New Ideas in Psychology 50 (2018) 73–79

Contents lists available at ScienceDirect

New Ideas in Psychology

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

A communicative approach to early word learning

Daniel Yurovsky

Department of Psychology, University of Chicago, 5848 S University Ave., Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Received 16 February 2017
Received in revised form
31 August 2017
Accepted 1 September 2017
Available online 12 September 2017

Keywords:
Language acquisition
Learning
Cognitive development

ABSTRACT

Young children learn the meanings of thousands of words by the time they can run down the street. Many efforts to explain this rapid development begin by assuming that the computational-level problem being solved is acquisition. Consequently, work in this line has sought to understand how children infer the meanings of words from cues in the communicative signals of the speakers around them. I will argue, however, that this formulation of the problem is backwards: the computational problem is communication, and language acquisition provides cues about how to communicate successfully. Under this framing, the natural unit of analysis is not the child, but the parent-child dyad. A necessary consequence of this shift is the realization that the statistical structure of the input to the child is itself dependent on the child. This dependency radically simplifies the computational problem of learning and using language.

> The infant's Language Acquisition Device could not function without the aid given by an adult who enters with him into a transactional format. That format, initially under the control of the adult, provides a Language Acquisition Support System, LASS. It frames or structures the input of language and interaction to the child's Language Acquisition Device in a manner to "make the system function."
>
> Bruner (1983).

1. Introduction

Humans get a lot of language learning done in strikingly little time. A useful comparison here is the relative rate of two of the most chronicled developmental milestones: language and locomotion. By the time she is a year old, the descriptive norm for a typically developing infant is to produce several words, and to know the names of many common objects. However, the same infant will still struggle to walk at all unless she is holding onto furniture with one hand. When this descriptively normative infant develops into a three-year-old toddler, she will be expected to produce multi-word utterances, to understand prepositions (e.g. on, under), and to describe scenes in picture books. However, this same toddler will still be unable to stand on one foot for more than one second at a time (Squires, Bricker, & Twombly, 2009). There is every reason to think that learning to walk should be a hard problem—it certainly has been difficult to build artificial systems that do it well (e.g Collins, Ruina, Tedrake, & Wisse, 2005). Walking is a problem that humans do not seem especially adept at solving relative to other species, particularly in comparison to their clearly unique trajectory in acquiring language (Capaday, 2002; Garwicz, Christensson, & Psouni, 2009; Hockett, 1959). In contrast, human children are uniquely adept at acquiring natural language—a hard problem that infants make look easy. Indeed, in the foreword to his seminal book on the topic, Bloom (2000) writes that "the child's ability to learn new words is nothing short of miraculous."

So what explains our precocious ability to acquire language? For the present paper, let us follow Bloom (2000) and focus specifically on learning words. And let us get even more specific: Concrete nouns. Of course, this does not exhaust the space of what children can or do learn in their first few years. But concrete nouns are a useful locus for two reasons: (1) Concrete nouns do make up a large slice of early vocabularies (Caselli et al., 1995; Gentner, 1982), and (2) The problem of acquisition should be even worse for more complex and abstract units of language.

2. The computational problem of language learning

Although details vary from analysis to analysis, roughly speaking there is broad consensus about the "computational problem of word learning" for concrete nouns (Marr, 1982). The child is an observer in a world filled with three kinds of

E-mail address: yurovsky@uchicago.edu.

http://dx.doi.org/10.1016/j.newideapsych.2017.09.001
0732-118X/

observables: words, objects, and referential cues. On any given occasion, the child hears a subset of the words, sees a subset of the objects, and also possibly sees one or more referential cues (e.g. a speaker's gaze) that point to a subset of the objects. The computational problem is to recover from these observables a lexicon—a latent structure that details the mapping between words and objects. The solution to this problem is to resolve the uncertainty about the lexicon by leveraging either the cues available on individual instances, the statistical relationship between words and referents across instances, or both (e.g., Blythe, Smith, & Smith, 2010; Frank, Tenenbaum, & Fernald, 2013; Kachergis, Yu, & Shiffrin, 2012; McMurray, Horst, & Samuelson, 2012; Siskind, 1996; Yu, 2008; Yurovsky, Fricker, Yu, & Smith, 2014; etc.).

Following this analysis, there is a growing body of experimental evidence that humans—both adults and children—are capable of using exactly this kind of information to learn words. For instance, infants are sensitive to cues like eye-gaze and pointing quite early in life, and can be shown reliably to use them to learn novel words early in the second year of life (e.g. Baldwin, 1993; Corkum & Moore, 1998; Scaife & Bruner, 1975; Tomasello, Carpenter, & Liszkowski, 2007). Similarly, adults have been shown to infer word-object mappings from co-occurrence information under a host of different conditions (e.g. Vouloumanos, 2008; Yu & Smith, 2007; Yurovsky, Yu, & Smith, 2013b), and many of these experiments have been extended to children and infants as well (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Vouloumanos & Werker, 2009).

Taken together, these and other similar results are taken as compelling evidence of movement in the right direction: Towards modeling the rapid pace of children's early word learning. There are skeptical arguments about this framework from the perspective of generalizability—e.g., will these same kinds of mechanisms explain the acquisition of verbs or adjectives (c.f. Scott & Fischer, 2012)? In this article, I will make a different kind of argument: Our optimism is misguided because of an unlicensed inference from early competence to expert performance (Chomsky, 1965). These and other demonstrations of early success in learning words from social and statistical cues are evidence of competence: they show that infants *can* learn from these regularities. But they have also been taken as evidence that humans excel at learning from these kinds of regularities—that they are subject to few performance constraints—and this inference is unwarranted. Many of these studies demonstrate that adults are not terribly good at learning words from social or statistical cues. And children are even worse.

Let us consider a representative case of social cues: The use of a speaker's eye-gaze to determine the target of her reference. As the title of their landmark paper says, Scaife and Bruner (1975) demonstrate the "capacity for joint visual attention in the infant." Their results show, for instance, that 30% of 2–4 month old children follow an experimenter's gaze in one or both trials on which they are tested; infants do not show levels of success near 100% until they are a year old. These studies demonstrate capacity; they do not demonstrate excellence. More recent studies using different paradigms show similar results: Young children succeed at above-chance levels, but there is significant development well into late childhood (Hollich et al., 2000; Moore, Angelopoulos, & Bennett, 1999; Yurovsky & Frank, 2017; Yurovsky, Wade, & Frank, 2013a). In all of these paradigms, success is defined as the ability to use the speaker's gaze and head direction to distinguish whether she is referring to an object on her left or an object on her right. In more complex visual settings, even older children and adults have difficulty using gaze to infer the target of a speaker's reference (Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Vida & Maurer, 2012).

The pattern of results for statistical word learning is strikingly similar. While infants demonstrate sensitivity to the co-occurrence information between words and objects, their memory for this information is quite fragile, even under low levels of ambiguity. For instance, in a study by Vlach and Johnson (2013), 16-month-old infants were able to learn word-object mappings through co-occurrence statistics only when the multiple occurrences of each word were blocked, but not when exposures to different words were interleaved. Vouloumanos, Martin, and Onishi (2014) found that 18-month-olds could distinguish words that had co-occurred many times with an object from those that had never co-occurred with that object, but could not distinguish words that had co-occurred 8 times with an object from those that had co-occurred twice with it (in contrast to adults, Vouloumanos, 2008). Even for adults, however, this process of statistical inference appears to be highly constrained by limits on memory and attention (Smith, Smith, & Blythe, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Yurovsky & Frank, 2015). As the number of referents available increases, adults track less and less information about each, and their ability to recall this information falls off precipitously with time between exposure and test. In contrast to domains like low-level vision, where human performance is often quite well described by ideal observer models (e.g., Najemnik & Geisler, 2005), human statistical word learning is markedly less efficient than should be expected from a system that makes optimal use of the available information (Frank, Goodman, & Tenenbaum, 2009; Yu & Smith, 2012b; Yurovsky & Frank, 2015). Several recent papers have shown that, under some working assumptions, human-scale lexicons are learnable from statistical dependencies between words and objects from approximately the amount of words heard by typically developing children (Blythe et al., 2010, 2016). However, there is little in the way of guarantees in these models that learning will be rapid (Reisenauer, Smith, & Blythe, 2013; Vogt, 2012), especially under the kinds of memory and attentional constraints found in young infants.

One should not conclude from this data that social cues and statistical cues are not useful for word learning, nor should one conclude that children do not use social cues or do not use statistical information to learn relationships between the words of their native language and the objects in the world. But the discrepancy between children's competence under ideal circumstances and their performance under more challenging circumstances raises a question: Why do children learn words so rapidly even though their learning performance is so constrained? The solution, I will argue, is that our consensus about the computational problem of word learning is incorrect. The right question is not "how do children learn the meanings of words," but rather "how do children and their parents develop systems for communicating successfully?" (Bruner, 1975). Put another way, we often think of the lexicon as the goal and the communicative moments as the tools through which the lexicon is acquired. I propose that we should make progress instead by inverting this relationship: Communication is the goal, and the lexicon is a tool for successful communication.

3. The computational problem of communication

The computational level description of word learning implicitly makes a strange kind of division: It divorces the problem of learning words entirely from the problem of using them; it assumes that the lexicon is a static property of the external world. That is, that there is some objectively "right" mapping between a word and its meaning in the same way that there is a "right" way to walk (c.f. Tomasello, 2001). But these are two very different kinds of problems. The solution for the problem of walking is constrained by biomechanics—the best way to walk is one that minimizes energy expenditure and probability of falling while maximizing distance traversed per unit time. Further, the right way to walk does not

depend on how other people are walking (at least in the absence of other social goals)—it depends on the infant's developing body (Cole, Lingeman, & Adolph, 2012; Garciaguirre, Adolph, & Shrout, 2007). Learning a word is the opposite. The only determiner of the "right" thing to call an object is what other people call it. Language is the solution to a *coordination* problem (Chater & Christiansen, 2010; Schelling, 1980).

This distinction has profound consequences for word learning. One straightforward consequence is that induction in coordination problems is easier because learners biases are more likely to be correct. In any learning problem, the goal is to minimize prediction error: The difference between what the system expects to happen and what actually happens. Error is minimized by using the data available to update the learner's model of the world so that it makes better predictions. But this learning process always faces a tradeoff between errors caused by two opposing forces: bias and variance (Hastie, Tibshirani, & Friedman, 2009). Variance reflects sensitivity to the data—how much different samples of input yield differences in learning. High sensitivity to data is good—after all, learning is fundamentally a process of changing in response to data. However, sensitivity to data also allows small fluctuations in input to have outsized effects on learning (what is sometimes called overfitting). To reduce variance, bias can be imposed on the learning system so that some kinds of data yield less learning. In regression models, for instance, the preference for parsimony employed in removing predictors that do not reach statistical significance is an example of bias. Because these biases reduce sensitivity to the data, they increase resilience to errors that would be caused by noise in the data. However, for the very same reason, bias reduces learning rate by making the system less sensitive to data that would drive learning.

Coordination problems are a serendipitous case where bias and variance may not pull in opposite directions. In the construction of statistical models, researchers use biases that are motivated by theoretical analyses or have behaved well empirically in the past (like the preference for parsimony). In natural learning systems, good biases need to have similarly been tuned over time by evolution to reflect the structure of the natural world. For instance, newborn babies prefer to look at face-like patterns over non face-like patterns (Johnson, Dziurawiec, Ellis, & Morton, 1991), a bias that facilitates attending to and learning to recognize a biologically and socially important stimulus. However, biases in a problem of coordinating with similar agents do not need to be tuned—they are right no matter what they are. This is because in a coordination problem, the goal is to learn the same thing as everyone else, so as long as biases are shared across players they will lead all players in the same correct direction.

While our model of the word learning problem assumes that the relationship between words and their referents is entirely arbitrary, in natural languages this is not the case (Saussure, 1960). Because our biases are the same as the biases of the other people we are coordinating with, these biases are likely to be reflected in the lexicon. We should thus predict, for instance, reliable sound-object relationships in the language around us that we essentially get for free and do not have to learn (Lewis & Frank, 2016; Maurer, Pathman, & Mondloch, 2006; Perry, Perlman, & Lupyan, 2015). These relationships can arise from virtuous cycles across generations of speakers that amplify subtle biases shared by all individual speakers and lead to more learnable languages (Kirby, Cornish, & Smith, 2008). Although there is of course variability in words used to refer to the same object across language communities, they are nonetheless not fully arbitrary, giving learners a leg up.

This same consequence of shared-biases is likely to help us in individual learning moments as well. While theoretical models generally assume that the referent is equally likely to be any of the objects around in the scene, in practice this is unlikely to be true. If people talk about the things that they find interesting, and different people find the same things interesting, then learners already have a leg up on knowing what object is likely to be the referent of a speaker's utterance (Frank & Goodman, 2012, 2014). This shared salience is what drives the disconnect between our lay intuitions and the formal intractability of Quine's (1960) indeterminacy problem. Quine asks the reader to imagine themselves as a field linguist hearing a speaker of a foreign language say "gavagai" while pointing in the direction of a rabbit as it scurries by. While this word could mean "rabbit," the experience is also consistent with an infinite set of other possible meanings (e.g. 'animal', 'white', and 'undetached rabbit parts'). But each of these other possibilities seems intuitively to be unlikely: We think that 'gavagai' means rabbit because rabbit is what we would want to talk about.

Of course infants and their parents need not find all of the same things interesting, and they likely do not. Infants might think that "rabbit" is most interesting, and their parents might be most interested in "dinner." Nonetheless, shared biases do the work of reducing the set of possible meanings of a novel word from an infinite set to a small finite set of plausible candidates. In order to adjudicate among these, infants and their parents need only share a common goal: The desire to communicate (Clark & Schaefer, 1989; Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004; Vouloumanos, Onishi, & Pogue, 2012).

3.1. Language learning in the context of communication

Because the lexicon was constructed by people, its structure depends on people. And because language is produced by people, the referents also depend on people. These dependencies make the inductive problems involved in learning the structure of language easier than the inductive problems involved in learning the structure of locomotion. But the language that an infant hears also depends on them in a more direct way: It is produced to them for motivated reasons. As Gleitman (1990) pointed out, the language that children hear is not a veridical running commentary on their visual world; we rarely come home and say "hello, I am opening the door." But neither is language random—it is motivated by the desire to communicate information (Grice, 1969).

If child-directed speech is intended to communicate information, the referential cues accompanying it should depend systematically on the child. For instance, consider the referential cues like speaker gaze that a child might use to infer the target of a speaker's utterance. These kinds of cues should be informative because speakers are likely to be oriented towards—and looking at—the objects that they are talking about. Indeed, corpus analyses of parent-child interactions show that there is a reliable relationship between the locus of a parent's gaze and the target of her reference (Frank et al., 2013; Yu & Smith, 2012a). However, a better predictor of the parent's reference is where the *child herself* is looking. Even without following their parents' eye-gaze, children would be right more often than not to just assume that their parents are talking about the object that they themselves are focused on (Tomasello & Farrar, 1986). Of course this simple solution would not resolve the discrepant cases, and indeed infants are sensitive to discrepancies between their attentional focus and the focus of their adult interlocutors (Baldwin, 1993). Nonetheless, child directed utterances are fundamentally dependent on the child's own attention, a dependency that does not exist in the standard description of the learning problem.

Similarly, if child-directed speech is goal-oriented—produced in the context of a desire to communicate—it will necessarily have structural features that make it different from the kind of structure that our models assume. It should, for instance, be structured in a way that makes it easier to extract the speaker's intended referent.

One example of this is that utterance-final words are easier to extract and learn from continuous speech (Endress, Scholl, & Mehler, 2005; Yurovsky, 2012). Predictably, parents tend to place target referents at the end of an utterance, even if this makes an utterance ungrammatical (Aslin, Woodward, LaMendola, & Bever, 1996). Further, child-directed speech should not be uniformly simpler, but rather fine-tuned to the particular child who is the recipient of the speech (Snow, 1972). Recent findings show at least two properties of child-directed speech that are tuned in this way.

First, one might intuitively predict that the length of child-directed utterances—which serve as a proxy for complexity—might grow monotonically over development. This turns out not to be the case (Newport, Gleitman, & Gleitman, 1977). But, utterances do vary systematically in a more subtle way. Roy, Frank, and Roy (2009) studied a high-density longitudinal corpus containing all of the speech heard at home by one child in the first three years of his life. They show that caregivers' utterances containing a given word (e.g. 'fish') follow a U-shaped trajectory with its trough centered at the point at which the child learns this word. Caregiver utterances are at their longest well before the child produces the word for the first time, become shorter in the months before the word is learned, and then lengthen once again. That is, utterances containing a word are at their easiest to process right around the period of time when the child is learning them—perhaps because caregivers are shortening them in response to the child's changing interest in the objects to which they refer.

Second, although utterances to young children are not systematically simpler, they are systematically more contingent. Yurovsky, Doyle, and Frank (2016) estimated the degree to which parents linguistically align to their developing children—contingently reproducing function word categories in their children's previous utterances. This analysis shows a high degree of alignment early in development that declines steadily over the course of the first 5 years, providing evidence that parents are contingently modifying their utterances on those of their children in a way that supports communication. These findings are consonant with earlier analyses of reformulations and expansions in child-directed speech (Chouinard & Clark, 2003; Saxton, 2000). Parents' speech depends fundamentally on children's own speech in a way that scaffolds them in order to maintain a consistent conversation.

Finally, the communicative context of language learning changes not just expectations for input and learnability, but also for what it means to "know" a word. Because the standard computational framework makes learning the lexicon the goal of language acquisition, many models are tested by comparing the lexicons they have inferred to a "gold standard" lexicon (e.g., Fazly, Alishahi, & Stevenson, 2010; Frank et al., 2009; Yu, 2008). And similarly, experimental participants are tested for their ability to select the right referent when cued with the right word (e.g., Smith et al., 2011; Yu & Smith, 2007; Yurovsky et al., 2014). But this is drastically different from the way that real children's language knowledge is tested outside the laboratory. What it means for a child to know a word is for that child to be able to use language to communicate successfully about its referent. We even know this fact implicitly as scientists, for instance, when we use the MacArthur Child Development inventories to measure children's knowledge. These parent-report vocabulary measures are a standard instrument in the field of developmental psychology, and a tool used by clinicians to assess children's language development. When we ask parents whether their children know the words on these forms, we instruct them that if their child "uses a different pronunciation of a word (for example, 'raffe' for 'giraffe' or 'sketti' for 'spaghetti'), mark the word right anyway" (Fenson et al., 2007).

The communicative nature of the learning context fundamentally changes the problem being solved. One way of diagramming this change is by formalizing the learning problem in a graphical model—a description of the statistical dependencies between the variables relevant for the learning problem. Fig. 1 shows three such models. In each model, the solid gray circles indicate observed variables—variables whose value the learner can observe directly. The hollow circles show latent variables—variables which have causal consequences for observed variables, but which cannot be observed directly, only inferred from the observed variables. The first model shows the framework implicit in many models of statistical learning. In each situation, the learner observes objects and hears words, and the words come from an unobserved lexicon that is the target of inference. The second model shows a framework proposed by Frank et al. (2009) that adds a second latent variable to the situation—the speaker's intention. Inference in this model is more powerful because it leverages a fundamental dependency between ambiguity in individual situations and ambiguity across situations. In each situation, the words a speaker produces are mediated by an intention to refer to only some of the objects. That means that if a learner can discover a speaker's intention, the learner need not consider mappings between the words and any of the unintended objects, reducing spurious correlations. Similarly, if a learner already knows some mappings, it is easier to discover a speaker's intention—because the learner can make a good guess about which object is being labeled. The third model instantiates the proposal in this article. In this model, the learning situation consists not just of a speaker, but rather a dyad—a parent and a child. Critically, the parent's intention to refer depends not only on the objects, but also on the parent's inferences about the child's intentions. The parent's goal is to communicate information that is interesting to the child, and thus a parent's intentions are partially predictable from a child's own words and intentions. Learning in this framework is even more efficient, because the child's own intentions are informative about the meaning of the parent's words. In this model, the child can not only make a good guess about a parent's referent if they know their parent's intention and vice versa, they can make a good guess about both on the basis of their own previous intentions.

3.2. Optimal for communication need not be optimal for learning

I have argued in this paper that much of the credit for children's

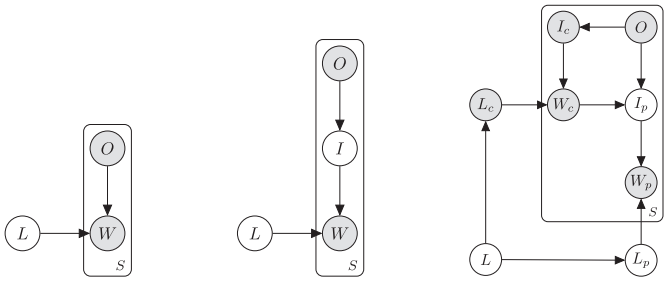

Fig. 1. Three computational-level descriptions of language acquisition. In the left-most model, each situation (S) contains some observed objects (O) that generate observed words (W) by their mapping through an unobserved lexicon (L). This is the framework implicit in many statistical approaches to word learning (e.g. Siskind, 1996; Yu, 2008; Yurovsky et al., 2014). In the middle model, words are generated by an unobserved intention (I) on the part of the speaker. This model allows the learner to leverage the inherent synergy between resolving uncertainty in-the-moment and uncertainty across multiple instances (Frank et al., 2009). The right model incorporates the communicative nature of the learning situation, noting that the parents' intention (I_p) depends not only on the objects, but also on their inferences about what the child is interested in (through their observed words W_p). The child in this model could learn by leveraging this information. Note that the child's lexicon (L_c) and intention (I_c) are observable to the child, even though they are not observable to the parent.

rapid word learning is due to parents rather than to children themselves. This is because the language that the child is learning is not a static, independent property of the world that is indifferent to the child's goals. Rather, it is statistically dependent on the child in two key ways. First, in the indirect sense that the target lexicon arose from the interactions of people with the same biases as the child. And second, in the more direct sense that the input the child hears depends on the parent's inferences about the child's intentions in the moment. The second half of the argument requires that the child's caregiver have a goal that makes the child's process of learning easier: The goal to communicate information. Because the child will need to be scaffolded for communication to succeed, language input will be easier to learn from.

Importantly, although the parent's speech needs to be goal directed, the goal does not need to be teaching. There may well be times when the parent's goal is to teach the child words, and these times are likely to be particularly informative about the meanings of words. Speech that is pedagogical licenses even stronger inferences because absence of evidence can be taken to be evidence of absence. For instance, if the parent chooses examples of a category with the intention to teach, these examples can license strong inferences about the extent of that category (e.g. a parent would be unlikely to choose three peppers to teach the child *vegetable*; Xu & Tenenbaum, 2007). Parents, particularly of children from some cultural backgrounds, may be motivated by pedagogical goals often, and children might be particularly adept at inferring when parents have these goals (Csibra & Gergely, 2009). But the argument in this paper does not require this strong assumption.

Indeed, there are many cases where optimizing for learning and optimizing for communication will be at odds (Kirby, Tamariz, Cornish, & Smith, 2015). Much of the early enthusiasm for the hypothesis that parents tune their speech in a way that helps children ran aground of this phenomenon. For instance, while parents often provide corrective input when children make semantic errors (e.g. calling a *dog* 'horse'), they tend not to provide straightforward corrective signals for syntactic errors. This is because semantic errors can produce failures in the communication, but syntactic errors generally do not (Brown, 1977; Newport et al., 1977). Similarly, if child-directed speech is optimized for learning, then parents should gradually increase the complexity of the words they produce to children over development (Elman, 1993). However, both early analyses of parental speech and more recent followups have shown that average complexity of words in child-directed speech remains relatively constant over development (Hayes & Ahrens, 1988; Yurovsky et al., 2016). This result is not surprising, however, if language is instead optimized for communication: most of the words used in typical communicative contexts are simple, and parents' goal is not to optimize for their child's learning.

If child-directed speech were optimized for learning, it would certainly make learning easier. But we should not take evidence that speech is not optimal for learning as evidence that it is not *better* for learning than it would be if it were not communicative (Eaves, Jr, Feldman, Griffiths, & Shafto, 2016; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Coordinating in-the-moment is only a piece of the language learning puzzle, but it is an important one (Tomasello, 2000). Understanding how both parent and child contribute to this coordination is critical for understanding why children's language learning is so rapid despite the gap between their competence and their performance in many of the relevant cognitive processes. In the final section, I flesh out this case, providing an example of a simpler computational problem that makes it easier to see why optimal for communication can be different from optimal for learning, but nonetheless be much better for learning than speech that is not intended to communicate.

4. The computational problem of searching an array

As an analogy, let us consider a simpler model system: The problem of searching an array. Suppose that you have been given an array of integers of some length n. And suppose that you can look at any of the integers you want one at a time. You are then asked about some particular integer i, and your goal is to determine whether i is somewhere in your array. How difficult would that be? To define difficulty, we will use a classic measure from computational complexity theory: The number of operations required. In this framework, every time we need to look one of the integers in the array, we pay a cost. The question of difficulty then becomes a question of how many of the n integers we will have to look at to guarantee that we can determine whether i is in the array.

Let us consider first the worst case scenario. The worst thing that could happen is that you have to look at every number in the array exactly once. No matter what strategy you have for checking the indices of the array—no matter the algorithm—the number could always be in the last place you look. If you start at the beginning, the number could be at the end. If you start at the end, the number could be at the beginning. And you can never stop early, because if there are any indices you haven't looked at yet, i could be in one of them. Therefore, the computational problem of searching this kind of array is said to have complexity of $O(n)$ —it scales linearly with the length of the array. This kind of worst-case analysis is often brought to bear in our descriptions of language as an adversarial problem, one in which the world is as unhelpful as possible (e.g. Blythe et al., 2010; Gold, 1967).

In contrast, consider the best case scenario. Suppose that the person who gives you the array and asks you the question is one and the same, and they also know your search strategy (and you know that they know). In that case, they might pick an order for the items in the array that makes the very first place you look the only index you need to examine. Either the number they are looking for is in that first index, and you can say yes. Or it is not, and you can say no, knowing that it is also not in any of the other indices. In this case, the length of the array does not matter, and the problem is said to have a complexity of $O(1)$ —it requires one operation no matter how many numbers are in the array. This is the kind of best-case analysis involved in descriptions of language as pedagogy—analyses under which the adult's goal is to structure input to maximize learning (Bonawitz et al., 2011; Csibra & Gergely, 2009; Shafto, Goodman, & Frank, 2012).

Finally, let us imagine a third scenario. Suppose that we make one key change in the set up of the problem: The array of integers you get is always sorted in ascending order. Now the worst case analysis is less bad, because you can apply a novel strategy: Binary search. You begin by looking at the integer in the middle of the array (index $\frac{n}{2}$). If this is larger than the number you are looking for, you no longer have to look at any of the numbers to the right of it. If it is smaller than the number you are looking for, you no longer have to look at any of the numbers to the left of it. You can then apply this strategy recursively, finding the number in the middle of the remaining half, again removing from consideration half of the remaining numbers. Now this problem is much easier. Even in the worst case, the number of indices you need to look at scales logarithmically the length of the array: $O(log_2 n)$ (Knuth, 1998).

Now let us suppose that we are conducting empirical research to understand the processes of list search. Unknown to us, the third scenario is the true state of the world (sorted arrays), but we believe apriori that we are observing the first scenario (unsorted arrays). What will happen? We first go out into the world and observe a group of searchers performing search in their natural ecology. We look at the lengths of the list they search, and we look at how long it takes them, and we are surprised to discover that they search

incredibly fast! So we bring these searchers into our labs, measure their search times on unsorted arrays, and discover that they are not terribly good at searching. Scratching our heads about why these searchers are so effective in their natural ecology while being so poor in lab, we decide that we must have been thinking about the problem all wrong: Perhaps the searchers' partners are trying to optimize their searching! We go back into the world to determine if we are actually in the second scenario—where best case and not worst case analysis is correct—and discover that the answer is no. At the end of all of this, we will end up in the puzzle where this paper began: Searchers are seemingly highly effective in the wild, not very good in the lab, and yet the discrepancy does not seem to come from a goal on the part of their partner to optimize their searching.

I argue that this is exactly the place in which we find ourselves in our formal analysis of how children learn the meanings of words. The key to resolving this puzzle is to understand that even if the language that children hear is not optimized for learning, it is also not random; it may be optimized for a related goal: communication. There may well be problems of language learning—especially in syntax—where these goals at orthogonal or even pull in opposite directions (Brown, 1977; Moerk, 1989; Newport et al., 1977). However, for the problem of learning the meanings of words, these two goals are likely to be highly aligned (Frank & Goodman, 2014).

5. Conclusion

Young children learn the meanings of thousands of words by the time they can run down the street (Mayor & Plunkett, 2011). The computational problem they solve is daunting: extracting discrete word forms from a sequence of continuous speech signals and mapping these forms onto their meanings. The explanation for this rapid learning has tended to come in the form of an appeal either to a precocious capacity to use social cues to rapidly infer speakers' intended meaning, or alternatively to a powerful capacity to learn from statistical relationships between words and objects in the world. Yet, the same children who solve this problem continuously forget where they left their coats and hats. How do children learn language so quickly despite their cognitive constraints?

The solution to this puzzle is to consider a second critical part of the language learning system: The parent. Although children are inundated with language from many sources—including overheard speech—much of their learning seems to be driven by the portion of their language input that is child-directed (Weisleder & Fernald, 2013). Decades of observational research have leveraged this idea, describing the ways in which parents talk to their children differently from the way that they talk to adults, and to trying to understand which of these differences support learning (e.g., Bloom, Margulis, Tinker, & Fujita, 1996; Moerk, 1989; Pan, Rowe, Singer, & Snow, 2005). My goal in this paper has been to try to integrate the insights of this body of work with the progress made by computational analyses of language learning. The core argument is that there is a unifying cause of these structural differences in speech to children, and that this cause is a part of the computational level description of the problem that children are trying to solve (Marr, 1982). The structure of this speech is different in a fundamental way than overheard speech, not just because it is in some ways simpler, but also because it depends on the child herself. Language directed at the child is purposeful—it is intended to communicate. If children are aware that this speech is communicative, their learning problem is radically simpler. Because children appear to be sensitive to the communicative nature of speech from quite early on, there is reason to be hopeful that this framework will give us a wedge into resolving our paradox (Vouloumanos et al., 2012, 2014). If there is anywhere in language acquisition where there is hope for finding optimality, it is not in the child's head, but in the coordination of the child-parent system.

References

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 117–134.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*, 395–418.

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

Bloom, L., Margulis, C., Tinker, E., & Fujita, N. (1996). Early conversations and word Learning: Contributions from child and adult. *Child Development, 67*, 3154–3175.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science, 34*, 620–642.

Blythe, R. A., Smith, A. D., & Smith, K. (2016). Word learning under infinite uncertainty. *Cognition, 151*, 18–27.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*, 322–330.

Brown, R. (1977). Introduction. In C. E. Snow, & C. A. Ferguson (Eds.), *Talking to children: Language input and interaction*. Cambridge, MA.: MIT Press.

Bruner, J. S. (1975). From communication to language—a psychological perspective. *Cognition, 3*, 255–287.

Bruner, J. (1983). *Child's talk: Learning to use language*. Norton.

Capaday, C. (2002). The special nature of human walking and its neural control. *Trends in Neurosciences, 25*, 370–376.

Caselli, M. C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanderl, L., et al. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10*, 159–199.

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

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

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language, 30*, 637–669.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Cole, W. G., Lingeman, J. M., & Adolph, K. E. (2012). Go naked: Diapers affect infant walking. *Developmental Science, 15*, 783–790.

Collins, S., Ruina, A., Tedrake, R., & Wisse, M. (2005). Efficient bipedal robots based on passive-dynamic walkers. *Science, 307*, 1082–1085.

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology, 34*, 28–38.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*, 148–153.

Eaves, B. S., Jr., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review, 123*, 758.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134*, 406.

Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science, 34*, 1017–1063.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates communicative development inventories*.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science, 336*, 998–998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology, 75*, 80–96.

Frank, M. C., Goodman, N., & Tenenbaum, J. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*, 578–585.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development, 9*, 1–24.

Garciaguirre, J. S., Adolph, K. E., & Shrout, P. E. (2007). Baby carriage: Infants walking with loads. *Child Development, 78*, 664–680.

Garwicz, M., Christensson, M., & Psouni, E. (2009). A unifying model for timing of walking onset in humans and other mammals. *Proceedings of the National Academy of Sciences, 106*, 21889–21893.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. technical report no. 257. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10*, 447–474.

Grice, H. P. (1969). Utterer's meaning and intention. *The Philosophical Review, 78*, 147–177.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised learning. In *The elements of statistical learning* (pp. 485–585). Springer.

Hayes, D. P., & Ahrens, M. G. (1988). Vocabulary simplification for children: A special

case of 'motherese'? *Journal of Child Language, 15*, 395–410.
Hockett, C. F. (1959). Animal "languages" and human language. *Human Biology, 31*, 32–39.
Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., et al. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*. pp. i–135.
Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1–19.
Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bulletin & Review, 19*, 317–324.
Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences, 105*, 10681–10686.
Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition, 141*, 87–102.
Knuth, D. E. (1998). *The art of computer programming: Sorting and searching* (Vol. 3). Pearson Education.
Lewis, M. L., & Frank, M. C. (2016). The length of words reflects their conceptual complexity. *Cognition, 153*, 182–195.
Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*, 297–307.
Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception, 37*, 1443–1457.
Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York, NY: W. H. Freeman.
Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound–-shape correspondences in toddlers and adults. *Developmental Science, 9*, 316–322.
Mayor, J., & Plunkett, K. (2011). A statistical estimate of infant and toddler vocabulary size from cdi analysis. *Developmental Science, 14*, 769–785.
McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*, 831.
McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition, 129*, 362–378.
Moerk, E. L. (1989). The LAD was a lady and the tasks were ill-defined. *Developmental Review, 9*, 21–57.
Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*, 60.
Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387–391.
Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. A. Ferguson (Ed.), *Talking to children Language input and interaction* (pp. 109–149). Cambridge University Press.
Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development, 76*, 763–782.
Perry, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in english and Spanish and its relation to lexical category and age of acquisition. *PLoS One, 10*, e0137147.
Quine, W. V. O. (1960). *Word and object*. Cambridge, Mass: MIT Press.
Reisenauer, R., Smith, K., & Blythe, R. A. (2013). Stochastic dynamics of lexicon learning in an uncertain and nonuniform world. *Physical Review Letters, 110*, 258701.
Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. In N. Taatgen (Ed.), *Proceedings of the 31st annual meeting of the cognitive science society*. Amsterdam: Cognitive Science Society.
Saussure, F. (1960). *Course in general linguistics*. London: Peter Owen.
Saxton, M. (2000). Negative evidence and negative feedback: Immediate effects on the grammaticality of child speech. *First Language, 20*, 221–252.
Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature, 253*, 265–266.
Schelling, T. C. (1980). *The strategy of conflict*. Harvard University Press.
Scott, R. M., & Fischer, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition, 122*, 163–180.
Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others the consequences of psychological reasoning for human learning. *Perspectives on Psychological Science, 7*, 341–351.
Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39–91.
Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science, 35*, 480–498.
Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558–1568.
Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development, 43*, 549–565.
Squires, J., Bricker, D. D., & Twombly, E. (2009). *Ages & stages questionnaires: A parent-completed child monitoring system*. Paul H. Brookes Publishing Company.
Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology, 126*, 395–411.
Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics, 10*, 401–413.
Tomasello, M. (2001). Could we please lose the mapping metaphor, please? *Behavioral and Brain Sciences, 24*, 1119–1120.
Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*, 705–722.
Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.
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*, 126–156.
Vida, M. D., & Maurer, D. (2012). The development of fine-grained sensitivity to eye contact after 6years of age. *Journal of Experimental Child Psychology, 112*, 243–256.
Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants cross-situational statistical learning. *Cognition, 127*, 375–382.
Vogt, P. (2012). Exploring the robustness of cross-situational learning under zipfian distributions. *Cognitive Science, 36*, 726–739.
Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition, 107*, 729–742.
Vouloumanos, A., Martin, A., & Onishi, K. H. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science, 17*, 872–879.
Vouloumanos, A., Onishi, K. H., & Pogue, A. (2012). Twelve-month-old infants recognize that speech can communicate unobservable intentions. *Proceedings of the National Academy of Sciences, 109*, 12933–12937.
Vouloumanos, A., & Werker, J. F. (2009). Infants learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611.
Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*, 2143–2152.
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*, 245–272.
Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development, 4*, 32–62.
Yurovsky, D. (2012). Statistical speech segmentation and word learning in parallel: Scaffolding from child-directed speech. *Frontiers in Psychology*, 374.
Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th annual meeting of the cognitive science society* (pp. 2093–2098). Austin, TX: Cognitive Science Society.
Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145*, 53–62.
Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science, 20*, e12349.
Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review, 21*, 1–22.
Yurovsky, D., Wade, A., & Frank, M. C. (2013a). Online processing of speech and social information in early word learning. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1641–1646).
Yurovsky, D., Yu, C., & Smith, L. B. (2013b). Competitive processes in cross-situational word learning. *Cognitive Science, 37*, 891–921.
Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.
Yu, C., & Smith, L. B. (2012a). Embodied attention and word learning by toddlers. *Cognition, 125*, 244–262.
Yu, C., & Smith, L. B. (2012b). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*, 21–39.