquiry.

In a complementary study, Bullens et al. (2009) adapted the object memory task for use in a 3-D space to examine the developmental trajectory of landmark and boundary information. In this version, the children entered an enclosed circular search space. The proximal landmark cue was a large traffic cone. Beyond the wall of the enclosure, distal orientation cues were displayed. Between trials, children were disoriented in a similar fashion to the standard orientation paradigm. Perhaps not surprisingly, adults were more accurate locating the target location (84%) than the 5- and 7-year-old children (26%). However, both groups searched in the correct location significantly more often than chance. There were also qualitative differences between the types of searches of adults and children. Adults were more dependent on boundary cues and were more accurate using angular estimates. In contrast, children evenly used boundary and proximal feature cues (albeit more weakly than adults) and were more accurate using distance estimates. No age-related differences in either accuracy or type of search were found. Thus, as the authors point out, it will be interesting to compare both younger and older children on this task to be able to understand which brain systems are developing when, and how cues are reweighted and integrated later in development.

This approach combines behavioral, neural, and developmental data to create a potentially elegant and fruitful model when applied to the traditional reorientation paradigm—which the paradigm used in the research was not. Thus, it not only remains to be seen whether it can explain all the phenomena of Table 3, including development but also whether it can explain the basic phenomena discovered in the original Cheng (1986) research. If it is successful, it may have the interesting implication that it will allow us to have our module and discard it too. That is, the processing of featural and geometric infor-

mation could, possibly, occur initially in two distinct brain areas, and yet, as needed, be combined and weighted in yet another. The way that the reorientation paradigm is conducted, it seems that there are two components to the task, first reorienting, and then navigating to the goal location. There has been quite a bit of research on the second component, navigating to a goal location, which may depend in part on place learning and response learning, which may be supported by the place cells of the hippocampus and the striatum, respectively. An additional layer is added to the reorientation paradigm, namely regaining a sense of direction, which is less well studied. As a start, it seems that the head direction cells, first proposed by Taube (1998) that are found in the Papez's circuit, including the postsubiculum, anterior thalamus, and retrosplenial cortex, would be important components of such a system. It will be interesting to explore the relative contributions of place cells, grid cells, head direction cells, and border cells to the reorientation task.

## 10. Conclusion

Massive modularity is a popular way to conceptualize human cognitive functioning, and it is attractively simple to explain development by postulating that modules are innately specified. Core knowledge positions share some of the same properties of modularity theory—namely innate endowment, areas of specialization, and characteristic limitations of each system. In contrast to massive modularity, the core knowledge position advocates a small number of modules, on the order of four or five areas of core knowledge. Many invocations of the term *modularity* are so vague as to be essentially untestable. A welcome exception has been the geometric module, which has been precisely defined and operationalized. We believe that the idea, once tested, has been found to be wanting, and that the majority of the empirical evidence is difficult to explain when one postulates modules, either from the massive modularity perspective, or from the core knowledge theory. Recently, the first proponent of the geometric module, Cheng (2008) has reviewed some of the evidence that makes the geometric module hypothesis quite unlikely to be true. What alternative model will take its place is not yet completely clear, but likely it will be an adaptive integrated model similar in spirit to ways of thinking about development suggested by connectionism, dynamic systems theory, and Siegler's (1996) model. Whether these doubts about one example of an innatist-modular account of development will extend to other hypothesized modules, such as theory of mind, cannot of course be stated from the present data. The example of the geometric module does, however, give us reason to be cautious about facile acceptance of the hypotheses of massive modularity or core knowledge and innate specification.

## Notes

1. Additionally, the early work with rats was conducted with all male subjects. Rats have been shown to exhibit stable sex differences in spatial learning, unlike mice (Jonasson, 2005). For example, in an experiment using a radial arm maze, a change in the

geometry of the room dropped the performance of control males and females who had been treated with estradiol benzoate. In contrast, the geometry change did not affect the performance of control females or males who had been neonatally castrated (Williams, Barnett, & Meck, 1990). Thus, findings of dominance of geometric information may depend on the sex of the animals. If this is correct, note that there is no easy way for the modularity position to explain how participant sex would affect reliance on geometric information.

2. However, there are a few things to note about Wall et al.'s study. First, accuracy with the feature panel in the first step of learning in the square was not impressively high. The rats were only 67% accurate. Thus, when they were transferred to the rectangular arena, additional learning was likely to be ongoing and could have included the geometric information. Unfortunately, the study is lacking a geometry-only control group. It would be helpful to compare the learning curve when geometry is presented on its own to one group of rats to the learning rate of geometry in rats that had previously learned that the black feature panel predicted the food location.
3. A reviewer pointed out that the small-large dot comparison may create an illusion of depth. It may appear to participants that the square is really a rectangle, and that perceived geometry that could be used for reorientation. If there is an illusion of size, it is particularly unlikely because of the small size of the search space. Illusions based on depth cues are particularly week in small spaces. Additionally, as multiple trials are administered with the children moving through the space, they would have an opportunity to interact with the space and identify the enclosure as a true square (S. F. Lourenco, personal communication, June 29, 2009).

## Acknowledgments

Preparation of this paper was supported by NSF BSC 0414302 and SBE 0541957. An earlier version of the paper was presented in a symposium at the Psychonomic Society in November 2007. Thanks to Vladimir Sloutsky for organizing the symposium and also assembling papers for a special issue, to the reviewers for their probing comments, and to the personnel and parents of the Temple Infant Lab for participation in experiments.

## References

Acredolo, L. P. (1978). Development of spatial orientation in infancy. *Developmental Psychology*, *14*, 224–234.

Acredolo, L. P., & Evans, D. (1980). Developmental change in the effects of landmarks on infant spatial behavior. *Developmental Psychology*, *16*, 312–318.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review*, *113*, 628–647.

Batty, E. R., Bloomfield, L. L., Spetch, M. L., & Sturdy, C. B. (2009). Comparing black-capped (Poecile atricapillus) and mountain chickadees (Poecile gambeli): Use of geometric and featural information in a spatial orientation task. *Animal Cognition*, *12*, 633–641.

Begley, S. (2009). Why do we rape, kill, and sleep around? *Newsweek*. Available at: http://www.newsweek.com/id/202789. Accessed June 26, 2009.

Benhamou, S., & Poucet, P. (1998). Landmark use by navigating rats (*Rattus norvegicus*): Contrasting geometric and featural information. *Journal of Comparative Pscyhology*, *112*, 317–322.

Biegler, R., & Morris, R. G. M. (1999). Blocking in the spatial domain with arrays of discrete landmarks. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 334–351.

Brooks, D. (2009). Human nature today. *New York Times*. Available at: http://www.nytimes.com/2009/06/26/opinion/26brooks.html. Accessed June 26, 2009.

Brown, A. A., Spetch, M. L., & Hurd, P. L. (2007). Growing in circles: Rearing environment alters spatial navigation in fish. *Psychological Science*, *18*, 569–573.

Bullens, J., Nardini, M., Doeller, C. F., Braddick, O., Postma, A., & Burgess, N. (2009). The role of landmarks and boundaries in the development of spatial memory. *Developmental Science*, doi: 10.1111/j.1467-7687.2009.00870.x.

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought*. New York: Clarendon Press/Oxford University Press.

Cheng, K. (1986). A purely geometric module in the rats spatial representation. *Cognition*, *23*, 149–178.

Cheng, K. (2008). Wither geometry? Troubles of the geometric module. *Trends in Cognitive Sciences*, *12*, 355–361.

Cheng, K., & Gallistel, C. R. (2005). Shape parameters explain data from spatial transformations: Comment on Peace et al. (2004) and Tommasi & Polli (2004). *Journal of Experimental Psychology: Animal Behavioral Processes*, *31*, 254–259.

Cheng, K., & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review*, *12*, 1–23.

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin*, *133*, 625–637.

Cheung, A., Sturzl, W., Zeil, J., & Cheng, K. (2008). The information content of panoramic image II: View based navigation in nonrectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavioral Processes*, *34*, 15–30.

Chiandetti, C., Regolin, L., Sovrano, V. A., & Vallortigara, G. (2007). Spatial reorientation: The effects of space size on the encoding of landmark and geometry information. *Animal Cognition*, *10*, 159–168.

Chiandetti, C., & Vallortigara, G. (2008). An innate geometric module? Effects of experience with angular geometric cues on spatial reorientation based on the shape of the environment. *Animal Cognition*, *11*, 139–146.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York: Oxford University Press.

Cressant, A., Muller, R. U., & Poucet, B. (1997). Failure of centrally placed objects to control the firing fields of hippocampal place cell. *Journal of Neuroscience*, *17*, 2531–2542.

Dawson, M. R. W., Kelly, D. M., Spetch, M. L., & Dupuis, B. (2008). Learning about environmental geometry: A flaw in Miller and Shettleworth's (2007) operant model. *Journal of Experimental Psychology: Animal Behavioral Processes*, *34*, 415–418.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding visual properties. *Psychological Science*, *19*, 189–195.

Diez-Chamizo, V., Sterio, D., & Mackintosh, N. J. (1985). Blocking and overshadowing between intramaze and extramaze cues: A test of the independence of locale and guidance leraning. *Quarterly Journal of Experimental Psycholgoy*, *37B*, 235–253.

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences*, *105*, 5909–5914.

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences*, *105*, 5915–5920.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (2000). Replies to critics. *Mind & Language*, *15*, 350–374.

Foster, T., Castro, C., & McNaughton, B. (1989). Spatial selectivity of rat hippocampal neurons: Dependence on preparedness for movement. *Science*, *244*, 1580–1582.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gouteux, S., Thinus-Blanc, C., & Vauclair, J. (2001). Rhesus monkeys use geometric and nongeometric information during a reorientation task. *Experimental Psychology General*, *130*, 505–519.

Graham, M., Good, M. A., McGregor, A., & Pearce, J. M. (2006). Spatial learning based on the shape of the environment is influenced by properties of the objects forming the shape. *Journal of Experimental Psychology: Animal Behavioral Processes*, *32*, 44–59.

Gray, E. R., Bloomfield, L. L., Ferrey, A., Spetch, M. L., & Sturdy, C. B. (2005). Spatial encoding in mountain chickadees: Features overshadow geometry. *Biology Letters*, *1*, 314–317.

Hamilton, D. A., & Sutherland, R. J. (1999). Blocking in human place learning: Evidence from virtual navigation. *Psychobiology*, *27*, 453–461.

Hardt, O., Hupbach, A., & Nadel, L. (2009). Factors moderating blocking in human place learning: The role of task instructions. *Learning & Behavior*, *37*, 42–59.

Hermer, L., & Spelke, E. (1994). A geometric process for spatial representation in young children. *Nature*, *370*, 57–59.

Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*, 195–232.

Hermer-Vazquez, L., Moffet, A., & Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition*, *79*, 263–299.

Hermer-Vazquez, L., Spelke, E., & Katsnelson, A. (1999). Sources of flexibility in human cognition: Dual task studies of space and language. *Cognitive Psychology*, *39*, 3–36.

Horne, M. R., & Pearce, J. M. (2009). A landmark blocks searching for a hidden platform in an environment with a distinctive shape after extended pretraining. *Learning & Behavior*, *37*, 167–178.

Hull, C. L. (1943). *Principles of behavior. An introduction to behavior theory*. New York: Appleton-Century.

Hupbach, A., Hardt, O., Nadel, L., & Bohbot, V. D. (2007). Spatial reorientation: Effects of verbal and spatial shadowing. *Spatial Cognition and Computation*, *7*, 213–226.

Hupbach, A., & Nadel, L. (2005). Reorientation in a rhombic environment: No evidence for an encapsulated geometric module. *Cognitive Development*, *20*, 279–302.

Huttenlocher, J., & Lourenco, S. F. (2007). Coding location in enclosed spaces: Is geometry the principle? *Developmental Science*, *10*, 741–746.

Huttenlocher, J., & Vasilyeva, M. (2003). How toddlers represent enclosed spaces. *Cognitive Science*, *27*, 749–766.

Jonasson, Z. (2005). Meta-analysis of sex differences in rodent models of learning and memory: A review of behavioral and biological data. *Neuroscience and Biobehavioral Reviews*, *28*, 811–825.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Kelly, D. M., Spetch, M. L., & Heth, C. D. (1998). Pigeons' (*Columba livia*) encoding of geometric and featural properties of a spatial environment. *Journal of Comparative Psychology*, *112*, 259–269.

Kinzler, K. D., & Spelke, E. S. (2007). Core systems in human cognition. *Progress in Brain Research*, *164*, 257–264.

Kjelstrup, K. B., Solstad, T., Brun, V. H., Hafting, T., Leutgeb, S., Witter, M. P., Moser, E. I., Moser, M. B. (2008). Finite scale of spatial representation in the hippocampus. *Science*, *321*, 140–143.

Landau, B., & Lakusta, L. (2009). Spatial representation across species: Geometry, language, and maps. *Current Opinion in Neurobiology*, *19*, 1–8.

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children's use of landmarks: Implications for modularity theory. *Psychological Science*, *13*, 337–341.

Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddler's use of metric information and landmarks to reorient. *Journal of Experimental Child Psychology*, *80*, 225–244.

Learmonth, A., Newcombe, N. S., Sheridan, M., & Jones, M. (2008). Why size counts: Children's spatial reorientation in large and small enclosures. *Developmental Science*, *11*, 414–426.

Lee, S. A., Shusterman, A., & Spelke, E. S. (2006). Reorientation and landmark-guided search by young children: Evidence for two systems. *Psychological Science*, *17*, 577–582.

Lee, S. A., & Spelke, E. S. (2008). Children's use of geometry for reorientation. *Developmental Science*, *11*, 743–749.

Levinson, S. C. (2003). *Space in language and cognition*. Cambridge, England: Cambridge University Press.

Lew, A. R., Foster, K. A., & Bremner, J. G. (2006). Disorientation inhibits landmark use in 12- to 18-month-old infants. *Infant Behavior & Development*, *29*, 334–341.

Lew, A. R., Gibbons, B., Murphy, C., & Bremmer, J. G. (2009). Use of geometry for spatial reorientation in children applies only to symmetric spaces. *Developmental Science*, doi: 10.1111/j.1467-7687.2009.00904.

Lourenco, S. F., Addy, D., & Huttenlocher, J. (2009). Location representation in enclosed spaces: What types of information afford young children an advantage? *Journal of Experimental Child Psychology*, *104*, 313–325.

Maes, J. H. R., Fontanari, L., & Regolin, L. (2009). Spatial reorientation in rats (*Rattus norvegicus*): Use of geometric and featural information as a function of arena size and feature location. *Behavioural Brain Research*, *201*, 285–291.

McComas, J., & Dulberg, C. (1997). Children's memory for locations visited: Importance of movement and choice. *Journal of Motor Behavior*, *29*, 223–230.

Miller, N. (2009). Modeling the effects of enclosure size on geometry learning. *Behavioural Processes*, *80*, 306–313.

Miller, N. Y., & Shettleworth, S .J. (2008). An associative model of geometry learing: A modified choice rule. *Journal of Experimental Psychology, Animal Behavior Processes*, *34*, 419–422.

Miller, N. Y., & Shettleworth, S. J. (2007). Learning about environmental geometry: An associative model. *Journal of Experimental Psychology, Animal Behavior Processes*, *33*, 191–212.

Nadel, L., & Hupbach, A. (2006). Cross-species comparisons in development: The case of the spatial ''module''. In M. H. Johnson & Y. Munakata (Eds.), *Attention and performance XXI* (pp. 499–512). Oxford, England: Oxford University Press.

Nardi, D., & Bingman, V. P. (2009). Pigeon (*Columba livia*) encoding of a goal location: The relative importance of shape geometry and slope information. *Journal of Comparative Cognition*, *123*, 204–216.

Nardini, M., Atkinson, J., & Burgess, N. (2008). Children reorient using the left/right sense of coloured landmarks at 18–24 months. *Cognition*, *106*, 519–527.

Nardini, M., Thomas, R. L., Knowland, V. C. P., Braddick, O. J., & Atkinson, J. (2009). A viewpoint-independent process for spatial reorientation. *Cognition*, *112*, 241–248.

Newcombe, N. S. (2005). Language as destiny? Or not—Essay review of space in language and cognition: Explorations in cognitive diversity by Stephen C. Levinson. *Human Development*, *48*, 309–314.

Newcombe, N. S., & Huttenlocher, J. (2006). Development of spatial cognition. In W. Damon & R. Lerner (Series Eds.) and D. Kuhn & R. Seigler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception and language*, 6th ed. (pp. 734–776). Hoboken, NJ: John Wiley & Sons.

Newcombe, N. S., & Ratliff, K. R. (2007). Explaining the development of spatial reorientation: Modularity-plus-language versus the emergence of adaptive combination. In J. Plumer & J. Spencer (Eds.), *The emerging spatial mind* (pp. 53–76). New York: Oxford University Press.

Newcombe, N. S., Ratliff, K. R., Shallcross, W. L., & Twyman, A. (2009). Young children really can reorient using features: Further evidence against a modular view of spatial processing. *Developmental Science*, *5*, 1–8.

Pearce, J. M., Graham, M., Good, M. A., Jones, P. M., & McGregor, A. (2004). Transfer of spatial behavior between different environments: Implications for theories of spatial learning and for the role of the hippocampus in spatial learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*, 135–147.

Pearce, J. M., Graham, M., Good, M. A., Jones, P. M., & McGregor, A. (2006). Potentiation, overshadowing, and blocking of spatial learning based on the shape of the environment. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 201–214.

Presson, C. C., & Montello, D. R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology*, *6*, 378–381.

Ratliff, K. R., & Newcombe, N. S. (2008a). Is language necessary for human spatial reorientation? Reconsidering evidence from dual task paradigms. *Cognitive Psychology*, *56*, 142–163.

Ratliff, K. R., & Newcombe, N. S. (2008b). Reorienting when cues conflict: Evidence for an adaptive combination view. *Psychological Science*, *19*, 1301–1307.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. G. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Roberts, A. D. L., & Pearce, J. M. (1999). Blocking in the Morris swimming pool. *Journal of Experimental Psychology—Animal Behavior Processes*, *25*, 225–235.

Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology—Animal Behavior Processes*, *23*, 110–118.

Sheynikhovich, D., Chavarriaga, R., Strösslin, T., Arleo, A., & Gerstner, W. (2009). Is there a geometric module for spatial orientation? Insights from a rodent navigation model. *Psychological Review*, *116*, 540–566.

Shusterman, A., & Spelke, E. S. (2005). Language and the development of spatial reasoning. In P. Carruthers, S. Laurene, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 89–106). New York: Oxford University Press.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Smith, A. D., Gilchrist, I. D., Cater, K., Ikram, N., Nott, K., & Hood, B. M. (2008). Reorientation in the real world: The development of landmark use and integration in a natural environment. *Cognition*, *107*, 1102–1111.

Solstad, T., Boccara, C. N., Kropff, E., Moser, M., & Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science*, *322*, 1865–1868.

Sovrano, V. A., Bisazza, A., & Vallortigara, G. (2007). How fish do geometry in large and in small spaces. *Animal Cognition*, *10*, 47–54.

Sovrano, V. A., & Vallortigara, G. (2006). Dissecting the geometric module: A sense-linkage for metric and landmark information in animals' spatial reorientation. *Psychological Science*, *17*, 616–621.

Spelke, E. S. (2008). The theory of ''core knowledge.'' *Annee Psychologique*, *108*, 721–756.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

Spetch, M. L. (1995). Overshadowing in landmark learning—touch-screen studies with pigeons and humans. *Journal of Experimental Psychology—Animal Behavior Processes*, *21*, 166–181.

Sturzl, W., Cheung, A., Cheng, K., & Zeil, J. (2008). The information content of panoramic images I: The rorational error and the similarity of view in rectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavioral Processes*, *34*, 1–14.

Taube, J. S. (1998). Head direction cells and the neurophysiological basis for a sense of direction. *Progress in Neurobiology*, *55*, 225–256.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 189–208.

Twyman, A., Friedman, A., & Spetch, M. L. (2007). Penetrating the geometric module: Catalyzing children's use of landmarks. *Developmental Psychology*, *43*, 1523–1530.

Twyman, A., Newcombe, N. S., & Gould, T. J (2009). Tale of two cities: Rearing environment influences spatial reorientation. In S. E. MacDonald (Chair), *Foraging and the evolution of cognition*. Symposium conducted at the meeting of the APA convention, Toronto, Canada.

Twyman, A., Newcombe, N. S., & Gould, T. J. (2009). Of mice (Mus musculus) and toddlers (Homo sapiens): Evidence against modularity in spatial reorientation. *Journal of Comparative Psychology*, *123*, 342–345.

Vallortigara, G., Sovrano, V. A., & Chiandetti, C. (2009). Doing Socrates experiment right: Controlled rearing studies of geometrical knowledge in animals. *Current Opinion in Neurobiology*, *19*, 20–26.

Vallortigara, G., Zanforlin, M., & Pasti, G. (1990). Geometric modules in animals' spatial representations: A test with chicks (*Gallus gallus domesticus*). *Journal of Comparative Psychology*, *104*, 248–254.

Vlasak, V. A. (2006). The relative importance of global and local landmarks in navigation by Columbian ground squirrels (*Spermophilus columbianus)*. *Journal of Comparative Psychology*, *120*, 131–138.

Wall, P. L., Botly, L. C. P., Black, C. K., & Shettleworth, S. J. (2004). The geometric module in the rat: Independence of shape and feature learning in a food finding task. *Learning & Behavior*, *32*, 289–298.

Williams, C. L., Barnett, A. M., & Meck, W. H. (1990). Organizational effects of early gonadal secretions on sexual-differentiation in spatial memory. *Behavioral Neuroscience*, *104*, 84–97.

Woodruff-Pak, D., Papka, M., & Ivry, R. (1996). Cerebellar involvement in eyeblink classical conditioning in humans. *Neuropsychology*, *10*, 443–458.

Wystrach, A., & Beugnon, G. (2009). Ants learn geometry and features. *Current Biology*, *19*, 61–66.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 34 (2010) 1357–1377

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2010.01131.x

# Domain-Creating Constraints

Robert L. Goldstone,[a] David Landy[b]

[a]*Department of Psychological and Brain Sciences, Indiana University*
[b]*Department of Psychology, University of Richmond*

Received 17 January 2010; received in revised form 29 April 2010; accepted 30 April 2010

**Abstract**

The contributions to this special issue on cognitive development collectively propose ways in which learning involves developing constraints that shape subsequent learning. A learning system must be constrained to learn efficiently, but some of these constraints are themselves learnable. To know how something will behave, a learner must know what kind of thing it is. Although this has led previous researchers to argue for domain-specific constraints that are tied to different kinds/domains, an exciting possibility is that kinds/domains themselves can be learned. General cognitive constraints, when combined with rich inputs, can establish domains, rather than these domains necessarily preexisting prior to learning. Knowledge is structured and richly differentiated, but its ''skeleton'' must not always be preestablished. Instead, the skeleton may be adapted to fit patterns of co-occurrence, task requirements, and goals. Finally, we argue that for models of development to demonstrate genuine cognitive novelty, it will be helpful for them to move beyond highly preprocessed and symbolic encodings that limit flexibility. We consider two physical models that learn to make tone discriminations. They are mechanistic models that preserve rich spatial, perceptual, dynamic, and concrete information, allowing them to form surprising new classes of hypotheses and encodings.

*Keywords:* Cognitive development; Constraints; Perception; Inference; Learning; Novelty; Embodiment; Domain-specificity

## 1. Introduction

Learning requires constraints. Gold (1967) and Chomsky (1965) formally showed that there are too many possible language grammars to learn a language in a finite amount of time, let alone 2 years, if there are no constraints on what those grammars look like. In a related analysis, Wolpert (1996) showed that there is no such thing as a truly general and

Correspondence should be sent to Robert L. Goldstone, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405. E-mail: rgoldsto@indiana.edu

efficient learning device. To be an efficient learner, one must make assumptions about the kind of structure one is expecting to find. Allegorically, if you are trying to find your favorite pair of socks and you only know that they are somewhere in your enormously large sock drawer, it will take you an enormously long time to find them. Some constraints allow you to limit your search to particular regions. Knowing that your socks are somewhere in the drawer means that you need never look anywhere outside it. Other, softer constraints simply determine an order in which the hypothesis space is searched. The physical structure of the drawer and the opacity of your socks incline you to consider the top of the drawer before the layers underneath. When the constraints coincide with reality—your favorite pair of socks really is at the top—they can turn unsolvable problems into relatively simple ones. Many problems in cognitive science, such as language learning and scene interpretation, apparently involve the cognitive equivalent of an infinitely large sock drawer, and hence require powerful constraints.

Psychologists have applied the formal results on the need for constraints to development and learning, concluding that different domains (including language, but also physics, biology, quantitative reasoning, social relations, and geometry) have their own special structures which must be exploited if learning is to be efficient (Spelke & Kinzler, 2007). Efficiently exploiting these kinds of structures entails having different kinds of constraints for different domains. The specific nature of many of these domains, and the corresponding nature of their internal constraints, was detailed in the contributed articles to the 1990 special issue of *Cognitive Science* devoted to structural constraints on cognitive development. The current special issue of *Cognitive Science* could be considered a 20th anniversary homage to this previous special issue. The contributions to the 2010 issue are no less concerned with constraints than the 20th century issue.

A deeper inspection of the current issue's contents does, however, show an evolution in how cognitive developmentalists conceptualize constraints. The articles in the 1990 special issue tended to posit internal constraints that paralleled structural characteristics of entities in particular evolutionarily important domains. A couple of examples provide helpful reminders as to their general modus operandi. Spelke (1990) argued that infants are constrained to assume that objects follow smooth trajectories through space and time. This is an eminently reasonable assumption because objects do not typically pop into and out of existence spontaneously, but rather move smoothly and vary conservatively. Markman (1990) argued that children need constraints on word meanings in order to learn them in a reasonable amount of time. For example, children assume that a word refers to a whole object rather than part of the object, *ceteris paribus*. They assume that words refer to taxonomic kinds rather than thematically related objects. In addition, they assume that a word will refer to an unlabeled entity, which allows them to overcome the first two constraints if necessary. Keil (1990) argued that ontological knowledge is better described by a tree structure than by a set of arbitrarily overlapping clusters. Predicates must apply to an entire subtree of a hierarchy, thus preventing ''M'' structures. Children, internalizing this ''M-constraint,'' assume that if some things (like mice) can fear but cannot be 2 hours long, and other things (like baseball games) can be 2 hours long but cannot fear, then there should not be still other objects that both fear and last 2 hours.

These constraints all follow the same pattern of postulating an internal bias that fits well with an external exigency. By this approach, we are capable of apt and efficient cognition because our internal structures have evolved over millions of years to reflect external structures of importance for survival (Shepard, 1984). A common conclusion of this approach is that humans, or any other cognitive learning system, cannot be general learning devices or tabula rasas. We need to have constraints like these built into us. Furthermore, because different aspects of the world manifest different structures, we need to have different evolutionarily bestowed constraints for different domains. Hence, Cosmides and Tooby (1992) compare the human mind to a ''swiss army knife'' of different tools that have each been adapted over evolutionary time to their task domain (see Twyman & Newcombe, 2010, for an extended discussion of this, and other, conceptions of modularity).

The exciting possibility raised in various ways by the articles in the current special issue is that experience with a richly and diversely structured world can allow people to devise some of the constraints that they will then use to make learning more from the world more efficient. Although some constraints are surely provided by evolution, others can be acquired during an organism's lifetime and are no less powerful for being learned. In fact, acquired constraints have the advantage of being tailored to an individual's idiosyncratic circumstances. At a first pass, humans seem to live in the same, reasonably fixed world, suggesting that adaptation across generations would be most effective. Indeed, many general environmental factors, such as color characteristics of sunlight, the position of the horizon, and the change in appearance that an approaching object undergoes, have all been mostly stable over the time that the human visual system has developed.

However, if we look more closely, there is an important sense in which people face different environments. Namely, to a large extent, a person's environment consists of animals, people, and things made by people. Animals and people have been designed by evolution to show variability, and artifacts vary widely across cultures. Evolutionary pressures may have been able to build a perceptual system that is generally adept at processing faces (Bruce, 1998), but they could not have hardwired a neural system to be adept at processing an arbitrary face, say that of Barack Obama, for the simple reason that there is too much generational variability among faces. Individual faces show variability from generation to generation, and variability is apparent over only slightly longer intervals for artifacts, words, ecological environments, and animal appearances. Thus, we can be virtually positive that hand tools show too much variability over time for there to be a hardwired detector for hammers. Words and languages vary too much for there to be a hardwired detector for the written letter ''A.'' Biological organisms are too geographically diverse for people to have formed a hardwired ''cow'' detector. When environmental variability is high, the best evolutionary strategy for an organism is to develop a general perceptual system that can adapt to its local conditions.

These adaptations, once effected, act as constraints at different levels of specificity. When adapting to a single person such as Barack Obama, our early expectations may constrain how we interpret his future actions and appearances. Learned constraints have far wider implications when they are distilled from experiences with many different objects. For example, Smith, Colunga, and Yoshida (2010) report earlier experiments that children

extended a label by shape and texture when the objects were presented with eyes (signaling animacy), but extended the label by shape alone when the target and test objects were presented without eyes. Likewise, for toys, children learn that shape matters, whereas for foods, material matters (Macario, 1991; see discussion by Sloutsky, 2010). Some evidence that these biases are learned is indicated by results showing that laboratory training allows students to acquire some of these biases at an age before they normally emerge (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). As a second example, Madole and Cohen (1995) describe how 14- month-old children learn part-function correlations that violate real-world events, whereas 18-month-old children do not learn these correlations, suggesting that children acquire constraints on the types of correlations that they will learn. As a final example, early language experience establishes general hypotheses about how stress patterns inform word boundaries (Jusczyk, Houston, & Newsome, 1999). Children are flexible enough to acquire either the constraints imposed by a stress-timed language like English or a syllable-timed language like Italian, but once they imprint on the systematicities within a language, they are biased to segment speech streams into words according to these acquired biases. In all these cases, constraints are acquired that subsequently influence how children will learn other materials from the same domain.

## 2. Learning overhypotheses

A learning system must have constraints on hypothesis formation in order to learn concepts in a practical amount of time, but a considerable amount of flexibility is still needed because different people face different worlds and tasks. Several of the articles in this special issue explore ways in which this dilemma can be resolved by making constraints themselves learnable. One way to think about this possibility is in terms of Nelson Goodman's (1954) notion of an overhypothesis, a hypothesis of the form ''All As are B'' where A and B are generalizations of terms used in any other hypothesis that we are interested in (Kemp, Goodman, & Tenenbaum, 2010; Kemp, Perfors, & Tenenbaum, 2007). One might have hypotheses that all dogs have four legs, all storks have two legs, and all worms have no legs. Generalizing over both animals and leg number, one could construct an overhypothesis that ''All animals of a particular type have a characteristic number of legs.'' The power of such a hypothesis is that upon seeing only a single six-legged beetle, one can infer that all beetles have six legs. Research indicates that adults employ probabilistic versions of overhypotheses such as these (Heit & Rubinstein, 1994).

Kemp et al. (2010) present a quantitative, formal approach to learning overhypotheses. Their Hierarchical Bayesian Framework describes a method for learning hypotheses at multiple levels, as with the legged animals' example provided earlier. Representations at higher levels capture knowledge that supports learning at the next level down. Learning at multiple levels proceeds simultaneously, with higher-level schemas acquired at the same time that causal models for multiple specific objects are being learned. This mechanism allows Kemp to accommodate, at least in spirit, the examples of constraint learning described in Section 1. Abstract knowledge supports causal learning involving specific objects, but critically, this

abstract knowledge itself can be acquired by statistical learning. Accordingly, it provides a way of learning, rather than simply declaring by fiat, the abstract domains that will govern causal inferences. Their schema-learning approach discovers causal types instead of stipulating them in advance. As an example, it learns that there are two types of blocks—ones that activate a machine and ones that do not, modeling experiments reported by Gopnik et al. (2004).

Kemp et al.'s models have multiple levels of abstraction, and so there might be a level that learns, for example, that pens are reliably pen-shaped and buckets are bucket-shaped, that these might both belong to a higher level that groups them together as artifacts, and at this level the constraint can be expressed that all artifacts have a characteristic shape whatever that shape is, thereby acquiring a shape bias for artifacts (Smith, Colunga, & Yoshida, 2010). Results like these suggest that association-learning devices are crucially undervalued if we only focus on token-to-token associations.[1] A child seeing a penguin is not just learning that penguins are black and white but is also learning about relations between coloration, shape, behavior, climate, diet, and so on, for birds, animals, and natural kinds.

These kinds of type-to-type associations do not release us from a dependency on constraints. In fact, given the unlimited number of abstract descriptions applicable to an observed event, constraints become particularly important in directing us toward useful levels of abstraction. For example, in the model presented by Kemp et al. (2010), the space of higher-level causal models must be fully specified ahead of time, leading to strong constraints on what kinds of abstract models the system can learn. Still, because constraints at this higher level will presumably apply to any novel particular domain, they are best seen as constraints on how experience drives the construction of special domains. The possibility of learning these type-to-type associations goes a long way toward severing the traditional connection between domain-specific constraints and innateness. Learning is not only caused by constraints, but it also causes constraints.

Ample empirical evidence for the flexibility of constraints is provided by Sloutsky (2010). For example, he reports recent experiments showing that people are highly flexible in attending to different features in different micro-contexts (Sloutsky & Fisher, 2008). In one context, shape is relevant, and in another context color is relevant. When a context is reinstated, people selectively weight the contextually relevant dimension. Impressively, this is achieved even with as minimal a manipulation of context as screen location and background color. Other related demonstrations have shown that people will selectively attend to different stimulus dimensions as a function of contextual cues that are provided by the features of the stimuli themselves (Aha & Goldstone, 1992). These manipulations of context fall short of genuine domains, but in some ways, the minimalism of the contextual manipulations is the strength. If people can learn to attend to different properties with arbitrarily created and minimally different contexts, then certainly domains as different as geometry and social relations would have considerably more internal structure that could be leveraged to self-organize a division between them.

Sloutsky suggests that contexts can be induced through a compression-based learning system even before a selection-based learning system has come online in an organism's development. This is particularly so for ''dense'' categories in which different dimensions

are highly correlated with each other. In practice, selection and compression will typically work in tandem, by creating categories that ignore some features (via selection) while at the same time creating compact representations (via compression) that represent an assembly of co-occurring features by a centroid (Love, Medin, & Gureckis, 2004).

One of the reasons why compression often seems to precede selection for natural categories is that selection requires that a categorizer has first differentiated their world's objects into dimensions. In some cases, early developed perceptual systems serve to split an object into separate dimensions. However, in other cases, much later experience provides the impetus to differentiate otherwise fused dimensions (Goldstone & Steyvers, 2001). Experience informs not only contexts and objects, as Sloutsky shows, but also the very descriptions along with the objects are encoded. Statistics from the world can clump situations into contexts, objects into categories, and parts of an object into features. Each of these clumps, once established, influences future learning.

Focusing on the development of infant visual perception, Johnson (2010) gives several compelling examples of learning to see as a constraint-creating activity. He provides evidence that infants originally see their world in disconnected fragments, and that exposure to faces and objects is necessary for infants to eventually come to see entities like these as coherent. In one reported paradigm, more 6- than 4-month-old infants show anticipatory eye movements that are initiated before a ball emerges from behind an occluder. This suggests that spatiotemporal completion strengthens during this 2-month period. Once the infant learns to correctly anticipate where an object will be, he or she is better able to look in the right place to extract more information about the object. In this fashion, learning begets still more learning. A large part of this rich-get-richer effect stems from the role that learning has in creating oculomotor patterns that appropriately constrain future information acquisition, and consequently learning. Johnson describes another excellent example of this dynamic in the work of Needham and Baillargeon (1998). Exposing infants to single or paired objects tends to lead the infants to parse subsequent events in terms of these familiarized configurations. Infants initially exposed to a cylinder abutting a rectangular box showed relatively long looking times, suggesting surprise, if one of the objects subsequently moved separately from the other. Consistent with many of the results described by Johnson, infants are surprisingly adept at adapting their perceptual systems to statistical regularities in their environment. As their visual systems become tailored to their world, they become constrained to see their world in terms of the regularities they have extracted. However, rather than viewing these acquired constraints as limiting perceptual abilities, it is more apt to view these constraints as permitting the infant to see a coherent and well-behaved world (Medin et al., 1990).

## 3. Active construction of entities and kinds

Thus far, the argument has been that any organism that would learn efficiently needs to have constraints that apply to learning particular domains, but at least some of these constraints are learnable. Different constraints can be simultaneously learned for different

contexts, object classes, modules, and domains because entities in the world naturally form recognizable clumps. Psychology is still important because the natural world can be carved into domains in many different ways depending on needs and goals. These goals shape the kinds of clumps that will be formed, but this is different from claiming that the clumps are preformed. Well-understood mechanisms of self-organization allow modules to be constructed for classes of objects based upon their constraints (Elman et al., 1996). By carving nature at its joints, clusters are formed such that the entities within a class are similarly constrained. Furthermore, once formed, the clusters reinforce and emphasize the distinctions between the entities. Joints are carved into nature where they were already incipient, making the joints sharper still (Lupyan, 2005). We do not need to start with domain-specific constraints. The specific domains can emerge from more domain-general principles of association, contingency detection, statistical learning, and clustering.

The current issue's articles propose a second way in which constraints are actively constructed rather than fixed. In particular, another recurring theme is that people play an active role in creating the entities to be learned. This theme is perhaps clearest in Chater and Christiansen's (2010) arguments that language speakers shape their language over generations in ways that make it more easily learned by others. Rather than language learning consisting of the acquisition of fixed structures in a natural, linguistic world, the task confronting language learners is typically one of ''C-Learning''—simply learning to coordinate with other individuals. This is a much easier task as long as the learner exists in a milieu in which the other individuals are generally configured similarly to the learner. On this view, languages are evolving to be learnable, at the same time that people are evolving to learn language. Language evolution assumes particular importance in this view, because languages change at a much faster rate than do genes. Certainly individuals still need to acquire their indigenous language, but this will be a language that has been rapidly evolved so as to be efficiently learnable by people with general perceptuo-motor, communicative, and cognitive constraints. The premise that languages can evolve relatively quickly is supported by documented reports that Nicaraguan sign language emerged in as little as three decades within a community of deaf children with little exposure to established languages (Senghas, Kita, & Özyürek, 2004).

Casting language as C-learning does not trivialize its difficulty. If language were purely a problem of coordination, it could be solvable by creating a very simple language containing only one word. But a language evolves under several distinct selection pressures, in addition to learnability. A language should be easily comprehended once learned, so ambiguities and confusions may make a language less successful, even if easily learned. Most important, a language must have sufficient power to express rich and structured thoughts. Language evolution is thus not completely untethered. Nonetheless, language does provide an excellent case study of a domain that is configured by general human constraints rather than existing as a preconfigured domain requiring language-specific constraints to acquire it.

Although focusing on language acquisition rather than language evolution, Smith, Colunga, and Yoshida (2010) nonetheless echo several of the themes raised by Chater and Christiansen. Smith et al. again point to ways in which language is shaped by general cognitive constraints, including general attentional effects, the highlighting of novel cues that

co-occur with novel outcomes, the illusory projection of cues that are associated with other cues but are not themselves present, and the construction of clusters reflecting correlated features (see also Sloutsky's compression mechanism). They show that language, as it is constructed by a child in a particular environment, reaches back to affect how that environment is coded. Unlike English, Japanese makes no distinction between count and mass nouns. However, giving Japanese children training with the kind of correlated linguistic cues that an English child might receive causes them to behave more as an English child might. In particular, they generalize names for solid things by shape and nonsolid things by material, rather than the less sharply delineated generalization pattern for an average Japanese child. Critically, this effect of linguistic training has a lasting influence on children even when the linguistic cues are no longer present. Language is instrumental in establishing categories like count, mass, and animate nouns. Although scaffolded, in part, by language, these categories remain in place when the scaffold is removed. It is this dynamic that leads Smith et al. to argue that children are forming the domains via which they will organize their world, and language both reflects and guides these acts of creative construction.

Applying this approach to domain-specificity, many of this issue's articles pursue the possibility that domain-general constraints are sufficient to produce what eventually become, or simply appear to be, domain-specific constraints, including Chater and Christiansen (2010), Twyman & Newcombe, Sloutsky (2010), and Smith, Colunga, and Yoshida (2010). Recall that the basis for the traditional link between domain-specificity and constraints lies in the idea that there are innate constraints associated with each kind of stuff—each domain, or area of ''core knowledge'' (Spelke & Kinzler, 2007). Empirical evidence indicates that we do not need to start with domain-specific constraints. The specific domains can come from more domain-general principles. Chater and Christiansen (2010) describe the existence of general cognitive routines for processing sequential information used in natural language reading and statistical learning; language is not as special as might have been believed. Some of the domain-general processes that they single out include encoding, organization, and production of temporally unfolding events. These processes are useful for finding structures in language and visual sequences alike, and positing domain-general constraints helps to explain patterns of correlation between language and visual temporal processing. Once it has been learned, language is highly constrained, but if we just look at the eventual constrained forms, it is easy to forget where they came from.

Likewise, for the domain of spatial navigation, Twyman and Newcombe (2010) argue that people, old and young alike, integrate across multiple cues, including featural cues, not just a single geometric module. The apparently constrained nature of children's navigation gives way to broader domain-general processes. Even children can use curtain color, a cue not considered to be part of the ''geometric module,'' when the room is large and they have experience with the task. We need to think about how cues are combined and integrated, which is both the bane and boon of domain-general processes. Information has to be combined across multiple sources, a process that is sometimes complex. However, the complexity is often more than justified by the benefits conferred by mixing expert systems, and by having these expert systems simultaneously train each other (de Sa & Ballard, 1998).

## 4. Adding skeletons to flesh

An important plank of the new approach to developmental constraints espoused in this issue's pages is that developing children learn to organize their world into the categories that they will use to guide their inferences. This New School of Constraints agrees with the Old School on the basic principle that different properties are important for different domains. For example, Gelman and Markman (1986) present evidence that children generalize from biological properties to objects with the same superordinate name but with a different appearance. For the biological property ''cold blooded,'' children extend more inferences from triceratops to brontosauruses than to rhinoceroses. However, for physical properties (like weighing 1 ton), they generalized more to the rhinoceros (which resembled a triceratops) than brontosaurus. Heit and Rubinstein (1994) find that adults are as flexible as children, for example generalizing an anatomical property more from chickens to hawks than from tigers to hawks, but generalizing feeding and predation properties more from tigers to hawks. Earlier, Nisbett et al. (1983) showed that reading about just one member of a tribe with a certain skin color makes people think that all tribe members have that skin color, but they are not as profligate with their inductions about the generality of obesity in the tribe upon seeing only one obese member.

One possible conclusion from these kinds of studies is that both children and adults come to their world already having broken it down into kinds of things, and this is critical because in order to know how something will behave, one needs to know what kind of thing it is. Domains provide the skeleton on which to hang knowledge. As Gelman (1990) writes, ''I find it helpful to think of a skeleton as a metaphor for my notion of first principles. Were there no skeletons to dictate the shape and contents of the bodies of the pertinent knowledge, then the acquired representations would not cohere'' (p. 82).

In referencing skeletons, Gelman clearly has in mind an a priori structure on which to hang experiential knowledge. However, to the extent that we, the authors, find skeletons to be an apt metaphor for knowledge, it is only when we reflect on the fact that skeletons themselves are not a priori structures, but rather unfold with a developmental process themselves, and are molded by need. The tennis player John McEnroe's right hand is significantly larger than his left hand, because his skeleton and musculature are not a priori givens. In general, bone mineral content is greater in the dominant arm of professional tennis players than in their contralateral arm, but not so for a control group (Calbet, Moysi, Dorado, & Rodríguez, 1998). Tennis players' literal skeletons have adapted to fit their tennis requirements. Neural network models provide working examples of skeletons forming because of the inputs provided to them. Bernd Fritzke's (1994) Growing Neural Gas model provides a compelling example of this (see Fig. 1). When inputs are presented, edges are grown between nodes that are close to the input, and new nodes are created if no node is sufficiently close to the input. The result is a skeleton that can aptly accommodate new knowledge because it was formed exactly in order to accommodate the knowledge. This skeleton-creating approach appears also in ''Rethinking innateness'' (Elman et al., 1996), where one of the primary ideas is that the existence of modularity does not implicate innateness. Modules can be learned because systems can self-organize themselves to have increasingly rich and differentiated structure.

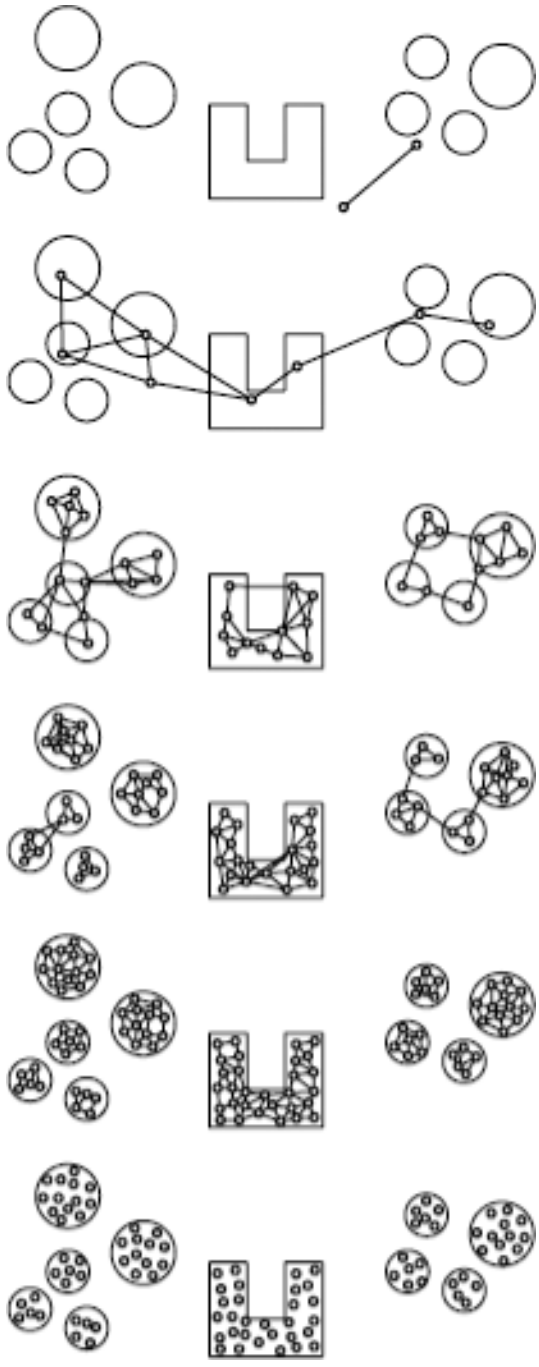

Fig. 1. In Fritzke's (1994) Growing Neural Gas model, the skeletal structure of the network is formed by the inputs themselves. (Figure adapted with permission by the author).

Computational modeling suggests that the eventual specialization of a neural module often belies its rather general origins (Jacobs, Jordan, & Barto, 1991). Very general neural differences, such as whether a set of neurons has a little or a lot of overlap in their receptive fields, can cause the two populations of neurons to spontaneously specialize for handling either categorical or continuous judgment tasks, or snowball small initial differences into large-scale ''what'' versus ''where'' visual systems (Jacobs & Jordan, 1992). At a higher level of abstraction, self-organizing neural network models have been proposed that account for how originally undifferentiated concepts become differentiated and increasingly structured with development (Rogers & McClelland, 2008). Without belaboring the details of these models, there are a sufficient number of examples of skeleton-creating working models to believe that to know how something will behave, one needs to know what kind of thing it is, but that these kinds can emerge through the progressive differentiation of objects into domains with experience.

## 5. Getting physical about constraint learning

To our mind, the articles in this special issue offer true advances in our understanding of cognitive constraints as developing over a person's lifetime. However, in the interest of urging the field to not rest on its laurels, we wish to point out that the models that have been presented to make this point strike us as incorporating input representations that are highly preprocessed and symbolic. This modeling decision hampers the models from producing as novel constraints and representational capacities as they otherwise might.

First, let us give some examples of what we feel are modeling choices that constrain too tightly the kinds of constraints that can be induced. Kemp et al.'s (2010) model succeeds in simultaneously learning object-level causal models, category-level causal models, and the categories that occupy the upper levels. However, to achieve these, the authors assume that the learner already has divided the world into domains (e.g., people and drugs) and events (e.g., ingestion and headache events). Furthermore, it is prebuilt to learn causal models that relate the ingestion of drugs to headaches. Finally, actual events like ''has headache'' are precoded as atomic, nondecomposable symbols. There is no perceptual apparatus that grounds objects, events, or causal relations, and hence no way to adapt perceptual processes to establish new kinds of entities. To be fair, the authors admit all these constraints, and they gesture to some possible ways of adding flexibility to their model.

To take a second example from a different modeling tradition, Rogers and McClelland's (2004, 2008) neural network model is an attempt to understand how domain-specific knowledge emerges solely from general learning principles. In this sense, it fits well within the current articles' leitmotif that preestablished constraints on preestablished domains is not required to account for the eventually structured form of cognitive representations. The particular structures that their PDP model, trained by back-propagation, acquires are as follows: progressively differentiated concepts, coherent categories characterized by clusters of co-occurring features, conceptual reorganization over time, and domain-specific attribute weighting. Their system takes statements such as ''Canaries can fly'' and ''Canaries can

grow'' as input statements and creates emergent and shifting clusters for different kinds of animals and attributes. From the perspective of learning constraints, though, we feel that the input representations are limiting. Single nodes are dedicated to each element in a proposition, such as ''Canaries,'' ''Can,'' and ''Fly.'' These nodes are not connected to a physical world via a perceptual system, so once again, there is no perceptual system to adapt.

A natural response to our objection is ''One has to start somewhere. Progress can be made on constraint learning without accounting for the organism's relation to its world.'' We do not deny that some progress can be made, but at the same time we feel that the most compelling examples of learning new objects and domains come exactly from situations where the organism's embedding in the world is rich, high-bandwidth, and dynamic (Beer, 2008). In this respect, we are in agreement with Johnson (2010) and Smith, Colunga, and Yoshida (2010).

### *5.1. Case studies of novel constraint generation in physical systems*

To show the kinds of novel objects and constraints that can emerge in situated systems, we will consider two working, physical devices, without asserting that they are formal models of learning. Fig. 2 shows the first physical model, developed by Gordon Pask (1958; see also Cariani, 1993), which features electrodes immersed in a dish containing a ferrous sulfate solution. Passing current through the electrodes caused dendritic metallic filaments to grow as precipitates from the fluid. Ferrous filaments could be adaptively grown to make the system sensitive to sounds. Early on, the system could only detect the presence or absence of sounds, but once filaments grew that joined electrodes and changed electrical conductance, the device was able to discriminate two frequencies. The conducting filament pathway is shaped by exactly the vibrational perturbations that it detects, and how it detects the perturbations is changed by its growth. This device has the capacity to represent things, like the difference between tones of 50 and 100 cycles⁄s, which it was not originally able to represent. This kind of model provides a compelling existence proof for a system that creates its own constraints—constraints that were not originally there before a certain physical connection was made. It is a device that, when (literally) immersed in the proper environment, develops its own concept of what is relevant.

Our second example is a physical device meta-designed to accomplish, perhaps coincidentally, a similar tone discrimination task. For this task, Thompson, Layzell, and Zebulum (1999) employed a field-programmable gate array (FPGA) containing a $10 \times 10$ array of programmable logic components called ''logic blocks.'' The logic blocks consist of multiplexers that act like switches to determine input–output relations between blocks. A computer is used to configure the multiplexers by sending them a stream of bits, thereby causing the multiplexer to physically instantiate a particular electronic circuit on the chip. FPGAs are thus integrated circuits with hardware that can be configured to specify how the logic blocks are connected to each other. Thompson et al. used a genetic algorithm to randomly alter the hardware of the FPGA, and then tested to see how well the resulting, automatically configured FPGA could accomplish the task of discriminating between 1- and 10-kHz square wave tones. After 5000 generations, the FPGA could solve the task well. The best performing hardware design is shown in Fig. 3. In this figure, the upper-left panel shows the

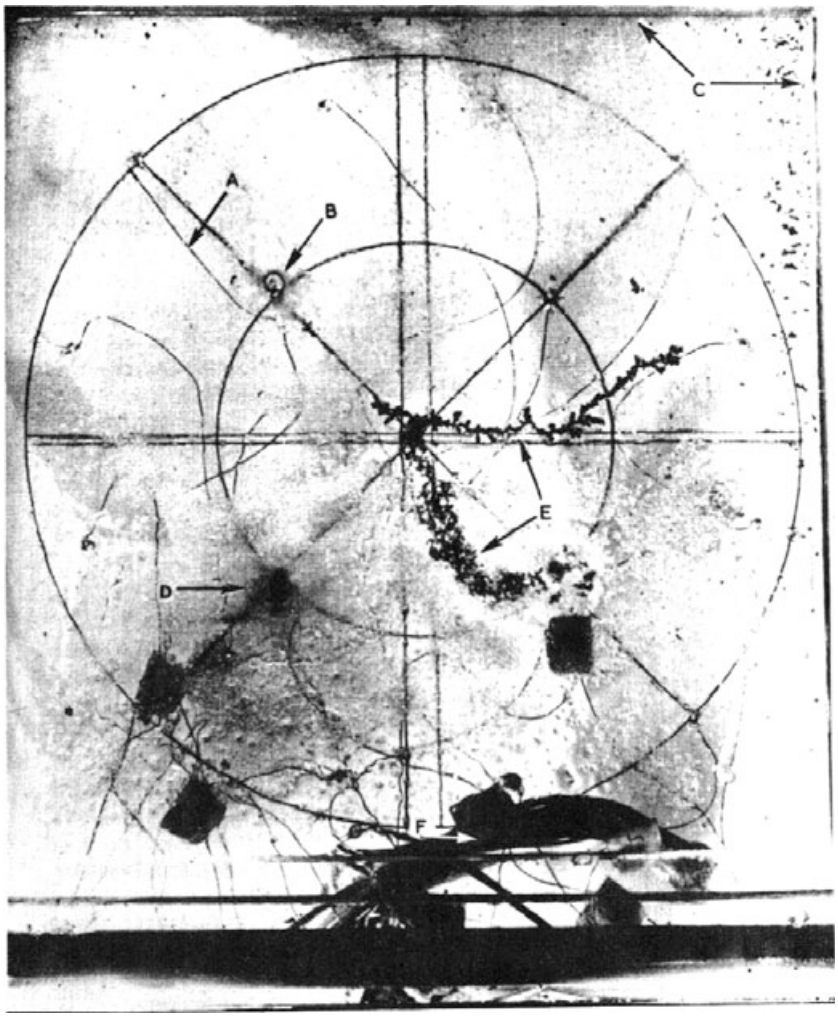


Fig. 2. Gordon Pask's electromagnetic ear. Dendritic filaments of ferrous sulfate (marked as E) grow to connect originally disconnected electrodes. The circular wires are a support frame. (Reprinted with permission).

entire set of connections among the logic blocks. However, because a random genetic algorithm was employed, there is no guarantee that this FPGA is actually a coherent electronic circuit at all. In fact, it was not, and some of the logic blocks are not part of the logical circuit of the FPGA that classifies a tone as high or low. These pruned blocks are not part of any connected pathway that connects to the output units. Pruning these blocks yields the upper-right panel. The remaining blocks are part of the logical circuit and could influence the FPGA's output, but that is no guarantee that they do. The authors clamped the values of each of the full $10 \times 10$ set of blocks—both those that were and were not included in the pruned circuit—to determine if a block did ever influence outputs. Removing the largest set of blocks that can be clamped without affecting performance from the electrical diagram

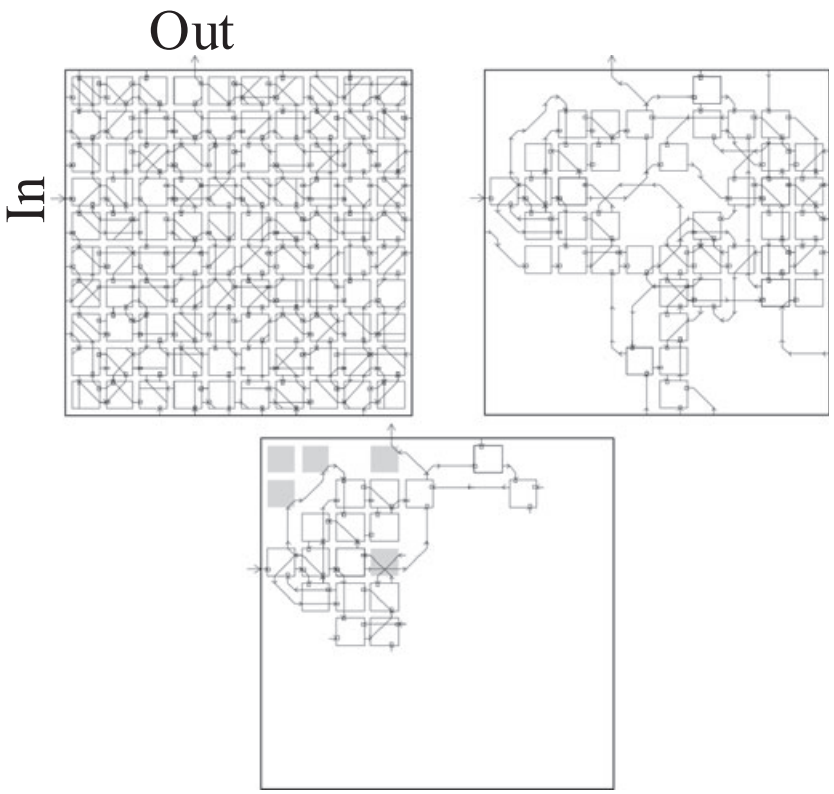


Fig. 3. A schematic description of FPGA (field-programmable gate array) circuits evolved by Thompson et al. (1999) to solve a tone discrimination task. The upper left circuit shows the electrical flow diagram of the FPGA unit that has been evolved to successfully solve the tone discrimination task. The upper-right circuit shows the pruned diagram that removes all blocks that have no connected pathway that could have affected the output. The lower circuit shows the functional part of the circuit that remains after the largest possible set of cells has been clamped without affecting performance. In the lower circuit, the gray cells are cells that influence the output but are not part of the logical circuit.

results in the lower panel of Fig. 3. Comparing the two methods of paring down the full $10 \times 10$ array of blocks reveals a surprising pattern. Some blocks (shown in gray) that were not part of logical circuit of the evolved solution are nonetheless important for the FPGA's success at the task. When they are clamped to constant but random selected values, performance suffers; hence, they influence the circuit's behavior and quantifiable performance. Furthermore, when only the logical circuit is implemented in a new, nominally identical FPGA, it no longer solves the task as well as the original circuit did. Likewise, a digital simulation of the evolved FPGA's circuit did not produce the same behavior as the physical FPGA.

This apparent contradiction is reconciled by understanding that although the FPGA is a digital chip, it also has analog properties as a physical device. Conventional digital circuit design makes the assumption that a given block will output only one of two states at any give time, when actually there are transition periods. The ''digital assumption'' has the

prominent advantage that it allows us to think in terms of a logic algebra. However, the multiplexer switches work according to laws of semiconductor physics. The circuit changes as a real-time, continuous-valued electrical system. For some of the cells that are part of the functional, but not logical, circuit, inputs are routed into them from the active circuit, but their outputs are not used by the active network. These cells influence the timing of the signals routed through or near them. If the configuration of the gray cells is changed, this affects the capacitance of these routes, and hence the time delays for signals to travel along them. These signals include not only on/off states but also transitional values between these two states, which are normally considered the only possible states in formal electronics. Thus, the cells that are not part of the logical circuit can still change the timing of the rest of the circuit by influencing their transient states, and the tone discrimination task is highly dependent on timing, so these changes are functionally important.

### 5.2. *Reliably traumatic learning in physical systems*

With these two examples in mind, we can address a question core for the field of cognitive development: How can systems develop genuinely new cognitive capacities? One answer is simply that they cannot. Fodor (1980) argues that learning a new concept necessarily involves first hypothesis formation and then testing the formed hypothesis. Therefore, a person could not acquire a new concept unless he or she already had the ability to represent its hypothesis. Learning can determine whether the hypothesis is true, but the fundamental ability to represent the hypothesis cannot be learned. If a person can learn that a square is a four-sided object with all angles equal and all sides equal, then it must be because the person already had the wherewithal to represent concepts such as ''four,'' ''angles,'' and ''equal.'' A system can increase its representational power by ''physical trauma'' like a blow to the head, but not through formal inductive learning processes. Inductive learning does not increase a system's representational ''vocabulary'' because mechanisms must already have been in place to express the ''new'' vocabulary, and so it has not been genuinely created.

A related argument is presented by Kemp et al. (2010) in defending their approach from the criticism that their model does not learn or discover causal schemata, but rather only selects one schema from a prespecified space of hypotheses. Kemp et al.'s response is that from a computational perspective, every learner begins with a prespecified hypothesis space that represents the abstract potential of the learner.[2] This hypothesis space includes all reachable states of knowledge given all possible empirical inputs. Both Fodor's and Kemp et al.'s arguments take the form: True novelty is impossible because all hypotheses (or their components, for Fodor) must exist to be (eventually) selectable.

Pask's device tidily points out the simplistic nature of this argument. There may be a sense in which all discriminations that the device can learn are present in the device when it is first constructed. However, that is only the same trivial sense in which all organs of all life forms that currently inhabit the planet were present in the earliest bacteria. This trivializes the physical changes that allow Pask's device to represent distinctions between sounds that it was originally unable to make. These physical changes may be dismissed as ''trauma,'' but they are nonetheless highly systematic, much as are the actual physical changes in the

brain that lead to long-term potentiation across the synapse between two neurons that are co-stimulated. The physicality of the change in Pask's device makes it instructively clear that at one point in its development, prior to the existence of a conductive filament that has grown to connect two electrodes, it is incapable of making a sound discrimination, but that after the physical change has transpired, it is capable.

We agree with Kemp et al. (2010) that a formal model of learning must originally contain all the hypotheses it will eventually be able to entertain. It is important to remember, though, that formal models are not the same thing as working, rigorous, and replicable models. Thompson's FPGA and Pask's device may well be models of the latter kind, and they offer the kind of physical models that may well be needed to yield genuine novelty with development. Thompson's evolved FPGA clearly shows the disadvantage of honing too closely to a formal algebra. The formal model systematically eliminates the possibility of discovering solutions to the sound discrimination task that are outside of the digital circuit framework. When the evolved circuit is reduced to its formal form, it no longer solves the task.

A physical device such as Thompson's FPGA can have more than one appropriate formal idealization. No doubt there is a more elaborate formal logic that captures via explicit representation the temporal properties of the FPGA, just as a sufficiently complex formal model could capture the chemical properties of Pask's device. Such a richer model could well solve the tasks these physical devices do, and all the capabilities of these devices would then be implicitly contained in that formal description. But, crucially, in order to design that formal model, one would have to know just which physical properties mattered to the behavior of interest. From this perspective, to observe that all of the possible hypotheses that can be entertained by a formal system are already implicit in the definition of the system is akin to noting that by the time one can build a satisfying formal model, one must already know how the real system works. On the other hand, a physical model or a model with rich, high-bandwidth connections to an external physical environment can exhibit properties that are not contained in any preexisting conceptualization. The moral, for us, is that we want our models, like our children, to be sufficiently flexible that they can surprise us with solutions that we had not anticipated by our formal analyses.

### *5.3. Constraint generation by interacting with other physical systems*

Physical interactions with an external environment may also play a role in constructing and constraining a hypothesis space. Several recent models of high-level cognitive activities depend on the routine application of deictic representations (Agre & Chapman, 1987; Ballard, Hayhoe, Pook, & Rao, 1997). For instance, Patsenko and Altmann (2010) describe a model of the process of solving the Towers of Hanoi. This is a quintessentially high-level task, but Patseno and Altmann's model performs remarkably little reasoning. Instead, a quite simple control system defines a method for manipulating visual pointers and physical objects. Assuming a reliably stable external environment, this control system, which includes rules for updating referential representations, suffices to perform the Towers of Hanoi task in a manner that closely matches human behavior. Landy, Jones, and Goldstone (2008) suggested similarly that deictic representations might drive human syntactic judgments in formal arithmetic.

To the extent that reasoning depends on deictic representations of this kind, the ability of an agent to maintain and coordinate multiple referential symbols provides a natural constraint on the kinds of hypotheses that can be maintained. There is little reason to think of the maintenance of these referential pointers as a purely internal, representational process. Rather, to the degree that hypotheses are represented via interactions with external representations, the maintenance of those hypotheses depends both on internal resources, and on various physical properties of the interacting systems, the external environment, and the ability of the agent to manipulate that environment (Kirsh, 1995). In this case, the constraints that limit the hypotheses that can be constructed are encoded in the physical structure of the environment and agent–environment interactions.

In some cases, hypotheses themselves may be built out of environmental components, with the result that constraints on one's ability to construct physical analogs will limit and constrain the hypothesis space that can be constructed. Seventeenth-century physicists generated their hypotheses about physical phenomena by mapping them onto known physical mechanisms (Bertolini Meli, 2006). Nersessian (2009) suggested that the physical models engineering scientists construct often serve a role in the generation and representation of hypotheses: Reasoners construct a physical model, which serves both as an object of study, and as a repository for theory. Given that the physical instantiation is not well understood, making a prediction from the theory may require consulting—running—the model. When hypotheses are partially external constructs, our ability to form hypotheses will be constrained not just by the limitations of a predefined hypothesis space intrinsic to an agent but also by our practical ability to build particular physical models that instantiate theories. For example, if a particular neural network is too complex to build, or would take to long to run on an available computer, one may well simplify it. The physical, external constraints limit the available hypothesis space in a manner neither fixed nor internal, but nevertheless quite restrictive.

### *5.4. Conclusions and caveats to the novelty that physical systems permit*

To the extent that concepts like squares and angles are rooted in one's ability to produce and perceive physical models of squares and angles, one's hypothesis space is at least softly constrained by the ability to coordinate fingers, limbs, and counting procedures. No less than scientists, children are situated, physical systems, and their physical presence is critical to their ability to develop their own constraints, and to increase their own representational capacities. This does not mean that we eschew computational and mathematical models of cognitive development because of their lack of physicality. However, we do recommend efforts to move beyond models, be they connectionist or Bayesian, that severely constrain the hypothesis space (Landy & Goldstone, 2005). We advocate models that preserve enough rich spatial, perceptual, dynamic, and concrete information for surprising new classes of hypotheses and encodings to emerge. We believe the hard problem for cognitive development will not be selecting from hypotheses or creating associations between already delineated elements, but rather constructing hypothesis spaces and elements in the first place.

We have argued for interpenetrations between people and their environments, and between the physical and functional descriptions within learning devices such as people. Furthermore, we have argued that these interpenetrations are crucial in developing learning systems that create genuinely new hypotheses, not just permutations of preestablished symbols. However, a critic might point out that these interpenetrations make for fault-prone and noise-intolerant devices. Thompson et al. (1999) not withstanding, most current electronics are designed to behave digitally precisely to provide tolerance to superficial variation in voltage signals that are irrelevant to the critical information. Even ''squishy'' organic life as we know it owes its longevity to a genetic code that closely approximates a digital code consisting of nucleotides and codons. Complex cellular machinery is dedicated to assuring that the code is relatively inert and is protected from many contextual influences (Rocha & Hordijk, 2005). It is reasonable to think that our cognitive system benefits from the same strategy of developing reusable, quasi-context-independent codes (Dietrich & Markman, 2003). Much of the benefit of discrete codes is exactly that they are not buffeted about by extraneous contextual influences.

We agree that systems that behave reliably often have evolved to have representations that are mostly sealed off from lower or outside pressures. Simon (1969) called such systems with multiple, encapsulated levels ''nearly decomposable,'' and we agree with their importance in cognitive systems but wish to equally emphasize the importance of the qualifier ''nearly.'' If Thompson et al. (1999) had designed electrically ''cleaner'' circuits with proper electrical shielding around cells, then their system would not have been able to evolve the same solutions to the sound detection problem based on influences of electric flow on nearby capacitances. Analogous advantages are found for systems that coordinate with their physical environment to solve problems that they could not otherwise solve. The advantage of creating new representations by recombining existing codes is too powerful to forego. However, systems are still more flexible when this combinatorial flexibility is extended by interpenetrations across levels and systems.

## 6. Conclusion

A good deal of cognitive development circa 1990 involved delineating specific constraints in domains such as language, motion, quantitative reasoning, social perception, and navigation. This focus on domain-specific constraints has now shifted to domain-creating constraints, and an interest in how general learning processes can give rise to learned domains, dimensions, categories, and contexts. This new movement need not argue that all domains are constructed. The major claim is simply that children need to learn to learn, and a key component of this second-order learning is organizing their world into kinds of things that are similarly constrained. If these caricatures[3] of the past and present of cognitive development bear any resemblance to actual history, then our speculative claim for the future of cognitive development research is that it will push to describe how genuine cognitive novelty arises. We expect that these efforts will not all be formal in the sense of involving only rules for manipulating symbols. However, they can still be rigorous and mechanistic, providing working models for how things such as concepts, domains, and formalisms themselves arise.

## Notes

1. Kemp et al. do not consider their models to be association learning, but the advantage of learning type-to-type relations can be exploited by non-Bayesian models.
2. The authors also argue that this computational-level response, though correct, could also be supplemented by algorithmic-level accounts that provide a process model for creating hypotheses. The authors do not provide such an account. We are arguing for far greater attention to algorithmic-level, and even implementational-level, process accounts of how new hypotheses are constructed in the first place. Attention to these levels is what allows truly emergent novelty to arise in a system.
3. One respect in which we acknowledge the caricatured nature of our historical comparison is that, even in the 1990s, there was a considerable interest in domain-general learning, and the notion that early constraints can serve to bootstrap the development of more sophisticated properties and constraints subsequently (Keil, 1990; Spelke, 1990).

## Acknowledgments

The authors wish to express thanks to Randall Beer, Lisa Byrge, Vladimir Sloutsky, Linda Smith, Adrian Thompson, and Peter Todd for helpful suggestions on this work. This work was funded by National Science Foundation REESE grant 0910218. More information about the laboratory can be found at http://cognitrn.psych.indiana.edu.

## References

Agre, P., & Chapman, D. (1987). *Pengi: An implementation of a theory of activity.* AAAI National Conference. New York: Morgan-Kaufman.

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. In *Proceedings of the fourteenth annual conference of the Cognitive Science Society* (pp. 534–539). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723–742.

Beer, R. D. (2008). The dynamics of brain-body-environment systems: A status report. In P. Calvo & A. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 99–120). Amsterdam, The Netherlands: Elsevier.

Bertolini Meli, B. (2006). *Thinking with objects: The transformation of mechanics in the seventeenth century*. Baltimore: Johns Hopkins University Press.

Bruce, V. (1998). *In the eye of the beholder: The science of face perception*. New York: Oxford University Press.

Calbet, J. A. L., Moysi, J. S., Dorado, C., & Rodríguez, L. P. (1998). Bone mineral content and density in professional tennis players. *Calcified Tissue International*, *62*, 491–496.

Cariani, P. (1993). To evolve an ear: Epistemological implications of Gordon Pask's electrochemical devices. *Systems Research*, *10*, 19–33.

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science*, *34*(7), 1131–1157.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York: Oxford University Press.

Dietrich, E., & Markman, A. B. (2003). Discrete thoughts: Why cognition must use discrete representations. *Mind and Language*, *18*, 95–119.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Fodor, J. A. (1980). On the impossibility of acquiring more powerful structures. In M. Piatelli-Palmarini, J. Piaget, & N. Chomsky (Eds.), *Langauge and learning: The debate between Jean Piaget and Noam Chomsky* (pp. 142–162). Cambridge, MA: Harvard University Press.

Fritzke, B. (1994). Growing cell structures—A self-organizing network for unsupervised and supervised learning. *Neural Networks*, *7*, 1441–1460.

Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.

Gelman, S. A., & Markman, E. (1986). Categories and induction in young children. *Cognition*, *23*, 183–209.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *16*, 447–474.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116–139.

Goodman, N. (1954). *Fact, fiction, and forecast*. London: University of London, Athlone Press.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.

Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 411–422.

Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias towards short connections. *Journal of Cognitive Neuroscience*, *4*, 323–336.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, *15*, 219–250.

Johnson, S. P. (2010). How infants learn about the visual world. *Cognitive Science*, *34*(7), 1158–1184.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.

Keil, F. (1990). Constraints on constraints. *Cognitive Science*, *14*, 135–168.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–1243.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307–321.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*, 31–68.

Landy, D., & Goldstone, R. L. (2005). How we learn about things we don't already understand. *Journal of Experimental and Theoretical Artificial Intelligence*, *17*, 343–369.

Landy, D. H., Jones, M. N., & Goldstone, R. L. (2008). How the appearance of an operator affects its formal precedence. In *Proceedings of the thirtieth annual conference of the Cognitive Science Society* (pp. 2109–2114). Washington, DC: Cognitive Science Society.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Lupyan, G. (2005). Carving nature at its joints and carving joints into nature: How labels augment category representations. In A. Cangelosi, G. Bugmann, & R. Borisyuk (Eds.), *Modeling language, cognition and action: Proceedings of the 9th neural computation and psychology workshop*. Singapore: World Scientific.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*, 17–46.

Madole, K. L., & Cohen, L. B. (1995). The role of parts in infants' attention to from-function correlations. *Developmental Psychology*, *31*, 637–648.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*, 57–78.
Medin, D. L., Ahn, W.-K., Bettger, J., Florian, F., Goldstone, R., Lassaline, M., Markman, A., Rubinstein, J., & Wisniewski, E. (1990). Safe takeoffs-soft landings. *Cognitive Science*, *14*, 169–178.
Needham, A., & Baillargeon, R. (1998). Effects of prior experience in 4.5-month-old infants' object segregation. *Infant Behavioral Development*, *21*, 1–24.
Nersessian, N. J. (2009). How do engineering scientists think? Model-based simulation in biomedical engineering laboratories. *Topics in Cognitive Science*, *1*, 730–757.
Nisbett, R., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339–363.
Pask, G. (1958). Physical analogues to the growth of a concept. *Mechanization of thought processes, Symposium 10*, National Physical Laboratory, November 24–27 (pp. 765–794). London: H.M.S.O.
Patsenko, E. G. & Altmann, E. M. (2010). How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General*, *139*, 95–116.
Rocha, L. M., & Hordijk, W. (2005). Material representations: From the genetic code to the evolution of cellular automata. *Artificial Life*, *11*, 189–214.
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Rogers, T. T., & McClelland, J. L. (2008). Precise of semantic cognition, a parallel distributed processing approach. *Behavioral and Brain Sciences*, *31*, 689–749.
de Sa, V. R., & Ballard, D. H. (1998). Category learning through multimodality sensing. *Neural Computation*, *10*, 1097–1117.
Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*, 1779–1782.
Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imaging, thinking, and dreaming. *Psychological Review*, *91*, 417–447.
Simon, H. A. (1969). The architecture of complexity. In *The sciences of the artificial* (pp. 192–229). Cambridge, MA: MIT Press.
Sloutsky, V. M., & Fisher, A. V. (2008). Attentional learning and flexible induction: How mundane mechanisms give rise to smart behaviors. *Child Development*, *79*, 639–651.
Sloutsky, V. M. (2010). From perceptual categories to concepts: what develops?. *Cognitive Science*, *34*(7), 1244–1286.
Smith, L. B., Colunga, E., & Yoshida, H. (2010). Knowledge as process: contextually cued attention and early word learning. *Cognitive Science*, *34*(7), 1287–1314.
Smith, L. B., Jones, S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, *13*, 13–19.
Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.
Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.
Thompson, A., Layzell, P., & Zebulum, R. S. (1999). Explorations in design space: Unconventional electronics design through artificial evolution. *IEEE Transactions on Evolutionary Computation*, *3*, 167–196.
Twyman, A. D., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, *34*(7), 1315–1356.
Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*, 1341–1390.