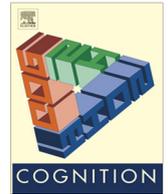




ELSEVIER

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Pigeons acquire multiple categories in parallel via associative learning: A parallel to human word learning?

Edward A. Wasserman^{a,b}, Daniel I. Brooks^c, Bob McMurray^{a,b,d,*}^a Dept. of Psychology, University of Iowa, United States^b Delta Center, University of Iowa, United States^c Dept. of Psychology, Tufts University, United States^d Dept. of Communication Sciences and Disorders, University of Iowa, United States

ARTICLE INFO

Article history:

Received 19 May 2014

Revised 23 September 2014

Accepted 17 November 2014

Available online 8 December 2014

Keywords:

Animal behavior

Associative learning

Categorization

Language

Word learning

Comparative cognition

ABSTRACT

Might there be parallels between category learning in animals and word learning in children? To examine this possibility, we devised a new associative learning technique for teaching pigeons to sort 128 photographs of objects into 16 human language categories. We found that pigeons learned all 16 categories in parallel, they perceived the perceptual coherence of the different object categories, and they generalized their categorization behavior to novel photographs from the training categories. More detailed analyses of the factors that predict trial-by-trial learning implicated a number of factors that may shape learning. First, we found considerable trial-by-trial dependency of pigeons' categorization responses, consistent with several recent studies that invoke this dependency to claim that humans acquire words via symbolic or inferential mechanisms; this finding suggests that such dependencies may also arise in associative systems. Second, our trial-by-trial analyses divulged seemingly irrelevant aspects of the categorization task, like the spatial location of the report responses, which influenced learning. Third, those trial-by-trial analyses also supported the possibility that learning may be determined *both* by strengthening correct stimulus–response associations and by weakening incorrect stimulus–response associations. The parallel between all these findings and important aspects of human word learning suggests that associative learning mechanisms may play a much stronger part in complex human behavior than is commonly believed.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Who was correct: Chomsky or Skinner? This question captures in stark terms one of the oldest debates in cognitive science. Are complex—perhaps uniquely human—behaviors like language acquired via specialized and perhaps innately constrained cognitive mechanisms? Or can such behaviors emerge from more basic and general mechanisms like associative learning (e.g., Fitch, 2010)? This debate spans virtually every level of language: from the

acoustic signal (Lieberman & Whalen, 2000; McMurray & Jongman, 2011), to word learning (Golinkoff & Hirsh-Pasek, 2006; Ramscar, Dye, & Klein, 2013; Xu & Tenenbaum, 2007; Yu & Smith, 2012), to syntax and grammar (Christiansen & Chater, 2008; Hsu & Chater, 2010; McClelland & Patterson, 2002; Pinker & Ullman, 2002), and to basic learning mechanisms that may underlie language acquisition (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran & Thiessen, 2007).

Such debates have resulted in considerable theoretical progress. Behaviorist notions of association have developed into more complex emergentist views (e.g., connectionism, dynamical systems); and, language-specific

* Corresponding author at: Dept. of Psychology, E11 SSH, University of Iowa, Iowa City, IA 52242, United States.

accounts have adopted more sophisticated symbolic and probabilistic algorithms. However, harkening back to Skinner (1957) and (Chomsky, 1958; but see, MacCorquodale, 1970), many of these debates rest on assumptions about what simple mechanisms like associative learning can or cannot do in a domain like language. Yet, there is little empirical basis for appreciating what associative learning can actually do. Studies of human learning can be ambiguous in this regard because the same learning problem may be solved via associative or inferential routes (cf., Medina, Snedeker, Trueswell, & Gleitman, 2011; Yu & Smith, 2007) (and see, Ramscar et al., 2013). Computational models have helped clarify what associative mechanisms can do (Mayor & Plunkett, 2010; Samuelson, 2002), but they also rely on controversial or simplifying assumptions that make them less than definitive. In view of these difficulties, animal models, particularly those that employ less cognitively sophisticated animals like pigeons or rats, may offer a complementary perspective by revealing which learning mechanisms are species-general and by distilling a purely associative learning paradigm through the careful control of inputs, outputs, and reinforcement schedules. Adopting an animal-model approach may prove to be crucial for understanding the potential of purely associative components to contribute to higher-level cognitive abilities like language learning.

This paper begins to address these issues in the context of word learning. As we argue below, no animal model currently exists that captures three critical aspects of word learning. First, existing animal models do not isolate a purely associative framework for word learning, but also include social interaction and an enriched learning environment. Although a purely associative animal model surely represents a pale replica of human word learning, it can also afford a great theoretical benefit by allowing us to isolate and refine our understanding of the associative mechanisms that, although embedded in the more sophisticated cognitive processes of humans, are likely to subservise at least part of word learning. Second, few animal models include tasks in which many inputs are mapped to many outputs. This is a critical problem that limits our ability to extend our knowledge of associative learning primarily in animals to those aspects of human word learning where it might be relevant. Third, most animal models progressively add associations over training. However, children are not successively taught their expanding vocabularies; rather, children are simultaneously barraged with thousands of words only a fraction of which are added as language learning proceeds.

Thus, the primary goal of our study was to develop such an associative learning model that meets these three criteria, using pigeons as experimental subjects. We specifically deployed this model to clarify a recent debate concerning the mechanisms of observational word learning in humans. We wish to stress at the outset that an animal model meeting the above criteria is insufficient to capture the full richness of human word learning. However, an animal model that satisfies these criteria is nevertheless a significant step toward capturing the types of associative processes that may form a crucial component of human

word learning; an effective animal model can therefore help us understand these foundational associative processes. We start our article with a broad review of association in word learning and the motivation for our animal model; we then detail this recent debate on observational word learning; finally, we present our experiment.

1.1. Association as a component of human word learning

In recent years, human word learning has offered an important domain in which the debate between language-specific and emergent associative approaches has played out. The prime challenge of word learning is to map phonological word-forms onto concepts and categories. This mapping seems straightforward, but it requires children to solve a range of problems. When children are confronted with a novel name, infinite interpretations are available involving: any available object, its properties, and so forth (Quine, 1960). Even if the learner can identify the referent of a word, they must generalize the word to new exemplars, with numerous available dimensions over which to generalize (color, shape, function, etc.). And, of course, they must ultimately do both of these things for tens of thousands of words.

Given the sheer magnitude of these problems, many theorists argue that word learning cannot be based on associative processes. Mechanisms such as constrained Bayesian inference (Xu & Tenenbaum, 2007), some form of social inference (Akhtar & Martinez-Sussman, 2007; Golinkoff & Hirsh-Pasek, 2006; Tomasello, 2001), or both may be needed (Frank, Goodman, & Tenenbaum, 2009). Other authors insist that word learning is largely a logical, symbolic problem (Halberda, 2006; Medina et al., 2011; Trueswell, Medina, Hafri, & Gleitman, 2013).

A rich discussion of these issues is underway. For example, empirical work is asking if complex social abilities may derive from attentional and/or associative factors (Akhtar, Carpenter, & Tomasello, 1996; Samuelson & Smith, 1998). Researchers are also asking if the way that words influence the formation of categories is best described by conceptual theorizing (Waxman & Gelman, 2009) or if labels simply guide attention (Robinson & Sloutsky, 2007). Further inquiry is exploring the learning mechanism itself, revealing that when we strip away many of the social or inferential cues, infants and adults can still master word-object mappings using only statistical co-occurrence (Smith & Yu, 2008; Yu & Smith, 2007), although it remains to be seen whether this type of learning by observation is associative or inferential in nature (McMurray, Zhao, Kucker, & Samuelson, 2013; Medina et al., 2011; Ramscar et al., 2013).

Similarly, computational work has explored the ability of simple associative systems to exhibit some of the hallmarks of more inferential processes, including: acceleration in the rate of learning (McMurray, 2007; Regier, 2005), abstraction of generalizable dimensions or selective attention (Colunga & Smith, 2005; Samuelson, 2002), and rapid inference of new names (McMurray, Horst, & Samuelson, 2012; Regier, 2005). There is also evidence that the componential structure of words can even emerge from simple, Rescorla–Wagner associative systems which

map letters or sounds directly to meaning (Baayen, Hendrix, & Ramscar, 2013; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010).

The combination of computational and empirical techniques can be particularly effective. For example Ramscar et al. (2013) used a computational instantiation of the Rescorla–Wagner learning rule (Rescorla & Wagner, 1972) to make predictions about the pattern of errors children and adults might make in a rapid word learning task. These researchers showed that children behaved in a way that more closely conforms to the predictions of that theoretical model than do adults, who appeared to adopt a more logical approach to inferring the name of a novel object.

Thus, although debate continues, considerable evidence is accumulating that emergentist associative accounts can explain many complex word learning phenomena and may thus underlie key aspects of word learning. Yet, all of these theoretical advances within emergentist associative accounts have relied either on researchers' intuitive predictions about what associative learning can do, on demonstrations that lower-level information (e.g., salience) can do the work of higher-level information (social cues), or on computational instantiations of associative learning. We know almost nothing about what *biological associative systems* can do in the context of problems similar to word learning. This lack of knowledge is crucial as animal models can strip away higher-level factors such as social pragmatics and use a pure reinforcement approach; they can ensure that the mapping between “word” and “meaning” is utterly arbitrary; they can arrange either successive or simultaneous schemes of programming the associations to be learned; and, they can use an organism that is unlikely to have the rational chops or evolved predilection to engage in hypothesis-driven inference. By understanding the nature of associative learning in a biological system under these circumstances, we may be able to focus and constrain theories positing the involvement (or absence) of associative learning in advanced cognitive processes like word learning.

Unfortunately, work in animal learning does not offer a characterization of biological association that is relevant to the problem of word learning, because work on categorization, attention, and the like has largely focused on paradigms in which animals must map a fixed number of stimuli to a small number of responses (Zentall, Wasserman, Lazareva, Thompson, & Rattermann, 2008). In contrast, word learning requires children to map from a very large number of possible objects or meanings to a large number of possible words. Although an associative model at this scale is impractical (it takes children 18 years hearing an estimated 17,000 words per day to learn such mappings), biological associative models that capture this many-to-many mapping are essential if we are to understand the biological basis of associative learning in a way that might apply to aspects of language acquisition.

Evidence from several celebrated animal “language” projects suggests that nonhumans may have the capacity for such learning, but it is not clear that purely associative mechanisms participated in the success of these projects. This limitation lessens their value as biological models of

associative learning. For instance, in Project Washoe and in several follow-up investigations, Gardner and Gardner (1984) trained four chimpanzees to perform up to 35 manual gestures in American Sign Language to refer to members of such diverse human language categories as balls, shoes, flowers, and cats. Similarly, an African grey parrot, Alex, learned to categorize about 50 different common objects (Pepperberg, 2002). More recently, two domesticated dogs have shown the ability to learn the names of several hundred different objects (Kaminski, Call, & Fischer, 2004; Pilley & Reid, 2011; though see, Griebel & Oller, 2012).

Whatever their other merits, these innovative projects do not effectively isolate the involvement of associative learning; thus, they do not offer a platform for understanding (or testing) the associative basis of word learning. In cases like Project Washoe, where gestural signs were taught, not all of the ASL signs were arbitrarily related to their referents; indeed, some of the signs had an iconic relationship to their referents, like “toothbrush” and “washcloth.” Perhaps more importantly, little attempt was made in this and other projects (e.g., Pepperberg, 2002) to rigorously control the associative structure of the learning environment of these organisms—they were taught via social interaction with humans or via observation of humans.

The solution, of course, is to strip the learning down even further: to teach an animal using the simplest possible stimulus–response–reinforcement paradigm. Studies with pigeons offer a basis for this approach. For example, Herrnstein and Loveland (1964) first showed that pigeons could sort photographs into 2 categories: those that contained a single kind of target object and those that did not. However, this noteworthy accomplishment fails to capture the many-to-many mappings that are of core interest here. Bhatt, Wasserman, Reynolds, and Knauss (1988) took a small, but important step in this direction, teaching pigeons to simultaneously sort several hundred photographs into 4 different categories. Still, 4 was a rather small number of categories, representing a many-to-few mapping.

Pilot work (reported in Wasserman, Brooks, Lazareva, & Miner, 2007) increased the number of training categories to 16 by introducing categories in a stepwise manner; pigeons were first taught to peck the correct pexigram for a single category, then a second was added, then a third, and so forth until the birds had mastered all 16. With each new category, a new report button and pexigram was introduced. Although pigeons were able to learn the task after extensive training, this 16-alternative forced-choice procedure suffers from two shortcomings. First, as the number of categories is increased, it becomes increasingly difficult for the animal to locate the added pexigram on its response panel. This creates methodological and statistical difficulties as the number of words grows larger.

Second, children are not typically exposed to words in the same strictly progressive fashion, adding a single word only after the previous one has reached criterion. Instead, children are often simultaneously exposed to a large number of words and must learn them in parallel. Despite this fact, we do not deny that children are introduced to new

words at important points in development (e.g., the onset of reading), and that the large number of words on which they may be currently working at any given time represents only a subset of their possible vocabularies. Thus, it is not possible to unequivocally claim that word learning proceeds in parallel or simultaneously.

Nevertheless, it seems highly unlikely that children are exposed to new words one at a time, and only after mastering the previous word. Thus, although we are not making strong claims about whether natural word learning is better captured by a simultaneous or sequential training paradigm, it seems clear that some element of simultaneous learning is needed in an animal model. Moreover, whether items are trained in blocks or interleaved can have important effects on learning (Battig, 1972; Carlson, Sullivan, & Schneider, 1989; McCloskey & Cohen, 1989; Mirman & Spivey, 2001; Perrachione, Lee, Ha, & Wong, 2011; Wulf & Shea, 2002). Although interleaved training is generally deemed to be more robust, sequential training may be easier (Perrachione et al., 2011). Therefore, as a first step in developing an animal model that is relevant to human word learning, it was important to capture the somewhat more natural (and challenging) case of learning multiple items simultaneously.

Thus, we sought to teach pigeons a much larger set of associations than in prior experiments, and to do so with a simultaneous training paradigm. Our experimental approach offers the possibility of explicit control of the training paradigm to ensure that the learning is associative in nature. Such a biological model of associative learning can effectively complement computational models (Mayor & Plunkett, 2010; McMurray et al., 2012; Ramscar et al., 2013; Yu & Smith, 2012) by probing the limits and nonlinearities of association learning, as well as by modeling the unexpected richness of such learning in a biological system.

In this regard, a critical focus of our project was to examine the time course of many-to-many learning on a trial-by-trial basis to gain insight into the factors that shape the acquisition of complex stimulus–response mappings. Our focus is highly relevant to recent debates about the role of associative processes in human statistical word learning paradigms. Thus, we use our animal model to add to this debate by illustrating how an animal model may help answer outstanding questions in human learning.

1.2. Learning by observation

Whereas the problem of referential ambiguity was long seen as an obstacle to unconstrained associative learning, a recent proposal by Yu, Smith, and colleagues offers an important alternative (Smith & Yu, 2008; Yu & Smith, 2007, 2012; see also, Siskind, 1996). They point out that, although the mapping between a novel word and its referent may be ambiguous in any one situation, words are more likely to co-occur with their referents than other objects. Consequently, if learners gradually acquire co-occurrence statistics between words and objects across situations, then the correct mappings will rise to the top. Such acquisition can easily be accomplished by associative mechanisms that gradually build links between words and

objects on the basis of co-occurrence (McMurray et al., 2012, 2013; Yu & Smith, 2012). Furthermore, such learning does not require feedback; purely unsupervised mechanisms can accumulate such statistics, thereby allowing children to learn a substantial portion of their vocabulary observationally. Even without overt feedback, however, it is possible that the underlying learning mechanism has access to an error signal, with learners predicting available objects from the heard word (or words from available objects) and evaluating these predictions to guide learning (Ramscar et al., 2010, 2013).

Recently, Medina, Trueswell, and colleagues (Medina et al., 2011; Trueswell et al., 2013) have pointed out that learners could also approach cross-situational learning paradigms with a more inferential strategy. When encountering a novel word, learners may “propose” a single referent or mapping for it. On subsequent trials, if this proposal is verified (e.g., the same object is present for this word), then the mapping is retained; if it cannot be verified (that object is not present), then a new proposal is made. This hypothesis was initially supported by data showing that if people arrive at the correct mapping for a referent earlier in training (e.g., during a more informative encounter), then they ultimately perform better than if that more informative encounter came later (because they would not spend as long revising and re-guessing). Medina et al. argued that this result was inconsistent with statistical or associative accounts, as the statistics by the end of training would be the same no matter when the highly informative event had occurred, thereby yielding identical performance.

Trueswell et al. (2013) further argued that, in statistical learning accounts, even if a learner had answered incorrectly on a prior trial, they should still reap the benefits of that earlier exposure as they still accumulated useful co-occurrence statistics, and thus show an improvement in performance. By contrast, in a propose-but-verify scheme, learners do not carry forward anything from previous trials about a word except its current proposal. So, in this case, learners should respond at chance if they answered incorrectly on a prior trial, because they would still be entertaining the wrong proposal. This notion was tested with a form of auto-correlation analysis, in which each trial's performance was predicted by the learner's accuracy on their last encounter with that word. This analysis supported the propose-but-verify account, with near-chance performance if the learner was incorrect on their last encounter with that word, but with substantially above-chance performance if they had been correct.

There have been a number of theoretical and empirical critiques of the propose-but-verify account in the word learning literature (Dautriche & Chemla, 2014; McMurray et al., 2013; Yurovsky, Fricker, Yu, & Smith, 2013) as well as findings that are difficult to reconcile with this interpretation (Ramscar et al., 2013; Vouloumanos, 2008). However, for the present purposes, it is important to point out several problems with the interpretation of associative learning to which propose-but-verify was offered as an alternative.

First, with regard to the timing of co-occurrence information, it has long been known that associative learning consists of more than co-occurrence counting or

correlation detection; precisely when information arrives can have a powerful influence on behavior (Rescorla, 1988). This fact is clearly illustrated by phenomena like forward and backward blocking (e.g., Wasserman & Berglan, 1998). Nonetheless, such order effects are rarely tested in complex associative paradigms that may be more analogous to word learning (e.g., Katagiri, Kao, Simon, Castro, & Wasserman, 2007). A richer animal model would be useful in generalizing these order-of-presentation effects to paradigms more closely resembling word learning.

Second, the auto-correlational analyses of Trueswell et al. (2013; and see also, Dautriche & Chemla, 2014) make the critical assumption that, in associative systems, the discrete choice behavior is not relevant—only the accumulating statistics are. This assumption is unlikely to be true. In supervised associative paradigms, the choice response and the reinforcer are highly relevant, and may importantly shape behavior. Moreover, even in unsupervised learning, it is not unreasonable to assume that learners allocate more attention to whichever referent is chosen, thereby building stronger connections between a word and that object (and see, McMurray et al., 2013, for a computational model). Thus, at least some aspect of choice behavior should influence associative learning, but it is not clear what form this influence should take. Here, again, an animal model—particularly one involving a well-controlled laboratory paradigm—could provide a crucial baseline by revealing if and how prior responding can influence the behavior of an associative learner.

Finally, associative learning is widely considered to be much richer than simple S-R mapping (Wasserman & Miller, 1997); factors like expectation (Ramscar et al., 2010; Rescorla, 1988), perceptual similarity (Soto & Wasserman, 2010), attention (Livesey & McLaren, 2011; McMurray et al., 2012), and learning sets or error factors (Harlow, 1959) all play a role. Thus, over and above the effect of prior choice behavior, one might expect to see the influence of a variety of factors such as the amount of experience with the object, its appearance as a foil on prior trials, and/or its spatial location. That is, a more fully-fledged associative account may predict a gradual learning effect, a prior choice effect, as well as other important behavioral effects. These possibilities were not examined by Trueswell et al.; and, it is entirely unclear how they may play out in a many-to-many animal model. Thus, we examine these possibilities in our animal model in order to lay the groundwork for more sophisticated analyses of human learning.

1.3. The present study

The present study sought to develop an associative learning task for pigeons that requires a mapping between many visual exemplars and many categories, a critical property of human word learning. We do not claim that we are attempting to teach words to pigeons. However, the foregoing review reveals a crucial gap in the literature on animal learning; namely, the absence of conditioning tasks in which many stimuli must be simultaneously mapped to many categories with no opportunities for social

interaction, imitation, etc. Using this model, we asked whether such learning is possible and whether it can lead to coherent categorization and generalization. We then investigated the trial-by-trial influences on learning to see if such a model can provide any useful insight into aspects of human word learning that may involve association formation. Specifically, we asked if a biological associative system is capable of divulging the influence of previous discrete choices and if this influence takes a similar or a different form from the results of Trueswell et al. (2013).

We trained 3 pigeons to name 16 categories of objects—*baby, bottle, cake, car, cracker, dog, duck, fish, flower, hat, key, pen, phone, plane, shoe, tree*—by pecking 16 arbitrary visual icons, or *pexigrams* (similar to the tokens, termed *lexigrams*, used in prior experiments with primates to stand for categories of objects; Rumbaugh, 1977; Savage-Rumbaugh, 2009). Each category contained 8 black-and-white photographs (128 total images). Although this feat would certainly be a far cry from human vocabulary learning capacity, it represents an exponential increase in associative complexity. In typical 2AFC associative learning tasks, there are four potential associations to enhance or eliminate (2 stimuli \times 2 responses). Here, if we consider each image, then there are 2048 (128 stimuli \times 16 responses) potential associations; even if we assume the individual category representations (rather than the individual exemplars) as the basis of association, there are 256 (16 categories \times 16 responses) potential associations. Thus, although not at the same scale as human word learning, such a task clearly represents a large, but tractable step toward understanding associative learning in a many-to-many framework.

It was important overcome the necessity of progressively adding categories/responses when training with so many categories (Wasserman et al., 2007) to achieve a simultaneous training paradigm. To do so, we borrowed an approach from human word learning, in which children and adults are simultaneously trained on *all* of the items, but on any given trial they receive only a *subset* of the available words as responses (Creel, Aslin, & Tanenhaus, 2006; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Yu & Smith, 2007). This experimental tactic simplifies the response options on each trial while permitting a large number of categories to be trained from the outset of training. Thus, the present method of simultaneously training a large number of categories using a simple 2AFC task represents the culmination of several developmental steps toward a realistic, yet tractable way to experimentally explore the associative substrates of natural categorization.

In our experiment, images of the items to be categorized appeared in center of a small screen. On either side of the training image were two report buttons depicting two colored pexigrams. One pexigram corresponded to the category of the displayed stimulus; the second was a foil corresponding to a different, randomly selected, category. Correct choices delivered food reinforcement; incorrect choices did not deliver food and required the pigeons to complete correction trials until the correct response was performed. Because generalization is the key signpost of concept learning, we tested the breadth of the pigeons' categorization behavior by showing them

4 novel exemplars from each of the 16 categories in later testing sessions.

Given our rich dataset of trial-by-trial responding, we next conducted an auto-correlation analysis similar to Trueswell et al.'s analysis of observational learning. Critically, we examined the effect of the pigeon's prior choice response (as did Trueswell et al. and Dautriche and Chemla). Here, our use of a supervised learning paradigm represents a departure from the analogous unsupervised learning work. This move was necessary, as it is quite difficult to make animals respond without an explicit reinforcer. Thus, we must be cautious in interpreting any differences we find in trial-by-trial behavior between supervised learning in animals and unsupervised learning in people. Nevertheless, to the extent that we obtain empirical similarities, it suggests a set of underlying regularities in the course of associative learning.

1.4. Goals and limitations of our animal model

Having described our basic paradigm, it is important to clarify just what it is trying to achieve. As with any experiment or computational model, an animal model must choose to ignore certain aspects of the problem to permit deeper study of others. Although our work is clearly motivated by issues in human word learning, our experiment is not setting out to capture the full problem. Our primary goal is rather to develop a biological model of the kind of associative learning that might be involved in word learning. There is currently no appropriate biological model which captures the essential property of simultaneously learning relationships between many stimuli and many responses, and which tests this learning in a purely associative framework. Only by understanding how such associative learning actually works (in a real organism) can we evaluate what role it might play in word learning.

That said, there are a number of things we are not (yet) attempting to capture. Word learning surely requires children to deal with referential ambiguity (the large set of possible referents for a new word). Although there is some ambiguity present in our paradigm (there are two available response options), there is also feedback as to which is correct. Of course, this feedback may not be so different from that given in common word learning situations in which children receive social cues pointing them the correct referent or verbally responding to a misnaming. Words must also support generalization. Although we are testing for generalization, it is all within a single "level" of categorization (at least to humans), the basic level. In contrast, children must flexibly generalize between different levels of categories. This should not be a critical concern, however, as prior work has shown that pigeons too can flexibly categorize the same stimuli at both basic and superordinate levels (Lazareva, Freiburger, & Wasserman, 2004). Finally, word learning requires children to extract a complex auditory word-form out of a time-varying signal and to map it to a novel complex visual object. We are necessarily simplifying these perceptual processes. Although the visual objects under investigation are complex (and novel) to the pigeons, we are in no way attempting to model the complex auditory processes.

Most importantly, our training uses an explicit reinforcement signal. Such explicit feedback is not typical of children's word learning studies, although it is common in studies of adults (e.g., Magnuson et al., 2003). But, the more important question is whether an explicit reinforcement signal is typical in a child's learning environment. Although many researchers operate under the intuition that it is not, there may be a number of important avenues for feedback. There is research suggesting that feedback is common after children name things (Bohannon & Laura, 1988; Chouinard & Clark, 2003; O'Hanlon & Roberson, 2007). Moreover, as we described, social cues like eye-gaze or ostensive naming may offer a sort of implicit supervisory signal. Finally, even observational learning may use a supervised learning rule (e.g., the Rescorla–Wagner rule) in the form of evaluating predictions (Ramscar et al., 2013). Thus, there is quite a bit of value in modeling this form of learning in animals – even if the mechanism of reward differs, the basic properties of learning may not.

To be fair, there are no human studies (or computational models) in which children learn many words in parallel, map them to completely novel objects, generalize responding at multiple levels, and do so purely via observation. Child studies are usually motivated to study one part of this complex system. Our work is meant to isolate the type of associative learning that would be needed for human word learning, to understand its capacity to support large numbers of associations, to assess its ability to generalize responding to novel stimuli, and to characterize the trial-by-trial influences on the learning process itself. None of these ambitious aims could be accomplished with prior animal models. We hoped that achieving these goals would yield a clearer picture of whether associative learning may be involved in at least some key aspects of human word learning.

2. Method

2.1. Subjects

Three pigeons (*Columba livia*) with various unrelated experimental histories participated. They were kept at 85% of their free-feeding weights and had ad lib access to water and grit.

2.2. Apparatus

The pigeons were trained in three conditioning boxes (Gibson, Wasserman, Frei, & Miller, 2004). The experimental stimuli were presented on an LCD monitor behind a transparent, resistive touchscreen. There were three critical areas of the touchscreen: a 5.4×5.4 cm central "display" area in which the photographic target stimuli appeared, plus two square "report" areas (2.1×2.1 cm) in which the pexigrams appeared placed 3.3 cm to the left and right of the central display area. The report buttons were aligned with the bottom of the picture, 20.0 cm above the wire mesh floor. A rotary food dispenser delivered 45-mg pellets into a Plexiglas cup on the rear wall. A houselight provided illumination during sessions.

2.3. Stimuli

The photographic stimuli came from 16 categories, each comprising 12 images (8 were used in training; 4 were used in testing). These images were taken from internet and from other databases and were resized to 185×185 pixels at a resolution of 300 dpi, converted to grayscale, and had their backgrounds removed by Knockout (Corel Corporation, Ottawa, ON, Canada). The full set of images can be seen in the [online Supplement \(S1\)](#); representative stimuli are shown in [Fig. 1A](#). The training and testing stimuli were the same for each pigeon. In addition, we used 16 highly distinctive visual patterns as pexigrams. These patterns were created by superimposing a colored shape outline onto a multicolored background (see [Fig. 1B](#) for examples); these images were then resized to 72×72 pixels at a resolution of 72 pixels per inch. The specific pexigrams that were assigned to each category were randomized for each bird.

2.4. Procedure

2.4.1. Acquisition

On each training day, pigeons were presented with 128 randomly ordered trials, 8 from each of the 16 categories. Within the 8 trials from each category, the specific exemplar was randomly chosen with replacement from the total pool of 8 exemplars. Although not all of the exemplars of a given category were shown each day (and some were sometimes shown more than once), they were shown on average the same number of times across consecutive training days. Of the two available report button locations, the left–right location of the correct pexigram was randomly chosen on each trial; a foil was randomly chosen as the alternative pexigram in the other report button location. Because of the large number of possible category–foil

combinations, the pairing of particular foil pexigrams with target categories was also randomized on each trial.

Each trial started with the presentation of a white square ($6 \text{ cm} \times 6 \text{ cm}$) containing a central black cross. Following a single peck to the square, a photograph was presented in the same location until the bird completed a fixed number of pecks on the photograph (different for each bird and ranging between 15 and 20). The bird then reported the category of this picture by pecking one of the two pexigrams presented on that trial. Correct choices produced 1–2 food pellets. Following a correct choice, there was an 8-s inter-trial interval before the next stimulus. Incorrect choices were punished with a dark timeout (ranging from 15 s to 25 s), followed by one or more correction trials; these correction trials repeated the trial photograph and the two response pexigrams until the pigeon executed the correct response.

2.4.2. Generalization

After accuracy reached a plateau for each bird (about 383 sessions of training or about 45,000 trials), generalization testing was begun. There were 10 four-session blocks of testing. One novel stimulus per category was shown each session; this method amounted to an additional 16 trials for a total of 144 daily trials. It thus took 40 days to complete the testing regimen, permitting each of the 64 testing stimuli to be seen 10 times. Food was given following whichever choice response the pigeon made to the generalization testing stimuli; no correction trials were given on novel stimulus trials.

3. Results

3.1. Acquisition

In some of the early training sessions, responding was occasionally a bit erratic; consequently the three birds



Fig. 1. (A) Three exemplars each from the dog and shoe categories, shown to the pigeons in black-and-white. (B) Four examples of pexigrams, shown to the pigeons in color. (For full color versions of these images, the reader is referred to the web version of this article.)

received slightly different numbers of total training trials (39Y: 45,988; 45W: 48,174; 66W: 44,768). Thus, to explore the course of learning, we examined only the first 44,700 trials (we will consider testing performance in the next section). Training trials were grouped into 15 equal-sized blocks of 2980 trials, approximately 186 trials from each of the 16 categories. The birds' accuracy during training is shown in Fig. 2A. Each of the pigeons showed clear evidence of learning, although they reached different levels of overall accuracy: Bird 45W performed best, reaching accuracy levels exceeding 80% correct; Bird 39Y reached nearly 70% correct; and, Bird 66W reached an accuracy level of 65% correct.

Because there were so many training categories, strong discrimination performance on this task could have been due to better than chance performance in classifying all of the categories, or to exceptional performance in classifying only a few of the categories. To further examine this issue, for each bird, we conducted a (one-tailed) binomial test against chance for each category in each block of training. Fig. 2B depicts the number of “learned” categories. The general trends closely mirror the overall acquisition functions for each bird (Fig. 2A); more importantly, these functions show that two birds did attain the maximum score of 16 by Block 7 and that the third bird attained a maximum score of 14 categories.

Thus, all three birds learned something about virtually all of the categories and were not merely masters of only a few of them (we examine this issue more closely below).

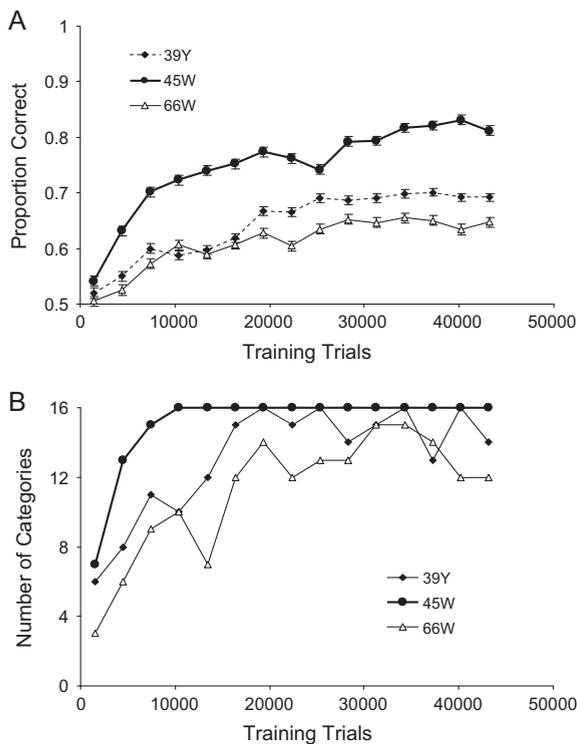


Fig. 2. (A) Overall accuracy as a function of training trials (binned into blocks of 2980 trials). (B) Number of categories significantly greater than chance.

The relatively modest accuracy of Birds 39Y and 66W may therefore somewhat underestimate their learning about the large number of training items (more on the matter of capacity follows).

3.2. Categorizing or memorizing?

The next question we addressed is whether the pigeons acquired this task as a categorization problem (e.g., learning 16 open-ended categories) or as a rote memorization problem (e.g., learning 128 individual photographs; Fagot & Cook, 2006). We addressed this issue in two ways. First, we assessed the cohesiveness of the categories; namely, whether responding was more similar to exemplars within the same category than to exemplars from other categories. Of course, it is possible that items are not learned as a category, but simply that items within a category are individually equally easy or hard to learn. Thus, we also examined generalization to untrained exemplars, to assess the breadth of original learning and to offer converging evidence for categorization. Both analyses examined the 5760 trials during the testing sessions: 5120 trials of the trained exemplars (128 stimuli over 40 sessions and differentially reinforced) plus 640 trials of the new exemplars (64 stimuli, 10 repetitions each, non-differentially reinforced).

3.2.1. Category cohesion

In the interest of cognitive economy, pigeons should learn this task as one involving a set of 16 coherent categories, each of which is associated with a distinctive pexigram, that is, as a categorization problem. Less parsimoniously, pigeons may learn 128 individual stimulus-response associations, that is, as a memorization problem. One experimental method that has been deployed to ascertain whether animals learn coherent categories has been to teach them either a true category task or a pseudocategory task constructed from randomly selected items that do not perceptually cohere. True category tasks are typically acquired far faster than pseudocategory tasks (Soto & Wasserman, 2010).

Although the present experiment did not explicitly train pseudocategories, the large number of true categories that the pigeons were taught offers the opportunity to apply an analysis entailing similar logic: if pigeons are learning the task on the basis of coherent categories, then the accuracy to any given exemplar should be more similar to other exemplars from that category than to a random combination of stimuli from the other categories. So, during acquisition, the variance of accuracy within a true category should be lower than the variance of accuracy within a randomly constructed pseudocategory.

There were 16 true categories of 8 exemplars. Thus, we randomly assigned items from different true categories to each pseudocategory. For example, one pseudocategory might contain: *Duck 1, Dog 5, Baby 3, Flower 7, Tree 2, Cake 1, Key 8, and Shoe 2*. This random assignment was constrained such that no two exemplars in a pseudocategory came from the same true category. We next computed the average proportion correct for each individual exemplar. Finally, we computed the standard deviation across

the 8 exemplars for each true category (for each bird) and for each pseudocategory. If true categories cohere, then the average SD within a true category should be less than the average SD within the pseudocategories. Because the pseudocategories were randomly constructed, we used a Monte-Carlo analysis to establish statistical significance, repeating this procedure for 1000 sets of pseudocategories, to ask how likely by chance it would be to obtain a SD as low as was observed for the true categories.

Fig. 3 shows the results. In each panel, the distribution of within-category SDs across all 1000 sets of pseudocategories is shown as a histogram. The corresponding SD for the true category is the dark vertical bar. For all three birds, the true category SD was significantly lower than the SD of the pseudocategories. In fact, there was not a single run of the Monte-Carlo analysis that yielded a pseudocategory SD which was less than that of the true categories, yielding a likelihood of achieving such within-category coherence by chance of $p < .001$ for each bird.

3.2.2. Generalization

The prior analysis suggests that our pigeons' discrimination behavior reflects the coherence of the training categories. But, equally important for establishing categorization is the issue of whether the categorization behavior generalizes to new, untrained exemplars. During the 40 testing sessions, in addition to the 8 familiar stimuli,

there were 4 extra transfer exemplars for each category presented 10 times each.

Fig. 4 plots accuracy across all 16 categories on training and generalization trials. Each bird behaved discriminatively to both kinds of stimuli, although they responded a bit more accurately on training stimuli; there was a marginally significant difference between training and generalization stimuli for each bird (39Y: $T(15) = 2.06$, $p = .057$; 45W: $T(15) = 2.07$, $p = .056$; 66W: $T(15) = 1.90$, $p = .081$). Still, binomial tests (across all 16 categories) showed that, for each bird, generalization performance was significantly greater than chance (39Y: $Z = 7.50$, $p < .001$; 45W: $Z = 16.10$, $p < .001$; 66W: $Z = 6.50$, $p < .001$). Thus, even when trained with a large number of categories, the pigeons were able to transfer discriminative responding to novel stimuli with only a modest decrement in accuracy.

3.3. Categorization capacity and distribution of responses

The preceding analyses suggest that the three pigeons performed this task as one of learning categories, not individual exemplars. Given the emphasis of this biological associative model on the many-to-many nature of the learning problem, we next turned to the question: how many categories and/or exemplars did the pigeons actually learn?

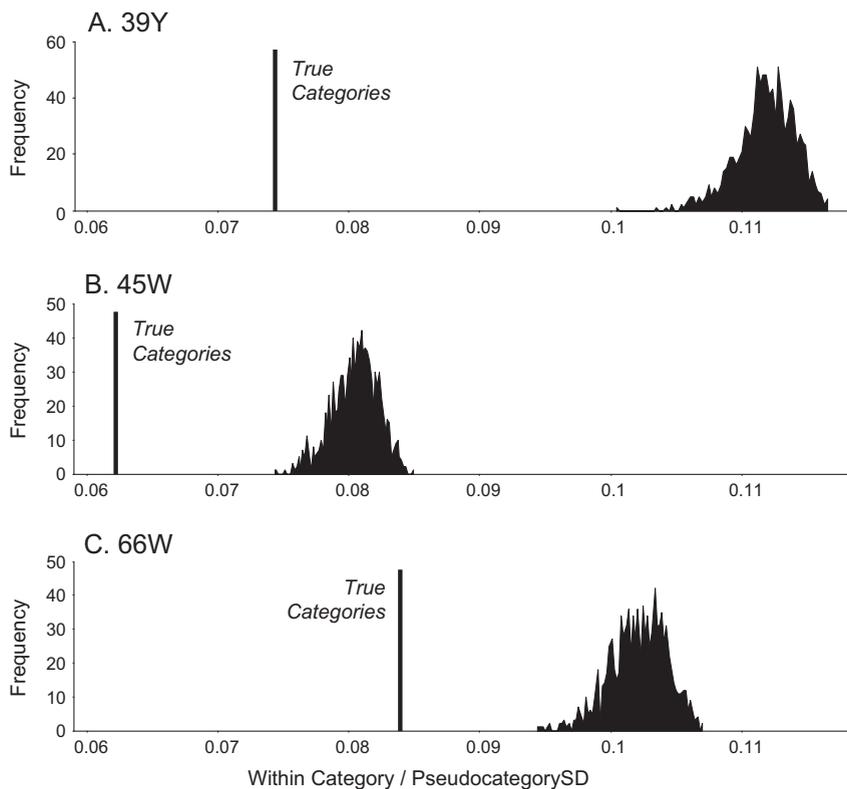


Fig. 3. Distributions of within-pseudocategory SDs (of accuracy) for 1000 runs of the Monte-Carlo simulations, plotted separately for each bird. The actual within-true category SD for that bird is shown as the single black line and was for all three pigeons substantially lower than the SD computed for any of the 1000 sets of pseudocategories.

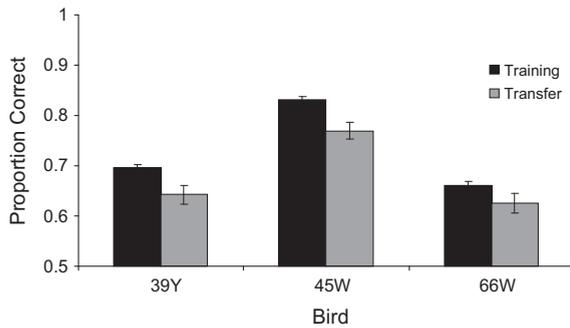


Fig. 4. Mean accuracy for trained and transfer stimuli in test.

We started by re-examining responding on training trials. Despite the birds' generally accurate responding during training, we did not find clear evidence that all three birds had learned all 16 categories (Fig. 2B), perhaps because of the small number of trials in each block (about 186 trials for each category for each bird). To increase power, we combined the last 40 training sessions with the 40 testing sessions (excluding generalization trials), yielding an average of 628 trials for each of the 16 categories for each bird. We then computed another binomial test against chance for each category for each pigeon. We found clear evidence of each bird having learned all 16 categories (Table 1 summarizes the binomial tests; means in Fig. 5).

We next did the same for each of the individual training exemplars. Again, we computed, for each pigeon, how many of the 128 exemplars were learned, also using a binomial test against chance, with an average of 78.5 trials for each exemplar. Overall, across the 384 statistical comparisons (3 birds \times 128 exemplars), 323 (or 84%) were significantly above chance. However, Table 2 suggests that inferential statistics may undersell this strong level of discriminative performance, because individual comparisons had rather low power. Here, Bird 45W continued to be the clear star, significantly learning 127 of the 128 exemplars. But, even for the other two birds, there were only 5 exemplars (each) which exhibited no evidence of learning – the remaining exemplars were marginally significant or at least above chance.

So, what accounts for the few exemplars that did not seem to have been learned? One possibility is that the birds have a limit on how many exemplars they can learn (either within or across categories). An alternative is that categories are learned as a whole and individual exemplar performance varies around some category mean. To

distinguish between these alternatives, we conducted an extensive analysis of the distribution of responding across exemplars. The idea was that a capacity-limited model should show a *bimodal* distribution with a cluster of exemplars around chance and a second cluster well above chance. We found no evidence for such a bimodal distribution (Supplement S2). The clearly *unimodal* response distributions strongly suggested that variations in behavior across exemplars simply reflect natural variation around the mean and do not represent an associative capacity limit.

3.4. Trial-by-trial analysis

The goal of the final analyses was to carefully document trial-by-trial dependencies in performance to gain insight into the learning mechanisms that underlie pigeons' ability to acquire complex many-to-many mappings. At the broadest level, this analysis serves to illustrate what an appropriately framed biological model of associative learning can reveal about an analogous problem in human word learning. Moreover, just as computational models of associative learning can often reveal unexpected emergent phenomena to arise out of simple learning principles (Mayor & Plunkett, 2010; McMurray et al., 2012; Ramscar et al., 2013; Regier, 2005), so a biological model can (and should) reveal similar emergent associative complexity.

Such a model may have important implications for human learning. Trueswell et al. (2013) posited that, in gradual or associative learning systems, performance on a given trial should be more strongly influenced by the amount of experience with that item, not by the learner's choice behavior on the previous encounter with the item. In contrast, in a propose-but-verify scheme, if the learner chose incorrectly on a prior encounter with that item, then they must have been entertaining the wrong proposal and should respond at chance level on the next trial. In other words, the experience of having just seen that object—even if the response to it was incorrect—should not improve the learner's current performance. However, as we described earlier, these assumptions (and see, Dautriche & Chemla, 2014) are not based on any actual associative system (either computational or biological), so an examination of pigeons' learning should reveal something novel concerning what to expect from a real associative system.

Trueswell et al. analyzed their data using a type of auto-correlation approach, predicting the response on each trial from performance the last time that word was heard. We began by replicating this analysis. For descriptive purposes, we used the same 2980-trial blocks depicted in Fig. 2. To start, we examined only the first block of training trials (~186 repetitions of each category), as Trueswell et al. only included a single block of 60 trials (5 repetitions of each word). Fig. 6A shows performance for each bird during this first block as a function of the bird's accuracy the last time it had seen that target category. A clear effect of prior responding can be seen. If the bird responded incorrectly on the previous trial with that category, then it performed at chance; however, if the bird responded correctly on the previous trial with that category, then it was more likely to respond correctly.

Table 1

Summary of binomial tests for each of the 16 categories for each of the three pigeons.

	Bird		
	39Y	45W	66W
Z: Range	2.2–20.3	15.4–36.3	2.2–17.8
Z: Average	15.4	23.2	8.5
P: Max	.014	<.001	.013
P: Average	<.001	<.001	<.001

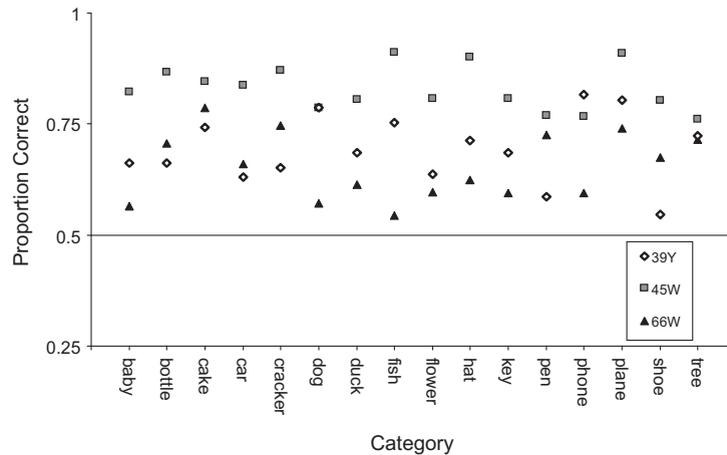


Fig. 5. Accuracy for each category for each pigeon.

Table 2

Performance (number of exemplars) across the 128 exemplars for each bird.

	Bird		
	39Y	45W	66W
Significant	104	127	92
Marginal	8	0	6
>Chance	11	1	25
Not learned	5	0	5

These findings were assessed with a logistic mixed-effects model, in which prior accuracy was the only fixed effect (coded 0/1) and the dependent variable was current trial accuracy. The random effects included bird (with a random slope of prior accuracy), the target category (intercept), and the foil pexigram, with only the first 2980 trials analyzed. Analyses were conducted in a linear mixed-models framework using the LME4 package (Bates & Sarkar, 2011) of R. By using random slopes for each bird, we were able to combine the data from all three pigeons to increase the power of this analysis while still accounting for the fact that each of the three birds learned at different rates and to different asymptotic levels of accuracy. By modeling individual learning curves in this way, it enables us to overcome some of the concerns raised by Gallistel, Fairhurst, and Balsam (2004), who suggested that averaging performance across birds could make learning look more gradual (associative) than it really is.

As expected, the intercept was not significantly different from 0 ($B = -.045$, $SE = .071$, $z = -.63$, $p = .53$); as pigeons' prior choice was dummy coded, this finding indicates that, when a bird was incorrect on the prior trial with that category, choice responding on the current trial was at chance, just like the humans in Trueswell et al. (2013). However, there was a significant effect of prior responding ($B = .258$, $SE = .053$, $z = 4.87$, $p < .0001$), indicating above chance responding when a bird had been correct on the prior trial with that category.

But, what happens with additional training? The effect of extended training was not assessed in the previously

published word-learning experiments, as there were only five repetitions of each stimulus. Consequently, such experiments could have missed a more gradual form of statistical or associative learning. Fig. 6B shows average accuracy across birds on each trial block as a function of prior choice accuracy. Here, we see very clear evidence for gradual learning on trials where the previous response was correct as well as on trials where the previous response was incorrect (see Roembke & McMurray, submitted for publication, for analogous findings with unsupervised human learning). Thus, we expanded our analysis to include the full training set. This model replicated the main effect of prior responding ($B = .219$, $SE = .049$, $z = 4.46$, $p < .0001$); however, now the intercept was significantly different from chance ($B = .582$, $SE = .14$, $z = 4.14$, $p < .0001$), suggesting that performance was above chance after prior incorrect trials.

To assess any gradual learning effect *over and above* the prior-choice effect, we next added the log of the number of repetitions of each category (how often the bird had been exposed to that category, what we termed the *gradual learning* effect) and its interaction with the prior choice response to the model (also included as random slopes on bird). These effects yielded a significantly better model fit ($\chi^2(9) = 2062.4$, $p < .0001$), suggesting that over and above the prior-choice effect (propose-but-verify), a strong effect of gradual learning was seen (Fig. 6B). In this model, prior-choice was still significant ($B = .138$, $SE = .024$, $z = 5.80$, $p < .0001$) and in the same direction as the prior models. Crucially, the gradual learning effect was also significant ($B = .284$, $SE = .042$, $z = 6.50$, $p < .0001$), as was its interaction with prior choice ($B = -.036$, $SE = .014$, $z = -2.47$, $p = .0136$). This interaction suggests that the effect of prior choice behavior is somewhat attenuated at later points in training (although difficult to discern in Fig. 6B).

It is intriguing that the prior-choice effect is still significant even when controlling for position in the learning curve. These factors were confounded in the Trueswell et al. (and the Dautriche and Chemla) analyses; by collapsing all of the data across the learning curve, trials on which

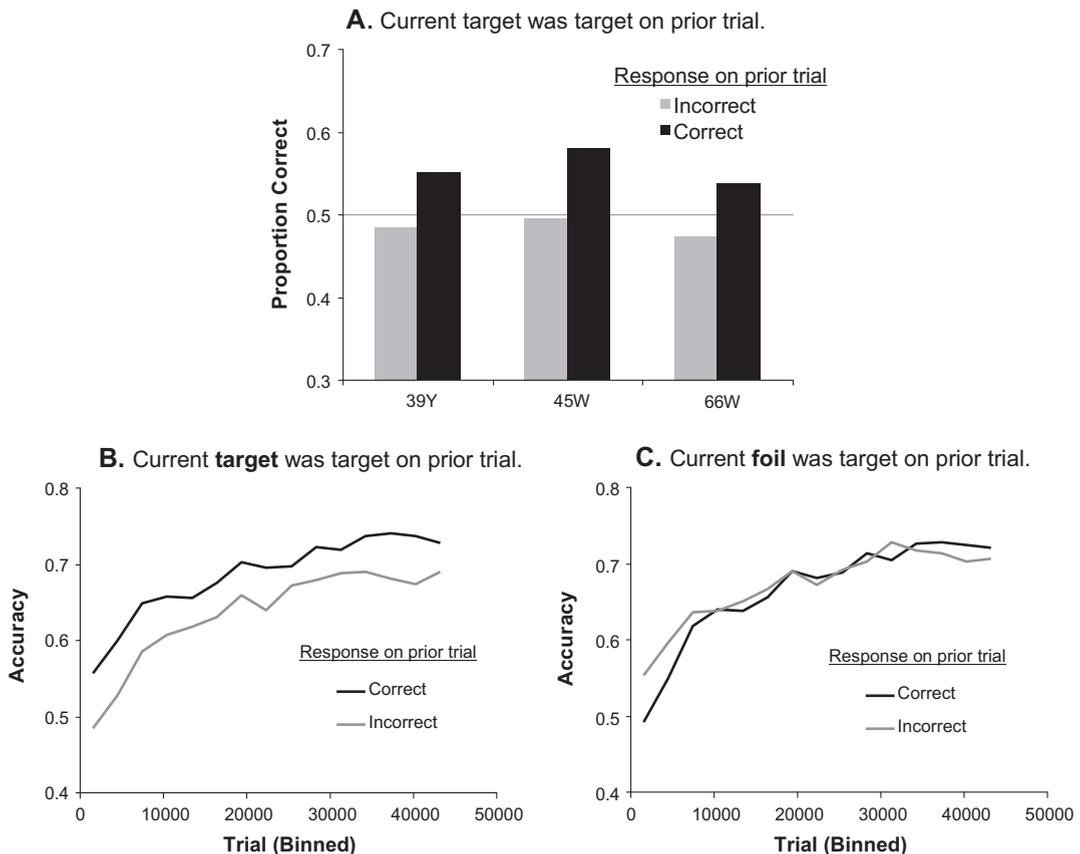


Fig. 6. Accuracy on the current trial as a function of responding to the target pexigram on a prior trial. (A) The first 2980 trials only for each bird; (B) mean responding across the three birds over the course of training when the target pexigram on the current trial was the target pexigram on a prior trial; (C) mean responding across the three birds over the course of training when the foil pexigram on the current trial was the target pexigram on a prior trial.

the previous choice was incorrect were more likely to have come from the beginning of training, whereas trials on which the previous choice was correct were likely to have come from the end. Thus, by only examining prior choice, Trueswell et al. could have mistaken an effect of prior choice for one of position in the learning curve (a gradual learning effect). Because both terms were included in our analysis, we have more confidence that the prior-choice effect is real.

In our data set, it is unlikely that the effect of prior choice behavior is a unique marker of a propose-but-verify learning strategy, as pigeons are not likely to engage in such inferential strategies; indeed, a closer inspection of our data confirms this suspicion. By the third block of trials (trial number 7450), the birds had responded correctly an average of about 40 times to each exemplar (39Y: $M = 38.9$ correct responses, minimum = 18 across 128 exemplars; 45W: $M = 43.7$, minimum = 26; 66W: $M = 37.3$, minimum = 20). Although all three birds had responded correctly to each individual exemplar many times before, all three birds were still performing well below their peak performance (Fig. 2). Indeed, after this third block, the birds would go on to respond incorrectly to those same exemplars about 80 times each (39Y:

$M = 96.6$, maximum = 155 across the 128 exemplars; 45W: $M = 65.9$, maximum = 117; 66W: $M = 103.4$, maximum = 143). Thus, it is more likely that this prior-choice effect derives from short-term memory of prior behaviors or prior stimuli, which interacts with slower-building long-term associations to affect behavior on any given trial. Although it is important to note that the availability of an explicit error signal in our paradigm limits a direct comparison to results such as those of Trueswell et al. (2013), what these results illustrate is that the strong effect of prior choice observed by such studies does not *uniquely* derive from logical inference models like propose-but-verify, and are consistent with at least some forms of biological associative learning.

Beyond the factors we have identified, there may be substantially more information in the trial-by-trial data than has been considered thus far. For example, does the distance between prior exposure to the target category and the present trial matter? Is this temporal distance effect modified by the prior choice? Other relevant factors may include the nature of the foil (and whether the bird correctly chose it the last time it was seen) and the spatial locations of the target and foil (relative to where they were last seen). Although our examination of such factors is

exploratory, it may more precisely characterize the learning curve and set the stage for richer investigations of the time course of human learning.

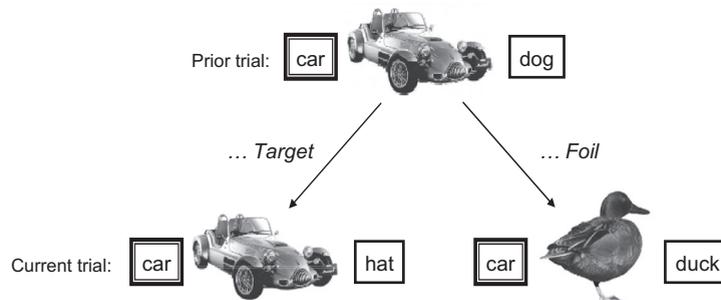
To investigate these issues, we ran a series of models based on the previously described model, but adding one or more factors. Although our analyses have thus far focused on repetitions of the same type of stimulus as the target across trials, there are actually four possible ways in which information can carry over from prior trials (Fig. 7). The *target* pexigram on the current trial could have served as the target pexigram on a prior trial (Fig. 7A, left path; the analyses reported thus far) or the *foil* pexigram on the current trial could have served as the target pexigram on a prior trial (right path). Similarly, the *foil* pexigram on the current trial could have served as the foil pexigram on the prior trial (Fig. 7B, left path) or the *target* pexigram on the current trial could have served as the foil pexigram on a prior trial. Separate analyses were conducted for each of these factors including one factor (dummy coded) along with the number of times the stimulus had been seen and their interaction. For each of the four factors, we also examined (in separate analyses) the number of intervening trials (coded as a z-score) and whether that pexigram had appeared in the same or in a different spatial location on prior trials (± 0.5), as well as their interactions with the other factors.

A complete description of all eight statistical models is outside of the scope of this paper, but several findings are noteworthy. First, of the four effects of prior trial

responding, only two were significant. If the pigeons responded correctly the last time the current target pexigram appeared as the target pexigram, then they were more likely to respond correctly on the present trial ($B = .27$, $SE = .042$, $Z = 6.5$, $p < .0001$; Fig. 6B, as discussed previously); conversely, if the pigeons responded correctly the last time the foil pexigram had appeared as the target pexigram, then they were *less* likely to respond correctly, although this effect was smaller ($B = -.09$, $SE = .015$, $Z = -6.1$, $p < .0001$; Fig. 6C). This latter result suggests that the pigeons may also have to learn to suppress responding to the foil pexigram to master the categorization task. Both effects waned over the course of training indicated by significant interactions of these prior-responding terms with the number of repetitions (target-was-target: $B = -.04$, $SE = .014$, $Z = -2.5$, $p < .0001$; foil-was-target: $B = .08$, $SE = .011$, $Z = 6.8$, $p < .0001$), although this diminution was again greater for the current foil pexigram than for the current target pexigram.

Thus, when either of the current pexigrams had appeared as a target on prior trials, prior responding to each of the pexigrams presented on the current trial appears to predict choice accuracy (Fig. 7B and C). In contrast, when either of the current pexigrams had appeared as a foil on prior trials, prior responding did not significantly affect categorization behavior. Pigeons therefore appear to retain more from responding to target pexigrams on prior trials than from responding to foil pexigrams on prior trials, although what they retain is relevant both for

A. Target pexigram on prior trial appears again as ...



B. Foil pexigram on prior trial appears again as ...

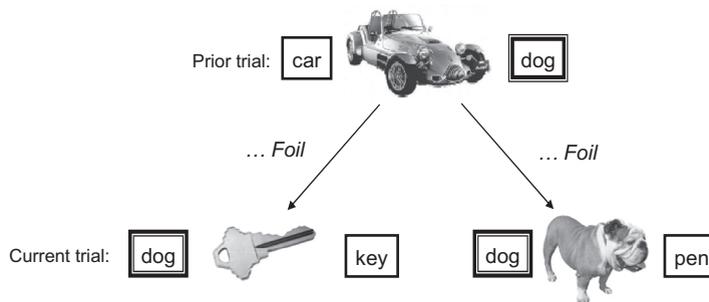


Fig. 7. Possible configurations of target and foil pexigrams across trials. (A) When the target pexigram (car) on a prior trial appears on the current trial, it can appear either as the target pexigram (left) or as the foil pexigram for a different target (duck, right). (B) When the foil pexigram on a prior trial (dog) appears on the current trial, it can again appear as the foil pexigram for a new target (key, left) or it can appear as the target pexigram (right).

enhancing target pexigram responding (Fig. 7B) and for suppressing foil pexigram responding (Fig. 7C) on the current trial.

Second, we also found significant effects of the number of intervening trials between trials sharing target pexigrams and foil pexigrams. These effects followed two broad patterns. If either the target pexigram or the foil pexigram played the same role on prior trials (the target was the target or the foil was the foil), then accuracy was enhanced for short lags regardless of how the bird had previously responded (target was target: $B = -.03$, $SE = .010$, $Z = -2.5$, $p = .014$; foil was foil: $B = -.08$, $SE = .011$, $Z = -6.9$, $p < .0001$; Fig. 8A and B), although the effect was much smaller for the target pexigram. In other words, there was some benefit to simply having seen the target pexigram or the foil pexigram in the same role in the recent past. Both effects waned over training. Thus, no matter how the bird had previously responded, recent experience with the pexigram may matter more than more distant experience. Moreover, the fact that this effect was particularly large for foil pexigrams suggests a strong role for recent experience in suppressing responding to

incorrect pexigrams. In contrast, when the current foil pexigram appeared as a target pexigram or the current target pexigram appeared as a foil pexigram (the roles were reversed), the birds showed poorer performance when this reversal occurred recently (target was foil: $B = .08$, $SE = .012$, $Z = 6.6$, $p < .0001$; foil was target: $B = .07$, $SE = .017$, $Z = 3.9$, $p < .0001$; Fig. 8C and D); that is, the bird was better able to cope with role reversals and to respond correctly if these happened after many intervening trials.

Finally, our analysis of the spatial location of the target and foil pexigrams suggests that, at least early in training, the birds were more strongly responding to spatial location than to the identity of the pexigram. In all four cases, we found strong interactions of prior responding and spatial location. For example, when the current target category had been the target on the prior trial and its corresponding pexigram was in the same location (Fig. 9), correct responding on that prior trial predicted higher accuracy on the current trial; the bird continued to peck at that location and responded correctly. In contrast, when the current target was the target on the prior trial and its corresponding pexigram was in a different location, incorrect

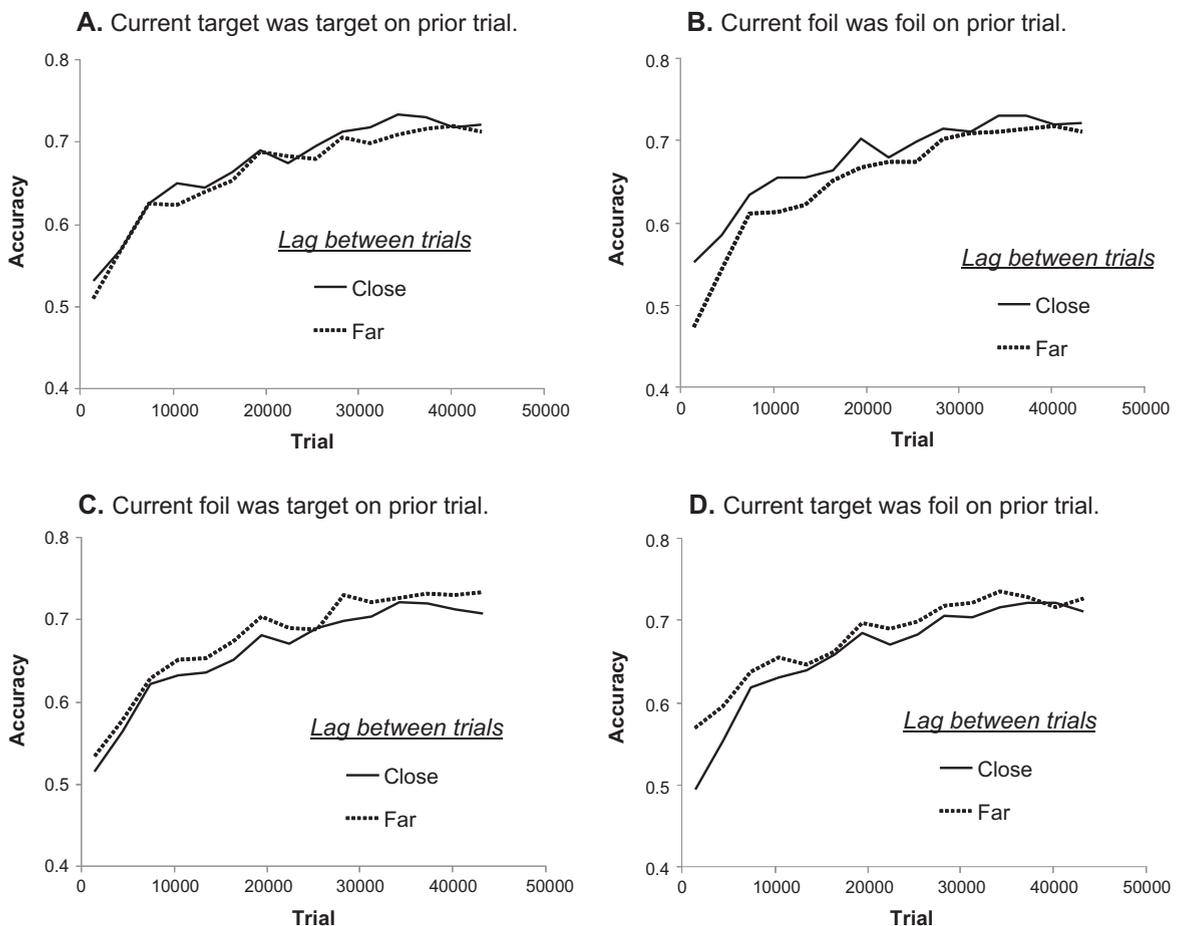


Fig. 8. Effect of the number of intervening trials on accuracy over training. Close refers to 15 or fewer trials between the relevant trials; far refers to more than 15. (A) When the current target was a target on prior trials; (B) when the current foil was a foil on prior trials; (C) when the current foil was a target on prior trials; (D) when the current target was a foil on prior trials.

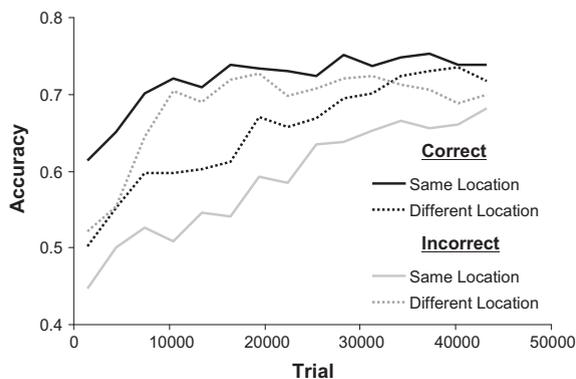


Fig. 9. The effect of prior responding and spatial location on accuracy as a function of training.

responding on the prior trial predicted higher accuracy on the current trial; the birds simply continued pecking what was now the correct location (although it was earlier incorrect). This two-way interaction ($B = .74$, $SE = .025$, $Z = 29.6$, $p < .0001$) further interacted with the number of exposures ($B = -.13$, $SE = .023$, $Z = -5.7$, $p < .0001$), indicating a diminishing effect over training. Similar patterns were observed for each of the other factors; in all cases, the birds appeared to be responding as a function of spatial location and this effect gradually waned over training.

These exploratory analyses offer a number of new factors that affect trial-by-trial performance during learning: distance between exemplars, spatial location, and the role of targets and foils on prior trials. Together, these factors suggest an unforeseen richness to biological associative learning, and prompt interesting “predictions” of our biological model for human learning.

4. General discussion

The goals of this study were twofold. First, we sought to determine if a biological associative system is capable of learning a many-to-many set of mappings between 16 target categories (comprising a total of 128 individual pictorial stimuli) and 16 arbitrary report responses, and whether such learning generalizes to new categories. Second, we sought to evaluate the trial-by-trial structure of the pigeons’ category learning—with an eye toward recent debates in human word learning—in order to gain a better understanding of associative learning in general. We discuss each of these questions in turn. We then end with a more speculative discussion of both the surface dissimilarities between our category learning paradigm and human word learning as well as what appear to be striking similarities between them.

4.1. How much did the pigeons learn and what does it mean?

The present results clearly document a substantial capacity of associative learning in nonhuman animals. All three pigeons learned to respond discriminatively to all 16 categories. The birds learned these categories as

coherent stimulus collections, reliably generalizing their discrimination to new exemplars, and showing more coherence in responding within a category than what would be expected by chance. Our pigeons also showed evidence of learning to categorize 97% of the 128 individual exemplars; indeed, the distribution of discriminative responding across the training exemplars revealed no sign of a capacity limit (Supplement S2). Finally, and of key importance, evidence of this large capacity was achieved while the birds were learning the 16 photographic categories in parallel, a training task that may be substantially more demanding than learning categories sequentially. Thus, our goal of establishing the viability of a many-to-many model of biological associative learning was clearly reached.

The contrast between our task and other animal models of category learning should now be clear. Although parrots, dogs, and chimpanzees have been shown to master more stimulus–response mappings than we attempted here, those earlier training tasks involved extensive social interaction and more functionally relevant referents and reinforcers, such as real toys and foods (Gardner & Gardner, 1984; Kaminski et al., 2004; Pepperberg, 2002). Moreover, most prior studies did not extensively investigate stimulus generalization or category coherence: two key behavioral phenomena that bear critically on the very nature of category learning.

An even more important contrast between our study and prior “word learning” projects with animals lies in the goals of these projects. Most earlier studies focused on whether animals have some capacity for language or lexical behavior writ large. In contrast, our study took a different tack, using an expanded animal conditioning task to elucidate associative learning principles that may be involved in children’s word learning. To do so, we moved from existing animal category learning paradigms (Wasserman et al., 2007) by adding parallel learning, a key facet of children’s word learning (Bloom, 2004). The results can be seen as a biological model of many-to-many associative learning. Not unlike computational models, we necessarily made many simplifying assumptions, but by doing so we were in a better position to isolate and understand the learning itself.

Thus, our present project represents a substantial advance in expanding the capacity of simple associative learning processes in animals and in promoting the applicability of those processes to important problems lying well beyond the realm of animal learning. Moreover, the tight laboratory control offered by our paradigm enabled us to conduct extensive trial-by-trial analyses that uncovered real richness to biological associative learning, even in this carefully constrained learning situation.

4.2. Time course and implications of category learning

Our analysis of the time course of learning also revealed the complexity of associative processes, which is relevant to current debates concerning observational word learning in humans. These analyses show a strikingly similar pattern to the findings of Trueswell et al. (2013). Early in training, pigeons were at chance if they had responded

incorrectly on the last trial featuring that target pexigram, whereas they were above chance if they had responded correctly to the target pexigram (Fig. 6A). Although Trueswell et al. interpret this result as evidence for a propose-but-verify strategy, their account seems highly unlikely in pigeons for several reasons. First, by the end of those first 2980 trials, the pigeons had seen each category many times and they had made multiple correct responses; nevertheless, the birds were still performing somewhat indiscriminately (Fig. 7A). Second, at a broader level, it seems unlikely that pigeons are capable of such an advanced cognitive strategy, particularly when confronted with so many exemplars and responses (we address this point in greater detail below).

Our pigeon project was not intended to mirror any specific human study, and the fact that an explicit reinforcer was present in our work (and not in the prior studies on observational learning), is an important caveat. Nonetheless, the fact that we obtained a highly similar pattern of findings in a slow, gradual associative learning system cautions against interpreting these results (and others like them) as strong support for a learning mechanism that is rapid, uniquely human, language-specific, or based on some form of logical inference. Rather, such results are also consistent with at least one form of associative learning.

This key point is underscored by our deeper investigations into the factors that shape learning. When we examined the entire time course of categorization learning, we found clear evidence for a gradual associative process over and above any prior-encounter effects: performance gradually rose throughout the experiment, even when we only considered trials preceded by an incorrect response to the current target (Fig. 7B). The presence of both a gradual learning effect and an effect of prior target responding suggests that both are natural properties of a purely associative model. Trueswell et al. (2013; and see also, Dautriche & Chemla, 2014), did not assess the effect of gradual learning in humans; however, they only gave five repetitions of each item, perhaps rendering such an effect difficult to observe. In fact, ongoing work with humans using longer training regimes documents a gradual learning effect in unsupervised word learning, suggesting that this finding is not unique to animal learning preparations (Roembke & McMurray, submitted for publication).

Moreover, prior target responding was not the only type of responding that affected later choice behavior. Pigeons' responding to the current foil pexigram depended on that pexigram having been the target pexigram on a prior trial (Fig. 7C); here, however, if pigeons had pecked the foil pexigram before (when it was the target pexigram), then they were more likely to (incorrectly) peck it again. This result suggests that suppressing responding to foil pexigrams is also an important component of many-to-many association formation, particularly early in learning, as the effect rapidly vanished. It is not clear whether this transient effect represents the birds' elimination of a spurious association between the foil pexigram and the target category (pruning) or whether the pigeons simply erroneously associated the foil pexigram with food reinforcement (regardless of the target category). The latter possibility would

represent what Harlow (1959) termed an error factor, an idea to which we later return.

However, substantially more than prior responding participated in our pigeons' performance. The lag between trials with matching pexigram components was highly predictive of accuracy (Fig. 8). Because this lag effect was not significantly affected by the choice that the bird had made on prior trials (and the feedback that followed this response), this result may implicate some form of unsupervised learning. This possibility is strengthened by a recent study suggesting that unsupervised learning in humans benefits from close proximity of related trials (Carvalho & Goldstone, in press). Moreover, the lag effect cannot be rooted in mere familiarity with the category stimuli or pexigrams; recent exposure to an item did not uniformly improve performance. Performance improved if the current target pexigram had recently been seen as a target pexigram (Fig. 8A), and especially so if the current foil pexigram had recently been seen as a foil pexigram (Fig. 8B); but, performance worsened if the roles of current and past pexigrams were reversed (Fig. 8C and D).

These effects of temporal proximity may not only be a matter of learning about the individual training stimuli and/or pexigrams, but also about learning how these items relate to one another to predict reinforcement (e.g., given a particular stimulus, which pexigram is the target and which is the foil). The more robust effect of the foil than the target again stresses the importance of learning to suppress foil responding in addition to enhancing target responding. However, as noted earlier, it is not yet clear whether this finding reflects the suppression of pecking the foil in general (an error factor) or the suppression of pecking the foil in the presence of a specific target stimulus (pruning).

Finally, we observed a large effect of the spatial location of the pexigrams, particularly early in training. Pigeons were clearly responding in a way that was largely dominated by a spatial strategy: if they had been reinforced on the left side the last time they saw that target, then they tended to peck there again, even if that response was now incorrect (Fig. 9). This spatial perseveration is clearly an error-factor, and the fact that it gradually declined over the course of training suggests that pigeons were suppressing it as a crucial part of learning.

This particular effect of spatial location, however, raises the possibility that, in addition to treating space as an overall error factor, these spatial locations may be temporarily associated with the target stimulus. A more extensive analysis of spatial bias (Supplement S3) showed that only one of the pigeons exhibited any overall bias toward the left or right response. Moreover, the "preceding trial" in this analysis is really the preceding trial entailing the same category—which was usually about 15–16 trials away—with multiple trials involving different spatial responses in between the preceding trial and the present trial. The very large magnitude of these spatial effects (in the face of such large intervening variability) suggests that pigeons may actually have been (erroneously) associating the stimulus with a particular pexigram location.

Thus, analysis of the trial-by-trial data suggests that, even in this seemingly simple associative system, we see

the simultaneous influence of multiple factors: a gradual learning effect, the prior response to a target or foil pexigram, the effects of repeating a pexigram in the recent past, and the influence of error factors like spatial location. All of these factors must be sorted out before accurate categorization can be achieved.

4.3. Association vs. logical inference

A critical part of our animal model is its ability to capture the potential richness of associative learning in a biological system. Thus, it is important to address whether or not our paradigm truly captures associative learning, before we can address its implication for human learning in many-to-many situations like word learning.

At the level of experimental preparations, ours was an unalloyed associative learning task. Unlike most prior attempts to teach large numbers of associations to animals (Gardner & Gardner, 1984; Kaminski et al., 2004; Pepperberg, 2002), we used neither elaborate shaping procedures nor social cues and reinforcers; instead, we relied solely on the basic dynamics of stimulus, response, and reinforcement (also see Fagot & Cook, 2006). However, is it possible that, despite this level of control, our pigeons were using a more inferential or rational strategy to learn these mappings. It has been proposed that, at least in some cases, animal learning may be better characterized in terms of information-processing or inferential learning strategies that discretely form and test propositional or symbolic hypotheses, rather than gradual association building (Gallistel et al., 2004; Mitchell, De Houwer, & Lovibond, 2009; but see, Castro & Wasserman, 2009).

This particular proposal for propositional learning is not to be confused with “rational” or Bayesian accounts of learning. These latter analyses can often complement associative accounts by offering a clear understanding of the functional goals of learning and the relevant information and/or expectations that can drive it; however, they still make use of accumulated input/output statistics and prediction/evaluation processes much like associative learning (Courville, Daw, & Touretzky, 2006).¹ In contrast, the alternative offered by Gallistel et al. (2004) represents a more logical or symbolic type of process.

The evidence for information-processing or propositional learning in animal learning has often been derived from paradigms featuring only a small number of stimuli and responses, as in classical conditioning (Gallistel et al., 2004; Gershman, Blei, & Niv, 2010). There is no evidence for this style of learning for many-to-many operant learning tasks such as ours. Thus, it is unclear whether this possibility of more symbolic learning might apply to our paradigm. However, when we consider the nature of our categorization paradigm, there are a number of reasons why information-processing or inferential learning models are implausible.

First, at the cognitive level, a logical or inferential account of this sort would be very difficult to apply in

our paradigm. Any specific trial configuration (photographic stimulus + target pexigram + foil pexigram) was virtually never repeated within a session (on average, a given configuration was repeated only about 10 times over the *entire experiment*), and the lag between repetitions of individual components of the trial configuration (e.g., a member of the same category) was large; the average lag between repetitions of the same target category (*not* the same target exemplar) was approximately 16 trials. These scheduling constraints highlight the inefficiency of simply retaining in memory the configuration or even the elements of recent trials, because they do not recur often enough or recently enough to guide accurate choice behavior (although they clearly exerted some effect on our pigeons' behavior). Rather, it seems likely that our birds were gradually building enduring category–pexigram associations. Thus, our austere discrimination task minimized other psychological processes beyond associative learning that might have contributed to our pigeons' highly discriminative categorization behavior.

Second, a critical argument offered in support of information-processing or inferential learning accounts comes from Gallistel et al. (2004; and Trueswell et al., 2013, explicitly build on this in their studies of humans), who contended that the apparently gradual learning curve observed in most animal learning studies may be the product of a more sudden insight by each animal (occurring at variable times). However, our autocorrelation analyses also offer evidence against this contention. The number of repetitions was modeled as a random slope on individual birds (accounting for such variability) and our analyses captured the effects of any sudden insight (the last-trial effects). Yet, our autocorrelation analyses still yielded evidence of gradual learning. Moreover, those analyses revealed a host of clearly irrelevant factors which shaped learning (e.g., the spatial location of the correct pexigram) and sometimes did so for quite a while. One would suspect that such factors would be quickly eliminated by more inferential approaches. Given these findings and observations, it seems unlikely that any sort of symbolic or inferential process can be the exclusive mechanism of pigeons' learning in our paradigm. Thus, we can treat our animal model as a biological instantiation of associative learning, one that may yield insight into domains like word learning that may utilize a similar form of many-to-many learning.

4.4. How our paradigm differs from human word learning

Our paradigm is not a direct analog of human word learning. Nevertheless, it does offer a unique biological model of a critical property of word learning: namely, the fact that a learner must map many exemplars to many categories. And, as we will next discuss, it illustrates striking parallels to several aspects of word learning. Our animal model, like all models (computational or biological), must simplify the problem; as a result, it does not capture some crucial aspects of word learning. At the highest level, the pigeons' categories and behaviors are not embedded in a communicative context – the bird are not learning and interacting via words. But, several more specific simplifications are worth discussing, as understanding them is

¹ Such analyses have not yet been applied to the case of many-to-many learning, making it unclear if they offer a similar insight here.

important for determining what we can glean from our project.

First, the learning paradigm itself was highly supervised, featuring both positive and negative feedback (respectively: food reinforcement and the darkening of the house lights plus a delay followed by a correction trial). Although we, and many others, have argued that word learning may to a large extent be learned observationally (Bloom, 2000; McMurray et al., 2012; Medina et al., 2011), there may be much more to the story.

There is evidence that children do receive feedback and can profit from it (Bohannon & Laura, 1988; Chouinard & Clark, 2003; O'Hanlon & Roberson, 2007). This feedback is particularly the case for naming situations in which children are corrected after erroneously naming an object (much as our pigeons must self-correct after selecting the incorrect pexigram). Moreover, it is commonly acknowledged that there are many social/pragmatic factors that influence child word learning, such as eye-gaze (Baldwin, Markman, Bill, Desjardins, & Irwin, 1996), pointing, and ostensive naming; these behaviors can be linked in part to the child's own lexical and attentional behavior (Pereira, Smith, & Yu, 2014). Such factors may offer an important supervisory signal to word learning. Finally, there is also computational work implicating a more implicit form of supervision, in which learners make predictions based on the input (e.g., I heard *dog*, so I should expect to see one); learners can then evaluate the validity of those predictions, yielding an error signal which can be input into traditional supervised learning models like the Rescorla–Wagner model (Elman, 1990; Ramscar et al., 2013). Thus, even though supervised learning models such as ours may not capture some of the surface properties of human word learning, there may be a deep core of error-driven learning to human word learning and we cannot dismiss such learning as entirely irrelevant.

Second, one might argue that 16 categories is a far cry from the tens of thousands of words children learn. Of course it is. But, despite decades of debate about the suitability of associative mechanisms for word learning, there are no current animal models that allow us to study those mechanisms because few animal models employ more than a handful of choices or responses to which inputs can be mapped. Sixteen categories is an important step toward understanding the nature of associative learning in a more complex, many-to-many learning problem like human word learning (and it is a comparable or larger number than in many studies of human word learning, particularly with children). In fact, if we consider the associations (rather than the items), then this limitation is brought into bold relief. A typical two-choice experiment may entail only 4 associations (2 stimuli \times 2 responses); in contrast, our paradigm may entail as many as 2048 associations (128 exemplars \times 16 responses). Thus, our task involves a dramatic upward shift in the scale of the associative learning problem.

Third, our birds' rate of learning appears to have been quite slow. Human adults regularly learn 16 categories in the space of an hour (Magnuson et al., 2003; Medina et al., 2011); yet, pigeons took 45,000 trials to reach their associative limits. Would children learn faster than

pigeons? Almost certainly. However, our pigeons came to the experiment with literally no background knowledge; they did not understand the nature of the "task," they had not encountered these categories before, and they had empty lexicons. Children, on the other hand, bring all of these things to bear on the problem of learning words. Thus, the more relevant comparison group may be newborn infants, who indeed take 6–9 months to learn their first words.

As noted by many authors, the critical development in children's vocabulary learning is not simply acquiring the associations, but learning a whole set of interconnected knowledge and acquiring an understanding of the word learning problem, both of which help the child to acquire words more efficiently. Indeed, our discussion of error factors in the next section suggests that, although pigeons may also embark on their task as naïve learners, they too may end this task with substantially more knowledge than merely the category-pexigram associations; they might now be "ready" to learn additional mappings even more rapidly. Still, accepting this difference at face value, it seems likely that pigeons will never learn category mapping as quickly or robustly as humans. Nonetheless, we must appreciate that the point of many models is not to match human performance, but to divulge and elucidate some critical mechanism of learning.

4.5. Implications for word and category learning

To repeat, we do not claim to have taught words to pigeons or that their acquiring our task captures all of what it means to learn a word. At the same time, our task clearly captures the many-to-many nature word learning. Of possibly greater importance, however, is just *how* our pigeons mastered this demanding task. Here, our trial-by-trial analyses may have disclosed deep similarities to several aspects of children's word learning.

4.5.1. Error factor theory and children's understanding of reference

Our autocorrelation analyses revealed that pigeons initially respond to several factors that are entirely irrelevant to the categorization problem. The pigeons attended to the location of the pexigrams or the recent conjunction of a particular pexigram with reinforcement (regardless of the pictorial stimulus), even though these factors are gradually ignored in favor of the procedurally correct task contingencies. Clearly, pigeons do not come to our category learning task with an understanding of what is expected of them; instead, they must gravitate toward the relevant task attributes over training, eliminating irrelevant error factors (Harlow, 1959) from consideration.

This error factor analysis is analogous to the common argument that children discover something fundamental about the nature of words, typically from 18 to 24 months of age. Whether it is the naming insight (Goldfield & Reznick, 1990; Stern, 1924), a move to a more social approach to word learning (Golinkoff & Hirsh-Pasek, 2006), a shift toward conceptual understanding (Namy, 2012), or simply an appreciation of which situations are referential and which are not (e.g., Fennell & Waxman,

2010), all of these approaches posit that children divine something fundamental about the basic problem of learning words.

This argument is often advanced to discredit an associative account of word learning: at this critical juncture, children may cease learning associatively and begin engaging more complex cognitive processes (Golinkoff & Hirsh-Pasek, 2006; Namy, 2012). But, given our evidence on the role of error factors in pigeons' categorization behavior, this purported associative → cognitive progression in children may be analogous to how our purely associative pigeons eliminate other possible interpretations of the task, finally to focus on the stimulus → pexigram mappings. Moreover, our research suggests that any seeming associative → cognitive progression could be achieved by entirely associative mechanisms (as also seen in connectionist models, e.g., Mayor & Plunkett, 2010). From this vantage point, what appears to be a shift to a more cognitive mode of learning in children may actually mark the elimination of error factors; rather than marking the end of associative learning, the elimination of error factors may simply allow associative learning to shift into high gear.

4.5.2. The richness of associative learning

Our categorization task implements an ostensibly reinforcement-driven or *supervised* form of learning. However, it is widely argued that children engage in at least some *unsupervised* learning, as there are too many words to be taught via purely explicit regimes (Akhtar, Jipson, & Callanan, 2001; Medina et al., 2011; Smith & Yu, 2008). However, this distinction is not as clear as it might seem for two reasons.

First, as we described earlier, a number of models argue that even without reinforcement, supervised learning (e.g., the Rescorla–Wagner rule or closely related back-propagation of error) can learn quite effectively by making predictions and evaluating them against the input (Baayen et al., 2013; Elman, 1990; Ramscar et al., 2013). Second, our learning paradigm may contain a component of unsupervised learning as well. Because not every pexigram was available on every trial, the probability of a particular category and its corresponding (target) pexigram being seen on the same trial was always 1.00, whereas the probability of that category and any other (foil) pexigram on the same trial was only 0.07. This arrangement is perfect fodder for unsupervised, cross-situational learning (Yu & Smith, 2007); the associative “boost” offered by this additional source of unsupervised category-pexigram learning may have played an important part in the pigeons' success in our task.

Indeed, such an effect of unsupervised learning may have been the source of the repetition effect that we observed, in which repeating the foil pexigram on recent trials improved categorization accuracy (Fig. 9B) (Carvalho & Goldstone, *in press*). Moreover, the fact that our repetition effects were larger for repeated foil pexigrams than repeated target pexigrams suggests that one role of unsupervised learning may be to prune irrelevant associations (as opposed to forging correct associations). As we discuss later, this form of unsupervised learning/

pruning may underlie a number of empirical phenomena in children's word learning, as revealed by a recent computational model (McMurray et al., 2012).

Nevertheless, this repetition effect, the effects of choice responding on prior trials, and the effects of space on prior trials, all suggest that associative learning cannot be conceptualized as solely building co-occurrence statistics between words and objects (e.g., Yu & Smith, 2012). This evidence is not news to the animal learning community (Rescorla, 1988). However, to account for data such as ours, animal learning researchers must begin entertaining more complex learning rules that embrace the contributions of both the statistics of the inputs (the unsupervised component of learning; specifically, the stimulus-pexigram correlations), the patterns of reinforcement (the supervised component of learning), real-time decision processes (choice responding), and other relevant factors (short-term memory, error factors).

Even while the animal learning community may have embraced some of these ideas, this more complete picture has not always informed debates about the role of associations in human word learning (e.g., Medina et al., 2011). Although we set out to investigate the “associative core” that may underlie many-to-many problems in word learning, it is clear that even in our simple biological model, this core may be embedded in a sophisticated system including multiple forms of learning (supervised/unsupervised, error factors) and real-time processes (effects of prior stimuli, responses, and locations).

A recent theoretical/computational model suggests a fresh way of thinking about such a system (McMurray et al., 2012, 2013). This model embeds associative (unsupervised) learning in a real-time competition or decision-making framework. It discloses that the interactions of these two processes can give rise to a range of complex developmental phenomena in children's word learning. And, although this model cannot capture the full complexity of our data set, it does underscore the power of embedding associative learning in a richer system, as illustrated by our trial-by-trial analyses.

Importantly, the McMurray et al. (2012, 2013) model helps us appreciate that, under the sorts of training regimes that have thus far been deployed, our pigeons are never going to be the proficient learners and word users that children are. Even though both humans and pigeons share an associative core to word learning, they are unlikely to share the sort of real-time processes that live on top of this core. Pigeons are unlikely to exploit social cues, to use words generatively, and so forth, phenomena which have been used to argue against associative accounts of children's word learning (Markman & Abelev, 2004).

However, the McMurray et al. (2012, 2013) model plus our rich dataset suggest that these arguments amount to an attack on a straw-man version of association formation (see also, Smith, 2000). Associations need not occur between raw inputs and outputs; factors like attention, competition, and other forms of cognitive processing may shape these inputs prior to (or during) the formation of associations. In addition, the presence of internal or mediating representations can participate in associations and

change learning dramatically (McMurray et al., 2013). Therefore, in this model, the added complexity of children's word learning may not necessitate a shift away from associative learning. Rather, word learning can be understood within a framework in which association formation is the core of learning, but this core is embedded in a system with more sophisticated real-time processes, richer inputs, and greater opportunities for building internal representations.

4.5.3. Enduring effects of prior trial events

Some of the most striking findings from our trial-by-trial analysis involved the enduring influence of prior events on categorization accuracy, including the photographic stimulus, the choice pexigrams, the spatial locations of the choice alternatives, and the birds' own earlier responses. When we looked at how these factors interacted with the birds' categorization behavior (and subsequent reinforcement or nonreinforcement), the effects of both spatial location (Fig. 9) and prior choice responses (Fig. 7A and B) persisted for quite a while.

Our analysis focused on how a prior trial with the same target (or foil) pexigram influenced performance on the present trial; on average, these prior trials occurred some 16 trials earlier and yet they still affected categorization accuracy, thus suggesting a surprising durability of this information. Such longevity is difficult to predict from most current animal learning rules; it implies a memory mechanism which can track the short-term reinforcement history of the birds' responses relative to multiple factors (e.g., was the bird reinforced for pecking a specific spatial location, or for pecking an individual pexigram, or for choosing an individual pexigram in response to a specific stimulus?). This memory for multiple contingencies between trial properties, responses, and reinforcers must then be integrated to make a momentary response decision on any given trial.

In our task, retaining the details of individual stimulus and response events some 16 trials ago is clearly not optimal. The identity of the foil pexigram or the identity or the specific location of the target pexigram several trials ago provides no useful information as to which pexigram is correct on the present trial; only the category-pexigram relation is relevant to solving the task. Yet, there are two reasons why the retention of prior trial events may be important.

First, because pigeons do not enter the experiment knowing that the task is going to entail category learning, this initial strategic multiplicity may ultimately allow a broader range of events and behaviors to be learned, thereby permitting the animal to entertain or to eliminate a wealth of possible error factors. Second, pigeons' memory of earlier trials is analogous to the concept of momentum in connectionist modeling. Momentum is a computational "trick" that is commonly used in supervised learning models. To implement momentum, the weight change (or associative change) from prior trials is partially carried over to the current trial; so, the weight change that is initiated on a prior trial will partially persist on later trials. Although the notion of momentum is not believed to have strong theoretical import (e.g., McCloskey, 1991), it improves network

learning by helping the network avoid "good, but not great" solutions (e.g., local minima). Our pigeons' clear retention over several trials of details concerning their history of response and reinforcement may represent an analog of this notion, which might help biological learning systems overcome similar problems.

4.5.4. Space matters

In our trial-by-trial analyses, a strong contributor to performance proved to be the spatial location of the correct pexigram (Fig. 9). Early in training, pigeons appeared to associate spatial location (not pexigram identity) with reinforcement, a tendency which declined throughout training. As we have discussed, this effect of spatial location likely represents a bias tied to a specific stimulus (a stimulus-location mapping).

In this case, spatial bias is clearly an error factor to be ruled out over training. However, it may closely accord with children's behavior in recent studies of word learning. Samuelson, Smith, Perry, and Spencer (2011) replicated a classic study by Baldwin (1993), investigating children's use of social cues to infer the meaning of new words. Samuelson et al. found that, during the learning phase in their project, children were using spatial cues to organize the visual scenes. Crucially, when referential cues were pitted against spatial cues (the speaker was clearly looking at and referring to an object in the same spatial location that the other object usually occupied), children followed spatial cues mapping the word to the incorrect object.

Thus, children, like pigeons, may also be associating space with objects and their names. Space may serve a critical role in organizing and binding elements of the visual scene to support children's word learning by combining multiple features into a single object or category that is available to map onto a word. There is now good reason to believe that the same may be true in pigeons' category learning; this close parallel between children's and pigeons' behavior suggests that a similar associative core may underlie aspects of category learning in both species.

4.5.5. Pruning incorrect associations

The other major finding that arose from our trial-by-trial analyses was that both the prior target and foil pexigrams played unique roles in categorization behavior. Seeing a foil pexigram as a foil pexigram on recent trials (Fig. 8B) yielded a dramatic benefit in performance throughout training; however, when the current foil pexigram had been a target, pecking it on those prior trials impaired performance (Fig. 6C, at least early in training). Although common conceptions of associative learning stress the building of linkages between discriminative stimuli and the correct response, our findings suggest that pigeons must also learn to suppress responding to incorrect choice alternatives. We envision two possible mechanisms mediating this suppression.

First, it is possible that these foil-based effects are an error factor: pigeons are associating reinforcement with a specific pexigram (not with a stimulus/pexigram linkage). They therefore continue to peck pexigrams that were previously followed by reinforcement (e.g., if the current foil pexigram had earlier been a target). This associative

account can explain our response-based effects (Fig. 6), but it is not clear if it can handle our distance effects (Fig. 8B).

Second, although we may envision our categorization task as one of merely requiring pigeons to learn to map images of, say, *cars* to the *car-pexigram*, it may also be important for pigeons to learn that images of *cars* should *not* be mapped to the *flower-pexigram* or to the *baby-pexigram* or to the *pen-pexigram*, and so on. Under this second scenario, it is possible that pigeons must learn to suppress partially and inappropriately formed links between the stimulus and incorrect pexigrams. In fact, spatial biases may also be of this same sort, as pigeons must learn that *cars* should *not* be mapped to either the left or right location.

In our paradigm, such a winnowing process may be effective, but it would not appear to be efficient: given a total of 16 pexigrams, for each category, there is only 1 correct category-pexigram association to build, but 15 irrelevant ones to ignore or to prune. Nevertheless, in the real world (e.g., in children's word learning), pruning may be even more efficient because many incorrect objects may be present in any given naming situation, not just one as in our categorization paradigm, allowing children to eliminate many associations in parallel.

The aforementioned computational model of McMurray et al. (2012; McMurray et al., 2013) underscores the importance of such a selectionist process (and see also Regier, 1996). When the nature of the model's learning was analyzed, one of the most important predictors of a number of findings was the degree to which *incorrect* associations were formed, retained, or pruned. For example, the strength of incorrect associations predicted the gradual improvement of the network's reaction time (e.g., Fernald, Perfors, & Marchman, 2006) better than did the strength of correct associations. Similarly, the model's ability to fast-map (a largely in the moment decision process, see Bion, Borovsky, & Fernald, 2013; Horst & Samuelson, 2008) was the product of the particular associations that were pruned, not any specific strategy or learning mechanism. In both cases, the pruning of incorrect associations proved to be the fundamental property of the model that enabled it to account for many of these developmental phenomena.

This model of human word learning in unsupervised paradigms largely uses the lack of co-occurrence between word and object (e.g., *tree* rarely co-occurs with a *table*) as the basis for pruning, whereas in supervised learning the nonreinforcement of specific incorrect responses likely plays a more crucial rule. In fact, it is possible that the mild punishment given after an incorrect response in our paradigm may have enhanced pruning effects (over what would be seen in humans), and it is clear that more human work is needed to document this effect. Nonetheless, the fact that pruning may be important in both biological and computational models of associative learning (each kind of model making different assumptions) raises important issues. Moreover, it is not the method by which associations are pruned, but the fact that they were pruned at all that predicted the aforementioned phenomena in the McMurray et al. model (e.g., fast-mapping and changes in processing speed); thus, while we await empirical confirmation of pruning in humans, our converging

evidence for this type of learning may speak to its importance more broadly.

4.5.6. What do pigeons do differently?

Although there are clear qualitative similarities between humans and pigeons, there are also clear differences. Humans likely learn words at a faster rate, as part of a system, without substantial supervision, and for the purpose of communication. Some of these differences are a matter of degree; primates can clearly learn categories for communication (Gardner & Gardner, 1984), and dogs, clearly can learn something of the system, showing phenomena like mutual exclusivity (Griebel & Oller, 2012; Kaminski et al., 2004). So, one answer is that species differences may be a matter of degree more than mechanism. There may also simply be developmental differences; by virtue of developing in a linguistic system, humans may, as a rule, eliminate error factors more efficiently than other animals to arrive at an understanding of the task (even though such error factor elimination can be performed by associative systems). There may also be more qualitative differences in the nature of the information sources that pigeons and humans can use as a source for associative learning; humans for example, may be more adept at using social cues as a form of feedback, complex inference to arrive at a better decision, or even unsupervised statistics to shape the associations than nonhumans. All of issues are surely speculative without a richer body of evidence from similar animal learning paradigms; but, such wide-ranging comparisons were not central to our project. Rather, if we consider the learning mechanisms in isolation, then our work suggests that the associative core of word/category learning may share many emergent commonalities between humans and pigeons, even if that core is embedded in a very different set of social, cognitive, and developmental systems.

4.6. Final remarks

Our work supports an emergentist approach to language acquisition. It adds to a growing body of evidence suggesting that word learning—and specifically the mapping of large numbers of categories to large numbers of responses—can operate via basic principles like associative learning (Horst, Samuelson, Kucker, & McMurray, 2011; McMurray et al., 2012; Samuelson & Smith, 1998; Samuelson et al., 2011; Smith & Yu, 2008). And, it demonstrates that trial-by-trial effects on learning which seem to support more inferentially-based accounts can also be supported by associative systems; it does so by stripping word learning to its core—mapping many categories to many responses—in a biological model of associative learning.

Although it is certainly premature to argue that human word learning is either purely or predominately associative in nature, it does not seem like a reach to suggest that human word learning involves associative processes. Moreover, as we come to understand the complexities of associative learning in both biological and computational models, as well as to appreciate the subtle sources of information in both stimulus and reward in the learning environment, we may find that such models become harder to discount.

The continuity across species that we observed—in terms of the factors that influence learning and the outcomes that emerge from learning a complex set of associations—suggests a core mechanism for children’s word learning that at least partially involves associative learning. The accomplishments of both children and pigeons, although likely served by such mechanisms, cannot be properly characterized as due to “mere associations.” Although children obviously learn to use spoken words in ways that our pigeons never will be able to approximate by tapping their pexigrams, our pigeons’ associative learning shows many hallmarks of children’s word learning, including categorical coherence and generalization, as well as the progressive elimination of error factors. Appreciating these complexities does not require us to abandon associative learning; rather, our animal model suggests that these complexities may emerge quite naturally in biological associative learning systems. In this regard, the much greater complexity of children’s lexical development may actually derive from domain-general emergentist developmental principles that serve as a crucial core to children’s sophisticated lexical capacities.

Acknowledgements

The authors thank Olga Lazareva, Michelle Miner, and Matt Manning for helping to pioneer early versions of the present categorization task, Leyre Castro for valuable assistance in all aspects of this project, Larissa Samuelson, Tanja Roembke, Darrin Miller and Karla McGregor for helpful discussions about the nature of associative learning in language development, and Mark Blumberg for posing the question that provided a key framing of the language debate. This research was supported by National Institute of Mental Health Grant MH47313 and by National Eye Institute Grant EY019781 awarded to EW and by National Institute of Deafness and Other Communication Disorders Grant DC008089 awarded to BM.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2014.11.020>.

References

- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 67, 635–645.
- Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, 72(2), 416–430.
- Akhtar, N., & Martinez-Sussman, C. (2007). Intentional communication. In C. Brownell & C. Kopp (Eds.), *Socioemotional development in the toddler years: Transitions and transformations* (pp. 201–220). New York, NY: Guilford Press.
- Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, 56(3), 329–347. <http://dx.doi.org/10.1177/0023830913484896>.
- Baldwin, D. A. (1993). Early referential understanding: Infants’ ability to recognize referential acts for what they are. *Developmental Psychology*, 29, 832–843.

- Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., & Irwin, J. M. (1996). Infants’ reliance on a social criterion for establishing word-object relations. *Child Development*, 67(6), 3135–3153.
- Bates, D., & Sarkar, D. (2011). *lme4: Linear mixed-effects models using S4 classes*.
- Battig, W. F. (1972). Intratask interference as a source of facilitation in learning and retention. In J. F. Voss (Ed.), *Topics in learning and performance* (pp. 131–159). New York: Academic Press.
- Bhatt, R. S., Wasserman, E. A., Reynolds, W. F., & Knauss, K. S. (1988). Conceptual behavior in pigeons: Categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 14(3), 219–234.
- Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. <http://dx.doi.org/10.1016/j.cognition.2012.08.008>.
- Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: The MIT Press.
- Bloom, P. (2004). Myths of word learning. In G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 205–256). Cambridge, MA: The MIT Press.
- Bohannon, J. N. S., & Laura, B. (1988). The issue of negative evidence: Adult responses to children’s language errors. *Developmental Psychology*, 24(5), 684–689.
- Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Practice and working memory effects in building procedural skill. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 517–526.
- Carvalho, P., & Goldstone, R. (in press). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin and Review*. <http://dx.doi.org/10.3758/s13423-014-0676-4>.
- Castro, L., & Wasserman, E. A. (2009). Rats and infants as propositional reasoners: A plausible possibility? *Behavioral and Brain Sciences*, 32, 203–204.
- Chomsky, N. (1958). A review of B. F. Skinner’s verbal behavior. *Language*, 35(1), 26–58.
- Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30(03), 637–669. <http://dx.doi.org/10.1017/S0305000903005701>.
- Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31(05), 489–509. <http://dx.doi.org/10.1017/S0140525X08004998>.
- Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, 112(2), 347–382.
- Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, 10(7), 294–300.
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, 54(1), 1–19. <http://dx.doi.org/10.1016/j.jml.2005.09.003>.
- Dautriche, I., & Chemla, E. (2014). *Cross-situational word learning in the right situations*.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179.
- Fagot, J., & Cook, R. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences*, 103, 17564–17567.
- Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, 81(5), 1376–1383. <http://dx.doi.org/10.1111/j.1467-8624.2010.01479.x>.
- Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge, UK: Cambridge University Press.
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. (2009). Using speakers’ referential intentions to model early cross-situational word learning. *Psychological Science*, 20, 578–585.
- Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36), 13124–13131. <http://dx.doi.org/10.1073/pnas.0404965101>.
- Gardner, R. A., & Gardner, B. T. (1984). A vocabulary test for chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, 98(4), 381–404.

- Gershman, S. J., Blei, D. M., & Niv, Y. (2010). Context, learning, and extinction. *Psychological Review*, 117(1), 197.
- Gibson, B. M., Wasserman, E. A., Frei, L., & Miller, K. (2004). Recent advances in operant conditioning technology: A versatile and affordable computerized touchscreen system. *Behavior Research Methods, Instruments, & Computers*, 36, 355–362.
- Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the vocabulary spurt. *Journal of Child Language*, 17, 171–183.
- Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current Directions in Psychological Science*, 15, 30–33.
- Griebel, U., & Oller, D. K. (2012). Vocabulary learning in a Yorkshire Terrier: Slow mapping of spoken words. *PLoS ONE*, 7(2).
- 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.
- Harlow, H. F. (1959). Learning set and error factor theory. *Psychology: A Study of a Science*, 2, 492–537.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. *Science*, 146, 549–555.
- Horst, J. S., & Samuelson, L. (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy*, 13(2), 128–157.
- Horst, J. S., Samuelson, L. K., Kucker, S., & McMurray, B. (2011). What's new? Children prefer novelty in referent selection. *Cognition*, 118(2), 234–244.
- Hsu, A. S., & Chater, N. (2010). The logical problem of language acquisition: A probabilistic perspective. *Cognitive Science*, 34(6), 972–1016. <http://dx.doi.org/10.1111/j.1551-6709.2010.01117.x>
- Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: Evidence for "Fast Mapping". *Science*, 304(5677), 1682–1683. <http://dx.doi.org/10.1126/science.1097859>.
- Katagiri, M., Kao, S.-F., Simon, A. M., Castro, L., & Wasserman, E. A. (2007). Judgments of causal efficacy under constant and changing interevent contingencies. *Behavioural Processes*, 74(2), 251–264.
- Lazareva, O. F., Freiburger, K., & Wasserman, E. A. (2004). Pigeons concurrently categorize photographs at both basic and superordinate levels. *Psychonomic Bulletin & Review*, 11, 1111–1117.
- Liberman, A. M., & Whalen, D. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4(5), 187–196.
- Livesey, E. J., & McLaren, I. P. L. (2011). An elemental model of associative learning and memory. In E. Pothos & A. J. Wills (Eds.), *Formal approaches in categorization* (pp. 153–172). Cambridge, UK: Cambridge University Press.
- MacCorquodale, K. (1970). On Chomsky's review of Skinner's verbal behavior. *Journal of the Experimental Analysis of Behavior*, 13, 83–99.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 133(2), 202–227.
- Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398), 77–80. <http://dx.doi.org/10.1126/science.283.5398.77>.
- Markman, E. M., & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, 8(11), 479–481. <http://dx.doi.org/10.1016/j.tics.2004.09.007>.
- 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., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6(11), 465–472.
- McCloskey, M. (1991). Networks and theories: The place of connectionism in cognitive science. *Psychological Science*, 2(6), 387–395. <http://dx.doi.org/10.1111/j.1467-9280.1991.tb00173.x>.
- McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In H. G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109–165). New York: Academic Press.
- McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631.
- McMurray, B., Horst, J. S., & Samuelson, L. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119(4), 831–877.
- McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219–246.
- McMurray, B., Zhao, L., Kucker, S., & Samuelson, L. K. (2013). Probing the limits of associative learning: Generalization and the statistics of words and referents. In L. Gogate & G. Hollich (Eds.), *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence* (pp. 49–80). Hershey, PA: IGI Global.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019.
- Mirman, D., & Spivey, M. J. (2001). Retroactive interference in neural networks and in humans: The effect of pattern-based learning. *Connection Science*, 13(3), 257–275.
- Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, 32(02), 183–198.
- Namy, L. (2012). Getting specific: Early general mechanisms give rise to domain-specific expertise in word learning. *Language Learning and Development*, 8(1), 57–60.
- O'Hanlon, C. G., & Roberson, D. (2007). What constrains children's learning of novel shape terms? *Journal of Experimental Child Psychology*, 97(2), 138–148. <http://dx.doi.org/10.1016/j.jecp.2006.12.002>.
- Pepperberg, A. (2002). *The Alex studies*. Cambridge, MA: Harvard University Press.
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, 21(1), 178–185.
- Perrachione, T. K., Lee, J., Ha, L., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *Journal of the Acoustical Society of America*, 130(1), 461–472.
- Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, 86(2), 184–195.
- Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456–463.
- Quine, W. V. O. (1960). *Word and object: An inquiry into the linguistic mechanisms of objective reference*. Cambridge, MA: The MIT Press.
- Ramsar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, 24(6), 1017–1023.
- Ramsar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, 34(6), 909–957.
- Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA, US: The MIT Press.
- Regier, T. (2005). The emergence of words: Attentional learning in form and meaning. *Cognitive Science*, 29(6), 819–865.
- Rescorla, R. A. (1988). Pavlovian conditioning: It's not what you think it is. *American Psychologist*, 43(3), 151–160.
- 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.
- Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. *Developmental Science*, 10, 734–740.
- Roembke, T., & McMurray, B. (submitted for publication). Learners consider multiple hypotheses in parallel during statistical word learning. *Journal of Experimental Psychology: General*.
- Rumbaugh, D. M. (Ed.). (1977). *Language learning by a chimpanzee: The Lana project*. New York, NY: Academic Press.
- Saffran, J. R., & Thiessen, E. D. (2007). Domain-general learning capacities. In E. Hoff & M. Shatz (Eds.), *Handbook of language development* (pp. 68–86). Cambridge, UK: Blackwell.
- Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15–20-month-olds. *Developmental Psychology*, 38, 1016–1037.
- Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: An alternative account of Akhtar, Carpenter and Tomasello. *Child Development*, 1, 94–104.
- Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PLoS ONE*, 6(12).
- Savage-Rumbaugh, S. (2009). Empirical Kanzi. *Skeptic*, 15, 25–33.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1–2), 39–91.
- Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing Group.
- Smith, L. B. (2000). Avoiding associations when it's behaviorism you really hate. In R. M. Golinkoff, K. Hirsh-Pasek, L. Bloom, L. B. Smith, A. L. Woodward, & N. Akhtar, et al. (Eds.), *Becoming a word learner: A debate on lexical acquisition* (pp. 169–174). New York, NY: Oxford University Press.
- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558–1568.

- Soto, F. A., & Wasserman, E. A. (2010). Error-driven learning in visual categorization and object recognition: A common elements model. *Psychological Review*, *117*(2), 349–381.
- Stern, W. (1924). *Psychology of early childhood* (A. Barwell, Trans.). London: George Allen & Unwin.
- Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In *Language development: The essential readings* (pp. 111–128). Malden: Blackwell Publishing.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.
- Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.
- Wasserman, E. A., & Berglan, L. R. (1998). Backward blocking and recovery from overshadowing in human causal judgement: The role of within-compound associations. *The Quarterly Journal of Experimental Psychology: Section B*, *51*(2), 121–138.
- Wasserman, E. A., Brooks, D. I., Lazareva, O. F., & Miner, M. A. (2007). A vocabulary test for pigeons (*Columba livia*). Paper presented at the meeting of the psychonomic society, Long Beach, CA.
- Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning? *Annual Review of Psychology*, *48*(1), 573–607. <http://dx.doi.org/10.1146/annurev.psych.48.1.573>.
- Waxman, S. R., & Gelman, S. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263.
- Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, *9*, 185–211.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.
- Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review*, *119*(1), 21–39. <http://dx.doi.org/10.1037/a0026182>.
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2013). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *1*–22.
- 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.