A distributional perspective on the *gavagai* problem in early word learning☆

Richard N. Aslin [a,b,c,d,*], Alice F. Wang [a]

[a] *Haskins Laboratories, New Haven, CT, United States of America*
[b] *Yale Child Study Center, New Haven, CT, United States of America*
[c] *Department of Psychology, Yale University, New Haven, CT, United States of America*
[d] *Department of Psychological Sciences, University of Connecticut, Storrs, CT, United States of America*

A R T I C L E   I N F O

A B S T R A C T

Word learning entails the mapping of an auditory word-form to its appropriate grammatical category (e.g., noun, verb, adjective), but before that mapping can occur, the naïve learner must infer which of the myriad of possible referents of that word was intended by the speaker. This creates a computational explosion of referential ambiguity referred to as the *gavagai* problem. In a set of corpus analyses of parent-directed speech to young infants, we describe the distributional information available to early word learners, with a focus on nouns and adjectives that refer to whole objects and object properties. And in two experiments on word-learning in adults spanning seven different distributional conditions, we document how variations in the ratio of novel labels for objects and properties affect the robustness of word learning. Our results suggest that the language input to 6- to 20-month-olds is robustly populated with high-frequency object words and high-frequency property words, but their co-occurrence is sparse. Although this distributional information slightly favors object words over property words, a more plausible account of the whole-object bias in early word learning is the inability to encode the details of an object/event during rapid naming. Our results from adults, presented with novel labels for multi-referent objects in a cross-situational statistical learning paradigm, also reveal this whole-object bias as well as the absence of property-label generalization to novel objects, even when the distribution of labels is shifted almost exclusively to property words. These results are discussed in terms of the relative ease of mapping auditory word-forms to whole objects vs. object properties, thereby limiting the combinatorics of the *gavagai* problem, especially in infants with immature encoding and memory representation abilities.

1. Introduction

One of the hallmarks of language acquisition is that children's first productions consist of content words – they label objects and events that are common in their immediate environment (Brown, 1973). Interestingly, infants readily distinguish between function words and content words as two separate classes long before they produce their first words (Marino, Bernard, & Gervain, 2020; Shi, Werker, & Cutler, 2006; Shi, Werker, & Morgan, 1999), in part because function words occur much more frequently than content words. Furthermore, Hochmann (2013) has shown that 7-month-olds treat extremely high-frequency auditory word-forms as less referential than word-forms of medium frequency. Nevertheless, the primary vehicle for understanding and conveying information to a communicative partner about a *referent* consists of nouns, adjectives, or verbs (roughly in that order). Moreover, when older

infants are taught a new word, they more quickly and robustly map that auditory word-form to the whole object (e.g., dog) than to a property of that object (e.g., *furry*; see MacNamara, 1972). The absence of function words in early child productions could be the result of their generally lower salience than content words – they are shorter in duration, unstressed, and co-articulated with their surrounding context – even though they are much more frequent than content words, suggesting that their ubiquity is insufficient to motivate their use. In contrast, the lower propensity of producing property words and mapping them onto novel auditory word-forms cannot be salience-based – the sound of *blue* is no less salient than the sound of *ball*. Thus, it is natural to look toward a distributional explanation to account for this difference. And, indeed, the fact that toddlers who fail to produce function words are nevertheless sensitive to their deletion in comprehension (Gerken, Landau, & Remez, 1990), suggests that distributional information is accessed prior

---

to production.

There is ample evidence that infants have access to robust distributional properties of linguistic input in natural-language corpora (Batchelder, 2002; Swingley, 2005) and can extract some of those statistics in artificial-language experiments, including at the level of phonetic categories (Maye, Werker, & Gerken, 2002), phonotactics (Chambers, Onishi, & Fisher, 2003), word segmentation (Saffran, Aslin, & Newport, 1996), and the frequency of function words as a cue to novel noun learning (Hochmann, Endress, & Mehler, 2010). One version of the distributional hypothesis for the whole-object bias is that property words are simply less frequent than basic-level object words in everyday parental language input to their infant. Moreover, perhaps *blue* is not only less frequent than *ball* overall but *blue* is less frequently paired with *ball* than all instances of *ball* (i.e., the conditional probability of *blue* given *ball* is low). But, of course, spoken language is not presented to infants devoid of visual context. Thus, another version of the distributional hypothesis is that parents, and infants themselves, structure their referential world to bias the mapping of auditory word-forms to whole objects rather than to object properties (Yurovsky, Smith, & Yu, 2013). When a parent holds up an object and elicits the infant's visual attention to it, the most "transparent" inference is that the spoken word refers to the whole object. A similar but less overt constraint on attention to a named object occurs when the infant rather than the parent is holding the object themselves. Regardless, the naming is correlated with visual attention to the object, thereby reducing the set of likely referents (see Smith, Jayaraman, Clerkin, & Yu, 2018).

But the two foregoing versions of the distributional hypothesis rely on what the parent says rather than on why the infant infers that the referent of that spoken word is the whole object and not one of its many properties. This is the *gavagai* conundrum (Quine, 1960) – given a novel word and a novel event in the world, why does the infant infer that the word refers to the whole object (e.g., a rabbit) and not some property of the event (e.g., hopping, white, furry, disembodied rabbit parts)? This conundrum is not resolvable in a straightforward way even though to most adults it appears to be trivial. What else would a listener infer when hearing "Oh, gavagai" than the obvious fact that "gavagai" is a synonym for rabbit? Unfortunately, intuitions are often faulty when dealing with sophisticated users of a particular language. In what has been called the Human Simulation Paradigm, Gleitman and colleagues (Gillette, Gleitman, Gleitman, & Lederer, 1999; Medina, Snedeker, Trueswell, & Gleitman, 2011) asked adults to report what a word (simulated by a "beep" inserted into a video of common everyday scenes) likely referred to. When the beep occurred as a person was hammering repeatedly, the beep was sometimes inferred as referring to the subject (hammer), the verb (act of hammering), and only occasionally as the object (thing being hammered) or a property of the event (heavy hammer, rapid/slow hammering). While there was, in fact, tremendous ambiguity about how to link the auditory word-form (simulated by the beep) with the referent of that word-form, the most common interpretation was nevertheless that the word referred to the whole object. Importantly, this ambiguity was also present when adults viewed videos of 14- to 18-month-olds interacting with their parents in natural settings (Trueswell et al., 2016), thereby confirming that even in simpler contexts typical of early development the presence of word-referent ambiguity is ubiquitous.

A common explanation for this bias to treat auditory word-forms as referring to the whole object consists of moving out of the linguistic domain and into the conceptual domain. That is, the whole-object bias in word-naming and word-learning could in essence be a conceptual bias to attend primarily to the object rather than to its myriad properties (Spelke, 1990). In turn, if parents share this conceptual bias, they may reinforce what for their infant begins as an initially small bias when a word is spoken in the presence of an ambiguous event, thereby enhancing the strength of the whole-object bias compared to an only slightly less salient property bias. Thus, we come back full circle to the second version of the distributional hypothesis – parents drive both sides of the equation by eliciting attention to an object (or waiting for the

infant to do so) and speaking the name that labels that whole object. It would be infelicitous for parents to do otherwise by, for example, consistently referring to the family pet as *furry*, because that property is not unique to that specific pet's basic-level category.

But even if parents consistently timed their utterances to take advantage of how the infant was allocating their attention on a moment-by-moment basis (which they clearly do not), how would the infant know to which of the many potential object properties that word-form should be linked? And even if an infant happened to be holding in their own hand a multi-dimensional object and the parent said "Look at the gavagai", how would the infant know that "gavagai" meant rabbit, white, or furry? A paradigm developed by Yu, Smith, and colleagues (see review by Smith, Suanda, & Yu, 2014), called cross-situational statistical learning, provided a partial answer to this question. If an auditory word-form was used in the presence of two objects, and across a series of scenes that same word-form was used to label scenes containing object A and not presented for scenes containing object B, then by a "process of elimination" adults and infants could generate a most likely hypothesis about which word-form referred to which object. Notice, however, a key difference between the cross-situational statistical learning paradigm and the Human Simulation paradigm – the former employs multiple instances of auditory-visual pairings that have a consistent correlation, whereas the latter relies on a single instance to elicit a causal explanation.

It is important to note that distributional information is not the only potential mechanism that could enable infants to infer whether a label refers to an object or a property. Infants appear to have a preference for attending to shape and size over pattern and color for objects undergoing brief occlusion (Wilcox, 1999) and they are less likely to rely on color than shape in a word-referent mapping task (Graham & Poulin-Dubois, 1999; Kandhadai, Hall, & Werker, 2017). However, these tasks do not provide extensive training to examine the role of distributional learning. There is also a long history of studies in older infants and toddlers that document the utility of *mutual exclusivity* as a way to link a novel auditory word-form to an object (Markman, 1990). And a variety of *syntactic bootstrapping* mechanisms have been shown to enable toddlers and young children to assign a novel label to the correct part of speech (Gleitman, 1990; Waxman & Booth, 2001). But in both of these latter cases, the additional mechanism that reduces ambiguity in linking a label to its referent do not appear to play a role until infants are well into their second year of life (Bion, Borovsky, & Fernald, 2013; Halberda, 2003).

The key question that motivates the present line of research is whether there is evidence in parental speech to young infants – in the early phase of word learning – that supports either version of the distributional hypothesis. To that end, we first examined language input to infants from two databases: CHILDES (https://childes.talkbank.org/) and SEEDLingS (Bergelson, 2017). We extracted all instances of the most frequent nouns as well as the most frequent verbs, prepositions, and adjectives. We then asked whether the spoken-word contexts for each of these parts-of-speech were stable enough to allow infants to extract common bigrams or trigrams (surrounding frames) from which the property names could be learned. For example, if *ball* was a high-frequency noun, we asked whether there were words immediately before or after *ball* that were also high enough in frequency to enable that adjacent word to be learned. In general, the answer was no. Specifically, the frequency of words surrounding a noun was either high (e.g., function words) or so low that it was implausible as a robust distributional source of information for a property word to be learned. This prompted us to ask whether the whole-object bias in word-learning can be overcome if we simply change the statistics of the corpus. In two online experiments with adults, we confirmed that learning property names and object names is readily achieved in a cross-situational statistical learning paradigm. However, performance for these two types of words did not mirror the statistics of their presentation during the learning phase. Moreover, even when property names far outnumbered

object names, object names were easily learned and property names were not learned better. And adults readily generalized object names to novel exemplars, but they did not readily generalize property names. While this does not provide definitive evidence for an intrinsic conceptually-based whole-object naming bias – because that bias might have been induced by early experience with objects and their labels – it does suggest that once formed, the whole-object bias is so robust that even extremely counter-biased property statistics cannot overcome it. Similar findings have been observed for non-native phonetic categories: adults have entrenched native-language phonetic categories that are more plastic during an early period of learning and can only be altered in adulthood after extensive training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999).

2. Corpus analyses of lexical distributions in parental speech

The fundamental question addressed by our corpus analyses is whether word frequency and/or the surrounding context of a word could serve as a reliable cue to infer whether that word was referring to the whole object or to a property of that object. We know from previous corpus analyses and novel word-learning studies, in both infants and adults, that the immediately surrounding context – including bigrams before or after the target word or the before-and-after frame that surrounds the target word – provides sufficient information to enable assignment of that word to a grammatical category (Chemla, Mintz, Bernal, & Christophe, 2009; Mintz, 2002, 2003; Mintz, Wang, & Li, 2014). However, this *frequent frames* hypothesis has never been applied to the question of resolving the ambiguity about object name vs. property name in young infants. Mintz (2005) reported evidence of novel property word-learning in 2- and 3-year-olds, but those results relied on children having at least minimal prior knowledge of grammatical categories (e.g., noun).

The basic idea motivating our corpus analysis was that three separate distributional cues might combine to provide infants with a mechanism for solving the *gavagai* problem. First, infants even in the first postnatal year rapidly assign content words and function words into different categories (Shi et al., 1999; Shi et al., 2006). Second, although young infants have auditory working-memory limitations (Benavides-Varela & Mehler, 2015), they are able to rank order a small subset of highly frequent content words from relatively infrequent content words (Vouloumanos & Werker, 2009). And third, a lower-frequency content word that was consistently paired with a higher-frequency content word, by a process of cross-situational statistical learning, could provide the infant with sufficient information to infer that the lower-frequency word referred to a property of the object.

2.1. Materials and methods

We conducted analyses on the Providence corpus from the CHILDES database, which contains transcriptions of spontaneous interactions between six monolingual English-speaking children (Alex, Ethan, Lily, Naima, Violet, and William) and their parents at home (Demuth, Culbertson, & Alter, 2006). We included only transcripts in which the target children were 20 months of age or younger, resulting in a total of 92 transcripts. We capped this age range at 20 months because this is the period of most rapid vocabulary growth (McMurray, 2007). Only the parents' utterances were analyzed, as we were interested in characterizing the linguistic input that infants received. Analyses were conducted using a combination of CLAN and Python. Our sample consisted of 7201 word types and 301,672 word tokens, of which 3184 types were nouns (63,198 tokens) and 799 types were adjectives (14,324 tokens).

We first extracted the most frequent words from the grammatical categories of noun, preposition, adjective, and verb across transcripts for each infant, then picked tokens that were within the top 20 most common across infants (Table 1). Then, the frequencies of different frames surrounding each of these tokens were recorded, with a frame defined as

Table 1
High frequency words by grammatical category.

Part of speech	Selected tokens
Noun	baby, ball, book, mommy
Preposition	in, on, to, with
Adjective	big, little, good, red
Verb	go, have, see, want

the pair of context words that appeared immediately before and after a token (Mintz, 2003). Frames could not cross utterance boundaries. In addition to frame frequencies, we recorded the frequencies of individual words that co-occurred with the tokens, coming either directly before or after the token word (within an utterance).

We also examined the SEEDLingS database, which contains video and audio recordings taken monthly of infants living in the upstate New York area (Bergelson, 2017). We used the 6-month-old sample for our analysis, which includes hour-long video recordings and day-long LENA audio recordings from 43 infants in their home environments. We extracted frames for all instances of *ball* and *book* for both audio and video recordings, including two words immediately before and two words after the token, and recorded the frequencies of unique words in each position.

2.2. Results

Table 2 shows the top-20 most frequent bigrams and most frequent-frames for the four target nouns in the CHILDES samples. As expected, the most frequent surrounding contexts for these nouns consisted of function words, which as discussed earlier, are treated as a separate auditory category from content words. The conditional probability of the most frequent bigrams containing a property word (*little baby, red ball, baby book, purple mommy*) given the target word (*baby, ball, book, mommy*) was 0.048, 0.055, 0.039, and 0.022, respectively, and none of the top-20 frequent frames contained a property word. Notably, this low conditional probability is not the result of property words being much less frequent overall than object words. The four most common property words (*big, little, good, red*) occurred 230, 342, 375, and 107 times in our CHILDES samples (7.36% of the adjectives in our sample), while the four most common object words (*baby, ball, book, mommy*) occurred 376, 348, 462, and 225 times (2.23% of the nouns in our sample).

If the low conditional probability of property words given an object word is a viable account of the object-word bias, then the inverse relation – the conditional probability of object words given a property word – should be substantially higher. Table 3 shows the top-20 most frequent bigrams and most frequent-frames for the four target property words in the CHILDES samples. Again, the most frequent surrounding contexts for these property words consisted of function words, which as discussed earlier, are treated as a separate auditory category from content words. The conditional probability of the most frequent bigrams containing an object word (*big girl, little piggy, good job, red ball*) given the property word (*big, little, good, red*) was 0.065, 0.069, 0.069, and 0.103, respectively. While the mean of the top four conditional probabilities for property words given object words is 0.041 and for object words given property words is 0.076, it is not clear that this twofold difference, even though in the predicted direction for an object-word bias, is sufficiently large to support the object-word bias unless the sample of parental speech was extensive.

A similar pattern was observed in the SEEDLingS samples (see Tables A1 and A2 in the Appendix). The two most frequent bigrams for the target word *book* (n = 420) were *baby* and *truck*, and the two most frequent bigrams for the target word *ball* (n = 241) were *beach* and *green*, with conditional probabilities for property words given object words of 0.019 and 0.017 for *book* and 0.058 and 0.008 for *ball*, respectively.

Table 2
Top 20 frequent frames for target nouns from CHILDES. Underscore indicates position of the target word. Absence of a word before or after an underscore indicates utterance boundary.

Baby	Freq	Ball	Freq	Book	Freq	Mommy	Freq
the_	135	the_	110	a_	91	to_	45
a_	68	a_	75	the_	82	want_to	39
little_	18	your_	37	this_	74	with_	23
my_	15	red_	19	that_	56	for_	16
the_in	15	big_	16	your_	26	that's_	12
that_	14	little_	14	another_	23	there's_	9
that_up	12	my_	10	baby_	18	the_	9
that_has	11	Koosh_	7	a_about	15	up_	8
the_a	10	a_too	6	the_about	12	_has	8
yes_	10	the_with	6	favorite_	8	does_love	7
your_book	9	that_	6	big_	7	tell_what	6
your_	9	the_in	6	train_	7	said_	6
hi_	8	beach_	6	the_to	6	a_	6
a_in	8	the_is	5	the_is	6	I'm_	5
say_	7	the_on	5	a_to	6	purple_	5
the_has	6	the_ride	5	different_	5	and_	5
the_book	6	her_	5	what_do	5	Elmo's_	4
that_all	5	mommys_	4	that_is	5	give_a	4
the_bunny	5	his_	3	a_for	5	on_	4
a_cow	5	runaway_	3	little_	5	can_have	4

Table 3
Top 20 frequent frames for target adjectives from CHILDES. Underscore indicates position of target word.

Big	Freq	Little	Freq	Good	Freq	Red	Freq
too_	89	a_bit	102	very_	126	a_one	13
a_girl	15	a_more	28	that's_	28	big_ball	11
too_for	14	this_piggy	23	_job	26	the_one	10
a_piece	12	a_	18	a_idea	24	bright_caboose	9
so_	10	a_boy	15	that_	23	is_	8
a_one	10	a_baby	14	it_	19	the_ball	8
the_ball	9	the_girl	14	very_honey	17	little_wagon	6
a_noise	8	hi_girl	13	a_one	16	the_car	5
the_red	7	this_piggie	11	a_job	14	big_barn	4
a_red	7	a_bag	11	very_sweetie	11	green_	3
a_mess	7	a_girl	10	a_day	10	a_ball	3
a_ball	6	a_piece	9	a_boy	8	big_dog	3
is_	6	the_boy	9	oh_	8	its_	3
a_book	5	a_while	9	very_job	7	the_barn	3
a_meal	5	too_	8	so_	7	that_	3
a_deal	4	the_ball	8	was_	7	a_balloon	3
a_rock	4	the_dog	8	is_	6	and_	3
boasted_tiger	4	your_guy	7	a_place	6	a_rose	3
a_kiss	4	a_mouse	7	a_breakfast	6	they're_	3
your_red	4			be_	6	that's_	3

2.3. Discussion

Our corpus analyses revealed, as expected, that property words are used slightly less often than object words in parental speech to young infants. But crucially, these property words are only rarely paired with a given object word. As a result, a plausible word-learning mechanism is for infants to compute (implicitly) the likelihood that a given content word (e.g., *ball*) was spoken in the presence of a multi-component object or event (e.g., *red ball*) and simply linking that content word to the most common referent (either red or ball) across all instances that contain that object or event. The problem with this putative mechanism is that the distributional evidence for the inverse case – the probability of a content word given a property word – is only slightly more likely. Thus, it is not clear, in the absence of an enormous amount of parental speech that contains these small and slightly different probabilities of co-occurrence (0.041 vs. 0.076), that distributional information provides young infants with a robust cue to explain the whole-object bias in word learning.

Of course, any corpus analysis is subject to several caveats. Transcriptions of parental speech to infants are a gloss on the nuances of acoustic/phonetic variability present in natural speech. This renders conclusions from transcriptions a best-case scenario as the infant is surely confronted with more variable input than what is captured by transcribed speech. However, prosodic and extra-linguistic information is largely absent from transcriptions, and even when audio-recordings overcome that limitation, the full context (e.g., which objects are present and at what point in the parental utterance) is missing unless there is a video recording as well. Thus, corpus analyses, at best, should be viewed as tapping into the most robust sources of information available to the infant learner and not a definitive demonstration that other sources of information are absent.

An alternative account of the whole-object bias in early word learning is that infants must conduct a "rapid mapping" task in which they are faced with large amounts of information. This task is, after all, what characterizes the *gavagai* problem – given the flow of events during natural word-learning contexts, infants must make inferences about the meaning of parental speech at a rate of about 2 words/s as they attend to visual object/events. While infants as young as 6 months show rudimentary word-referent mapping for highly familiar stimuli (Bergelson & Swingley, 2012), an unfamiliar word-form paired with an unfamiliar visual object creates a challenge even for 14-month-olds (Stager & Werker, 1997). Thus, the fundamental problem facing infants who are

attempting to map a novel word-form to a novel object/event is to retain the most robust visual information in memory so that it can be repeatedly linked to that same word-form across future repetitions of that object/event.

The role of memory constraints on learning in infants, and in fact even in naïve adults, has been termed the "less is more" hypothesis (Newport, 1990). The basic idea is that when confronted with too much information, a reasonable strategy is to not attempt to retain all of that information in memory. Adults who try to retain all of the details present in a complex event often focus on the exceptions rather than the dominant rules (Hudson Kam & Newport, 2005), whereas infants who have such constraints "built in" because of their limited working memory and limited experience with creating robust sets of visual features for object recognition, fail to access all of the details and only retain "global" information. The whole-object bias is potentially an outcome of this limited ability to rapidly encode more than the most global form of an object when it is presented briefly in a word-naming context. As a result, information that is reliably present in the learning context but is at a "fine grained" level of detail, may simply be lost from the array of potential word-referent mappings, thereby reducing the combinatorics of the *gavagai* problem and highlighting the global information that leads to the whole-object bias.

The cross-situational statistical learning paradigm provides a relevant testbed for examining this "rapid encoding and global memory" hypothesis. Interestingly, studies using the cross-situational statistical learning paradigm have focused almost entirely on mapping auditory word-forms to whole objects (i.e., unique combinations of shape, color, texture, size, and viewpoint). Thus, in Experiment 1 we turn to the case of learning property words and we modify the paradigm slightly to vary the distributional information present for adult learners.

3. Experiment 1: on-line learning of object and property words in adults

The cross-situational statistical learning paradigm was introduced by Yu and Smith (2007) to study how sparse labeling of objects might enable learners to associate nonsense words with scenes containing multiple objects. First with adults, using scenes with four objects, and then with 12-month-old infants (Smith & Yu, 2008), using two objects, they showed that when these multi-object scenes were labeled with the same number of words as objects present in the scenes, learners could induce which label was reliably mapped to each object. Here we ask whether adults can learn property words when presented with simple 1-object scenes. Crucially, each object has two potential referents – its shape, which is most commonly associated with the whole-object bias, and its color, which is a salient object property. However, rather than providing a pair of labels for each object – one label for its shape and one label for its color – we varied the distribution of these object and property labels. For example, on some learning trials only one label was provided, and the adult had to infer whether it was an object/shape word or a property/color word. On other learning trials, two words were presented, but again adults had to infer which one was the object/shape word or the property/color word. Thus, the overall learning situation confronting the adults was more akin to the ambiguity present in real-world word-learning, thereby affording us access to any biases that adults might bring to this cross-situational statistical learning context and the *gavagai* problem that it represents.

There is a surprising paucity of research on this question of learning property words via cross-situational statistical learning (see review by Zhang, Chen, & Yu, 2019). Only two studies have examined directly the trade-off between the learning of object labels and property labels. Chen, Gershkoff-Stowe, Wu, Cheung, and Yu (2017) created a novel set of visual stimuli that had both whole-object names and names for a visual feature of a subset of those objects (e.g., a hook). The feature name was contained in a single syllable that formed part of the whole-object name, much like a morphological marker. The visual stimuli were presented

across learning trials in groups of four and each was labeled with a bisyllabic word that contained information that mapped onto both the objects and their features. Adults exhibited above-chance learning of both object-level and feature-level labels, although feature-level performance was less robust, especially when contained in final-syllable position. Chen, Zhang, and Yu (2018) employed a similar design using real-world objects but now the object name (e.g., *vamy* = beagle) and the category name (e.g., *zorch* = dog) were separate words. Most relevant to the present experiment, in a mixed condition, three objects were shown on each learning trial while two object labels and one category label were spoken. Adults showed above-chance performance on the test trials for both object words and category words, but word-learning performance for a given object was quite poor (~15%) when both types of information were present.

It is important to note that one component of the design of Experiment 1 differs from all previous cross-situational statistical word-learning experiments. Rather than exhaustive labeling of each object in the display (e.g., 2 or 4 objects), only one object was presented and labeled on each trial. We chose this design because pilot studies revealed that labeling only one object in a 2-object scene led to poor learning of the object/property labels across trials, especially in an on-line data collection platform. While this single-object design is admittedly simpler than the traditional multi-object design, it nevertheless captures the essential ambiguity of the *gavagai* problem and represents the canonical case of parental labeling when the infant's attention is directed to a single object in an otherwise cluttered environment. Moreover, the mixture of one- and two-word labels across learning trials more closely approximates the distributional properties of parental input, since it would be infelicitous to use property labels in natural contexts where it is not required (e.g., referring to every instance of a dog as *furry dog*).

3.1. Materials and methods

Six different groups of participants were recruited via SONA and Prolific to participate in an online experiment. There were 6 different conditions in the experiment, each of which had a different distribution of shape and color names presented as participants viewed images containing a single object. There were 28 adults (21 female) in Condition 1, 28 adults (15 female) in Condition 2, 27 adults (9 female) in Condition 3, 28 adults (9 female) in Condition 4, 27 adults (9 female) in Condition 5, and 28 adults (10 female) in Condition 6. Participants were largely White or Asian, at least 18 years or age, and not required to be native speakers of English (although all instructions were only in English).

The inventory of objects consisted of six different shapes, each of which was paired with one of three possible colors. This ratio was chosen to better reflect the greater relative frequency of whole-object labels than property labels in natural discourse. Each shape and each color was associated with a unique artificial word label (see Fig. 1). Auditory stimuli consisted of a female voice that named the object/shape or property/color for the learning phase of the experiment, followed by a test phase in which trials asked, "Which one is (object/property label)?" (see Fig. 2). Note that this 6:3 ratio of shapes to colors created greater memory demands on learning object names than on learning property names, thereby biasing performance toward better learning of color labels. Auditory stimuli were generated using an online text-to-speech synthesizer. The experiment was coded using PsychoPy3 and hosted online via Pavlovia. Participants completed the experiment independently using their own computers.

During the learning phase of the experiment, participants were instructed to simply watch and listen to the objects and labels so that they could "learn some new words." They saw one object per trial and heard either the shape label, color label, or both labels (see Fig. 2), for a total of 432 trials (with the exception of Condition 4, which was half the length of the others). When both labels were presented, the order was always the color label followed by the shape label to compensate for the
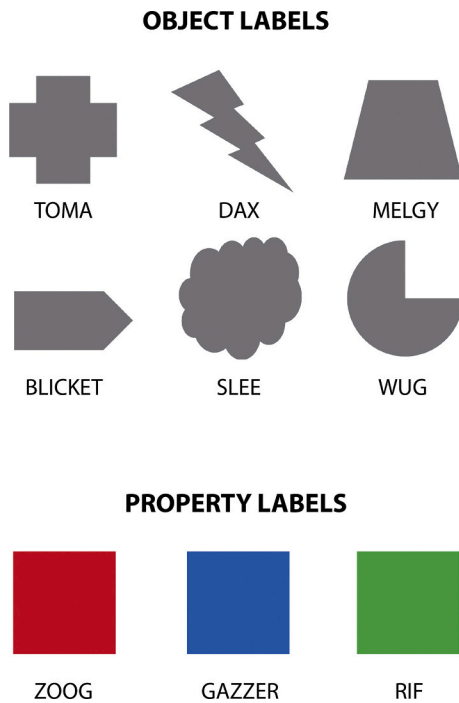
OBJECT LABELS

TOMA   DAX   MELGY

BLICKET   SLEE   WUG

PROPERTY LABELS

ZOOG   GAZZER   RIF

Fig. 1. Stimuli for online Experiment 1.

canonical ordering of adjectives and nouns in English-speaking participants. Because a single instance of each label was extremely brief, the label (either one or two words) was repeated twice. Two pseudorandomized orders of the trials were created in advance and counterbalanced between participants; stimulus presentation was constrained such that the labeled shape and color did not appear in consecutive trials.

The six conditions were identical except for the proportion and number of trials in which the object label, property label, or both labels were presented. In Condition 1, participants were provided only object labels for 25% of the trials, only property labels for 25% of the trials, and both labels for 50% of the trials – this first condition of the experiment served to examine differences in learning object versus property labels when participants had equal exposure to both. In Condition 2, participants were provided with object labels for 50% of the trials, property labels for 25% of the trials, and both labels for 25% of the trials. In Condition 3, participants were provided with object labels for 25% of the trials, property labels for 50% of the trials, and both labels for 25% of the trials. In Condition 4 the same proportions of object, property and both labels as in Condition 3 were used, but with half the number of trials (for a total of 216 trials) to see if increasing task difficulty (by shortening the learning phase) would highlight differences in learning object versus property labels. In Condition 5, object labels were presented on 12.5% of

the trials, property labels were presented on 75% of the trials, and both labels were presented on 12.5% of the trials. And in Condition 6, property labels were presented on 75% of the trials and both property and object labels were presented on 25% of the trials.

Following the learning phase, participants completed a test phase in which they saw two stimuli per trial and were asked to choose which one was described by an object or property label (see Fig. 2). Participants were instructed to press "1" or "2" on their keyboard to select a stimulus. Each object and property label was presented three times, for a total of 27 test trials. Test trials were pseudorandomized such that the same label was not presented on two consecutive trials.

3.2. Results

Fig. 3 shows the mean test trial accuracy for object and property words in each of the six experimental conditions. The most robust object-word learning was present in Conditions 1 and 2 where object labels (either alone or in combination with property labels) were present on 75% of the learning trials. In Conditions 3 and 4, where object labels were present on 50% of the learning trials, test trial accuracy for object words declined slightly, and remained at the same level (~70%) in Conditions 5 and 6 despite object labels being present on only 25% of learning trials. This effect of the prevalence of object labels on learning object words was reflected in a significant Condition effect in a 1-way ANOVA [$F(5,165) = 4.413$, $p < 0.001$].

Fig. 3 also shows that the accuracy of learning property words was fairly consistent with the likelihood that property labels (alone or combination with object labels) were present on the learning trials. The proportion of learning trials with property labels in Conditions 1–6 was 75%, 50%, 75%, 75%, 87.5%, and 100%, respectively, and a 1-way ANOVA revealed that the Condition effect was not significant [$F(5,165) = 2.014$, $p = 0.079$]. Interestingly, it was only in Condition 6 where property labels alone were presented on 75% of the learning trials and object labels were always presented in combination with property labels on the remaining 25% of the learning trials, that accuracy for property words exceeded accuracy for object words [$t(27) = -2.21$, $p < 0.05$]. Further evidence of a relationship between distributional information and learning for object labels but not property labels is shown in Table 4.

Finally, we asked whether individuals who learned object words showed a trade-off in their learning of property words (and vice versa). That is, was there competition between the two types of labels present in the task as observed by Benitez, Yurovsky, and Smith (2016) for 1-word versus 2-word labels. As shown in Table 5, the answer was 'no', with participants showing consistently positive correlations between their learning accuracy for object-labels and property-labels.

3.3. Discussion

The findings from Experiment 1 show that adults can readily learn both object/shape words and property/color words in a simplified cross-
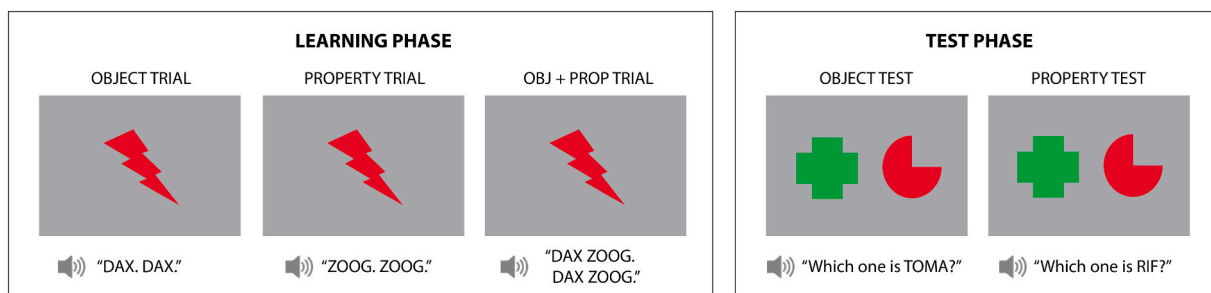

Fig. 2. Example of each type of trial in the exposure phase of Experiment 1. Auditory stimuli are indicated by quotations. Target object is the one on the left for both of the depicted test trials. (TOMA = shape; RIF = color).
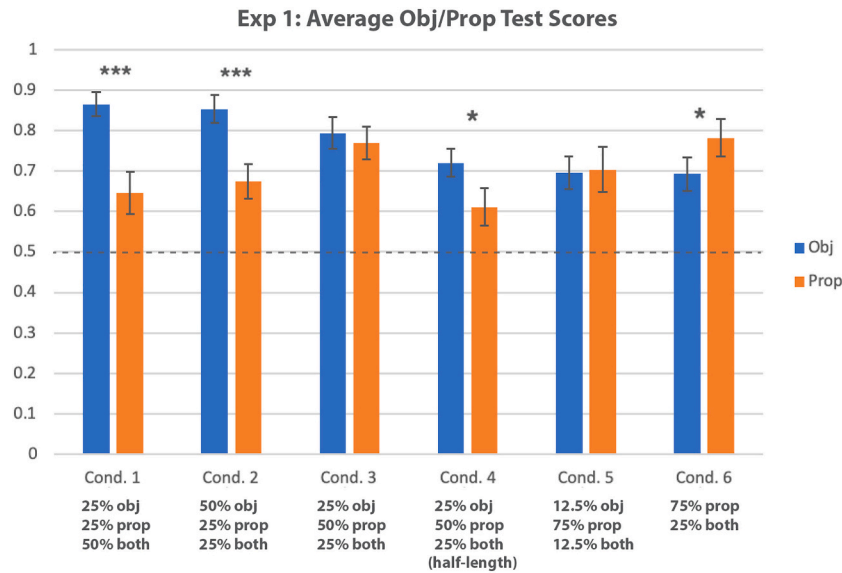
Fig. 3. Average scores across participants on object and property tests for all conditions. Dashed line represents chance, error bars represent S.E., * indicates p < 0.05, *** indicates $p < 0.001$.

Table 4
Pearson correlations between label proportions and test scores across all conditions in Experiment 1.

	Object test	Property test
object labels	r(164) = 0.265, p < 0.001	r(164) = −0.120, p = 0.124
property labels	r(164) = −0.329, p < 0.001	r(164) = 0.133, p = 0.088
both labels	r(164) = 0.239, p = 0.002	r(164) = −0.079, p = 0.310

Table 5
Pearson correlations between accuracy of learning object names and accuracy of learning property names for each condition in Experiment 1. Raw object and property percentages were converted to log-odds to linearize scores.

	Object vs. property
Cond. 1	r(26) = 0.536, p = 0.003
Cond. 2	r(26) = 0.594, p < 0.001
Cond. 3	r(25) = 0.535, p = 0.535
Cond. 4	r(26) = 0.478, p = 0.478
Cond. 5	r(25) = 0.613, p < 0.001
Cond. 6	r(26) = 0.649, p < 0.001

situational statistical learning paradigm, in which only one object was present but a mixture of 1- and 2-word labels were provided on each learning trial for the object shape and/or object color. Moreover, the results reveal that the accuracy in learning labels for objects and properties generally tracks the distributional statistics with which these labels are presented in the input, with some evidence that property labels are even less preferred in the early phase of learning (e.g., Conditions 3 and 4 had the same distributions but the learning phase in Condition 4 was half that of Condition 3). However, even in Condition 6 where object and property words were presented on 25% of learning trials and property words on 75% of learning trials, adults still learned objects words with above-chance accuracy: mean = 0.69 [$t(27)$ = 4.57, $p < 0.0001$]. Moreover, there was no evidence that learning one type of word interfered with learning the other as learning accuracy for object and property words was positively correlated.

4. Experiment 2: generalization of object and property labels

While the results of Experiment 1 show a clear bias in adults for learning object/shape labels over property/color labels, a stronger test of the productive use of such labels involves their generalization to novel exemplars. For example, a child who labels their favorite stuffed animal as "teddy" could be restricting that name to a single unique object, whereas a child who extends that label to a large set of bears of different sizes and textures has clearly generalized that label to a category of objects. Experiment 2 provides such a test of generalization using the same cross-situational statistical learning paradigm as in Experiment 1. In addition to a test of generalization to novel exemplars of the object/ shape and property/color labels, Experiment 2 divided the exposure phase into thirds, with a test of learning for the shape and color labels of the trained stimuli after each of these three exposure sub-phases. The goal here was to determine whether the object bias emerged early in learning and whether learning of property labels grew stronger with additional exposure to the distributional information contained in the exposure phase.

4.1. Materials and methods

A total of 28 participants were recruited via Prolific to participate in Experiment 2. The stimulus inventory was the same as in Experiment 1 for the exposure phase, but an additional 24 novel stimuli were generated for use in the generalization test at the end of the experiment. Of these novel stimuli, half had a novel color but a familiar shape and half had a novel shape but a familiar color (overall two novel colors and four novel shapes, see Fig. 4). Audio stimuli were the same as in Experiment 1. That is, a female voice presented the object or property label twice for each trial during the learning phase and asked, "Which one is (object/ property label)?" for the two-object test trials (see Fig. 2).

The design included 216 learning trials during the exposure phase (the same as the half-length exposure in Condition 4 of Experiment 1). The exposure phase was divided into three blocks of 72 trials each. During these learning trials, participants saw one object per trial and heard an object label on 50% of the trials and a property label on the other 50% of the trials. The 6 shape labels and 3 color labels were balanced within the 50–50 distribution of object and property words presented in each learning block, although this resulted in twice as many instances of each color label than each shape label. Notably, participants were only exposed to single word labels; there were no trials that included shape+color labels as in Experiment 1. To assess the time course of learning, we included a brief test after each of the three
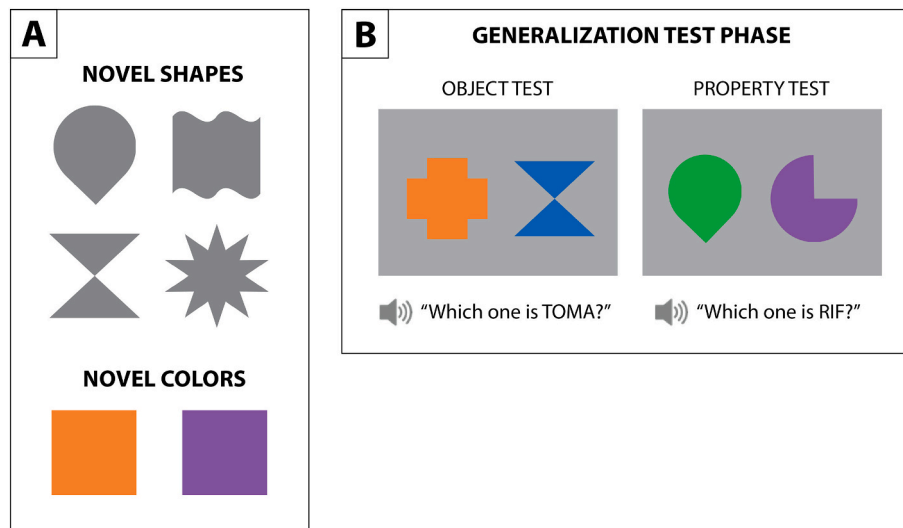
Fig. 4. Novel object shapes and colors used in the generalization test (A). Examples of two generalization test trials (B), in which the correct answer is on the left.

learning blocks, during which participants saw two familiar stimuli per trial and were asked to select which one was described by an object or property label. Participants were tested on each of the 6 object/shape words and each of the 3 property/color words for a total of 9 trials per test. Counterbalancing of learning and test trials was otherwise the same as in Experiment 1.

At the end of the third testing phase, participants completed a generalization test in which they saw two novel stimuli per trial and were asked to choose which one was described by an object or property label (see Fig. 4). Each of the 6 shape labels and 3 color labels was tested twice, for a total of 18 generalization trials. Test pairs included one stimulus with a novel color/familiar shape and one stimulus with a novel shape/familiar color, thereby preventing participants from inferring whether a test label referred to an object or a property during this generalization test. Test trials were pseudorandomized such that the same label was not presented on two consecutive trials.

4.2. Results

The results of the tests after each third of the exposure phase are shown in Fig. 5. Performance on object/shape test-trials was above 80% correct even after the first block of 72 learning trials in the exposure phase and rose to over 90% after the third block of learning trials. In

Fig. 5. Accuracy on object and property tests after each third of the learning phase of Experiment 2, as well as the final generalization test. Dashed line represents chance, error bars represent S.E., * indicates $p < 0.05$, *** indicates $p < 0.001$.

contrast, performance on property/color test-trials never reached 70% correct. This difference in performance between the learning of object and property labels was significant [$F(1,2) = 44.55$, $p < 0.001$].

Fig. 5 also shows the results of the generalization test at the end of the entire exposure phase (216 learning trials). Performance on object/shape generalization trials (93.7% correct) was not only significantly above chance [$t(27) = 31.69$, $p < 0.001$] but was indistinguishable from performance on the trained stimuli (91.1% correct). In contrast, performance on property/color generalization trials (31.5% correct) was significantly below chance [$t(27) = -4.13$, $p < 0.001$] and also below performance on property/color test trials for familiar stimuli (63.1% correct). Presumably, this further decline in performance on generalization trials was the result of confusion when a weak representation of the color label was confronted with novel shapes, perhaps leading to the inference that the same label referred to two different colors (as in bilingual infants: see Padmapriya, Hall & Werker, 2017). However, there were relatively few property-generalization test trials, so this interpretation should remain tentative.

4.3. Discussion

The results of Experiment 2 provided even clearer evidence than Experiment 1 that adults have a strong bias to map novel labels to object/shape than to property/color in a cross-situational statistical learning paradigm. When the distributional statistics of these two mappings were balanced (50–50 object vs. property), adults learned the labels for the 6 shapes at near asymptotic levels (> 90% correct) and generalized those shape labels to objects with novel colors with equal performance (> 90% correct). However, adults learned the labels for the 3 colors with less accuracy (< 70% correct) and failed to generalize those color labels to novel shapes.

5. General discussion

The *gavagai* problem has been recognized by linguistics, philosophers, and psychologists for 60 years as a potentially serious – some would claim intractable – challenge for language learning. Given the potentially infinite number of possible mappings between auditory word-forms and their intended meanings, how does a naïve learner "break this code" and settle on the demonstrably correct mapping that enables reliable communication? Theorists have posited a variety of constraints (Yurovsky & Frank, 2015), from the presence of innate categories at the structural level to powerful general-purpose learning
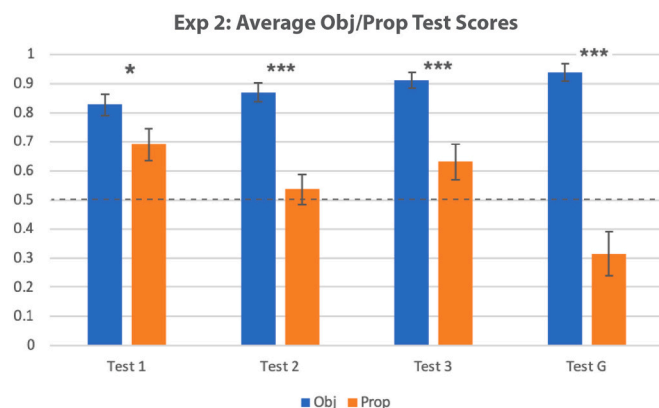
mechanisms fueled by extremely large datasets. There is little debate that the rapid acquisition of language by children in the first few years of life requires some set of constraints, whether structural or learning or both. In the present study, results from corpus analyses and two experiments have added some further insights on what those constraints are and how they might operate.

First, it is clear from our corpus analyses that the distribution of words in the early linguistic input to infants is skewed to include the names for whole objects, with property words that describe some feature of those objects being slightly less frequent. Moreover, the conditional probability of a property name given an object name is extremely low, suggesting that knowing an object name would not enable learning a property name. However, the inverse – the conditional probability of an object name given a property name – was also extremely low, although about double that of the conditional probability of a property word given an object word. This implies that even if infants had no bias, whether perceptual or attentional, to seek matches between whole objects and auditory word-forms, the input to which they are exposed would disfavor mappings that involve properties, but only if infants were exposed to a sufficiently large corpus to enable these low conditional probabilities to be relevant for learning.

Second, the results from Experiments 1 and 2 provide compelling evidence that adults have little difficulty mapping auditory word-forms to both objects and their properties. Despite this robust evidence of learning, they exhibit a bias in favor of object labels over property labels. Even when a property label was used 100% of the time during the object+label learning phase, and object labels were used on only 25% of the learning trials, adults were able to reliably map words to objects. Moreover, under these highly property-biased circumstances, the accuracy for property labels was only slightly (though significantly) higher than for object labels. This is similar to evidence from Monaghan, Mattock, Davies & Smith (2015) that adults learn noun mappings more readily than verb mappings, even when those mappings are not intermixed. Finally, performance on both object- and property-word mappings was positively correlated, suggesting that good/poor learners were equally adept at extracting both types of information. This is particularly interesting in light of evidence that "mixture" designs, in which more than a single source of information is available to map words to their referents, can lead to competition between alternative mappings (Benitez et al., 2016; Yurovsky, Yu, & Smith, 2013).

The results from Experiments 1 and 2 are consistent with the following overall account of how the *gavagai* problem is overcome in natural word-learning contexts. First, infants attend to a limited set of objects/events that are in close proximity to themselves and often in their grasp as parental speech is used to label that object/event. Parental speech is dominated by whole-object (basic level category) words, with property words used only slightly less frequently, but with pairs of object and property words having very low conditional probabilities. Thus, the distributional properties of the word-learning context are slightly object-biased, but that bias is unlikely to be sufficient to create a whole-object bias in word learning unless the infant is exposed to a very large corpus of parental speech input.

Second, a small set of high-frequency object words are present in consistent contexts (Roy, Frank, DeCamp, Miller, & Roy, 2015), thereby enabling those words to become firmly established in long-term memory and allowing them to serve as "anchors" for learning the low-frequency bigrams and frequent-frames that surround these object words. In addition, shared contexts create competition for lexical access, even when the images depicting a word's referent are easily discriminable (Bergelson & Aslin, 2017). Both computational models (Yu & Smith, 2012) and empirical findings from adults (Clerkin, Hart, Regh, Yu, and Smith (2017) reveal that low-frequency words can, indeed, be learned as long as there is sufficient overall input to enable the extraction of sparse

but consistent distributional information. Moreover, the Zipfian distribution of object names does not prove to be an impediment to this word-learning process.

Third, although distributional information is available and could, in principle, contribute to the whole-object bias, in the early phase of word learning infants are being exposed to sparse language input. Thus, their initial word-learning task requires them to attend to, perceptually encode, and retain in memory a large set of potential matches between the auditory word-form and its intended referent. This mapping task, because of its high demands on working memory, creates a "representational fidelity" challenge. If infants attempt to encode all available information, the *gavagai* problem becomes exponential. But if their limited cognitive abilities constrain the number and diversity of object/event characteristics that can be encoded into memory, then the combinatorics of the *gavagai* problem are reduced. Although infants have impressive encoding and short-term memory abilities (Blaser and Kaldy, 2010), they are unable to retain that information without extensive repetition. Moreover, infants have very poor working memory (Kaldy & Leslie, 2005; Ross-Sheehy, Oakes, & Luck, 2003). In fact, even retaining two object properties (color and texture) in working memory after each property has been learned via repeated exposures is a challenge for 9-month-olds (Piantadosi, Palmeri, & Aslin, 2018). These cognitive limitations are a hallmark of the "less is more" hypothesis (Newport, 1990) and have been shown – somewhat counterintuitively – to facilitate learning (Hudson Kam & Newport, 2009). The relevance of this hypothesis for the whole-object bias in word learning is that memory limitations are more likely to direct infants' attention to the "global" characteristics of visual objects/events in a rapid word-object mapping task. This demand on rapid encoding and working memory is present even when only a single object is in the infant's focal attention while being labeled, as in Experiments 1 and 2, especially when neither the object's visual features nor the label's auditory features are highly familiar.

Finally, it is important to note that natural objects, at least in the shape and color dimensions, have distributional characteristics that favor the whole-object bias. The defining features of most basic-level objects rely on the obligatory conjunction of its parts – a cup without a handle is a bowl. Thus, the global features that are encoded in memory for a basic-level object have a canonical representational format. In contrast, the defining features of object properties such as color or texture are often non-uniformly distributed – a blue and white striped T-shirt is often labeled as *blue*. This difference in the degree of consistency in how an object's shape or color is present when it is labeled as a *cup* or as *blue* provides another level of distributional information that contributes to the whole-object bias in word learning.

There are, of course, several limitations to the present corpus analyses and experiments with adults. Given relatively small samples of parental speech per infant, we collapsed across infants to increase the robustness of our estimates of parental speech distributions. Thus, some parent-infant dyads with more highly differentiated distributional cues for object and property labels could benefit from this information. We did not conduct a detailed analysis of the contexts in the CHILDES and SEEDLingS samples to fully understand how infants might have interpreted non-linguistic information that could potentially be available for word learning. Thus, we cannot know precisely how much these contextual cues facilitated or impeded the infant's interpretation of parental speech. We also did not have access to the time-course of acquiring an accurate mapping of words to objects/properties in the cross-situational statistical learning paradigm, although the multiple testing phases employed in Experiment 2 revealed little evidence that initial biases were readily overcome by distributional properties of the learning input. Finally, we did not have detailed information about the English-language proficiency of the participants, which could have

interacted with the ordering of object and property labels in Experiment 1 (i.e., we used a property-last word order that differs from English but is consistent with other languages).[1] These and other more detailed questions about the biases that enable infants to resolve the *gavagai* problem await further research.

6. Conclusions

Distributional information about the likelihood of object and property words, as well as their co-occurrence, is present in parental speech to infants. While this information is biased toward whole objects over object properties, those distributional cues are unlikely to play a dominant role in the whole-object bias in word learning. Rather, a more likely mechanism leading to the whole-object bias are the intrinsic demands placed on the infant during rapid naming events. These demands consist of constraints on working memory and representational fidelity for multi-dimensional objects/events to which an auditory word-form is mapped. These cognitive constraints – consistent with the "less is more" hypothesis (Newport, 1990) – result in a bias to map auditory word-forms to the "global" features of objects (i.e., those that are easily encoded), which are correlated with the whole object. Results from adults in a series of cross-situational statistical learning experiments support this memory-based account of the whole-object bias in word learning.

Appendix A. Appendices

Table A1

Top 25 most frequent words co-occurring with instances of "book" in SEEDLingS audio files. Numbers in parenthesis indicate ordinal position, e.g. before (1) indicates word that occurs two words before "book" and before (2) indicates the word that comes directly before "book."

Before (1)	Freq	Before (2)	Freq	After (1)	Freq	After (2)	Freq
read	63	the	112	and	17	you	16
a	26	a	67	is	11	we	7
the	18	your	50	that	8	the	5
all	17	this	35	about	8	read	5
at	15	that	27	to	8	bears	4
like	13	some	13	are	7	a	4
eat	11	good	11	you	7	and	4
with	11	those	11	up	7	us	4
get	10	my	10	for	6	then	3
your	9	baby	8	upstairs	6	that	3
of	8	these	8	I	6	this	3
have	8	another	7	but	5	to	3
for	7	truck	7	wit	5	called	3
put	7	more	7	in	5	I	3
got	7	library	6	like	5	there	3
hold	7	of	6	out	4	your	3
reading	7	many	5	or	4	got	3
it's	7	our	5	too	4	silly	3
here's	7	new	4	down	3	over	3
my	7	her	4	with	3	Mommy	3
want	6	little	4	the	3	me	3
this	6	what	4	back	3	with	2
see	6	any	3	here	3	maybe	2
you	5	picture	3	again	3	down	2
one	5	three	3	of	3	my	2

Table A2

Most frequent words co-occurring with instances of "ball" in SEEDLingS audio files.

Before (1)	Freq	Before (2)	Freq	After (1)	Freq	After (2)	Freq
roll	43	the	133	and	17	you	16
get	21	your	19	is	11	we	7
gimme	13	a	18	that	8	the	5
the	8	beach	14	about	8	read	5
that	7	my	14	to	8	bears	4

(continued on next page)

---

[1] The IP-addresses of the 194 participants across all experimental conditions were located in 23 countries, 6 of which are predominantly English-speaking. We assumed that all 85 participants from these 6 countries were native speakers of English and asked whether performance on the property-label test was correlated with the proportion of subjects in each condition who were putatively native speakers of English. That correlation was not significant [$r = -0.21, t(5) = -0.47, p = 0.658$], suggesting that language-specific ordering of grammatical categories did not contribute to differences across experimental conditions.

Table A2 (*continued*)

Before (1)	Freq	Before (2)	Freq	After (1)	Freq	After (2)	Freq
with	7	that	8	are	7	a	4
of	7	ball	5	you	7	and	4
a	7	little	5	up	7	us	4
like	6	this	3	for	6	then	3
got	6	green	2	upstairs	6	that	3
see	5	her	2	I	6	this	3
and	5	giant	2	but	5	to	3
there's	5	yellow	2	with	5	called	3
have	5	giraffe	2	in	5	I	3
me	5	playing	2	like	5	there	3
you	4	his	1	out	4	your	3
your	4	cannon	1	or	4	got	3
at	4	cotton	1	too	4	silly	3
kick	4	blue	1	down	3	over	3
put	4	they're	1	with	3	Mommy	3
throw	3	another	1	the	3	me	3
ball	3	understand	1	back	3	with	2
this	2	bouncy	1	here	3	maybe	2
that's	2	tennis	1	again	3	down	2
it's	2	rolling	1	of	3	my	2

Table A3
Statistics for object and property tests from Experiment 1.

Cond.	Distribution	Object	Property	Paired sample *t*-test
1	Both-50, Obj-25, Prop-25	M = 0.87, SD = 0.16	M = 0.65, SD = 0.28	t(27) = 4.49, p < 0.001
2	Both-25, Obj-50, Prop-25	M = 0.85, SD = 0.18	M = 0.67, SD = 0.23	t(27) = 4.58, p < 0.001
3	Both-25, Obj-25, Prop-50	M = 0.79, SD = 0.20	M = 0.77, SD = 0.21	t(26) = 0.56, p = 0.58
4	Both-25, Obj-25, Prop-50 (half exposure)	M = 0.72, SD = 0.18	M = 0.61, SD = 0.25	t(27) = 2.36, p = 0.03
5	Both-12.5, Obj-12.5, Prop-75	M = 0.70, SD = 0.21	M = 0.70, SD = 0.29	t(26) = −0.17, p = 0.86
6	Both-25, Obj-0, Prop-75	M = 0.69, SD = 0.22	M = 0.78, SD = 0.25	t(27) = −2.21, p = 0.03

Table A4
Statistics for object and property tests from Experiment 2.

Test.	Object	Property	Paired sample t-test
1	M = 0.83, SD = 0.20	M = 0.69, SD = 0.29	t(27) = 2.52, p = 0.018
2	M = 0.87, SD = 0.17	M = 0.54, SD = 0.28	t(27) = 5.80, p < 0.001
3	M = 0.91, SD = 0.14	M = 0.63, SD = 0.31	t(27) = 5.44, p < 0.001
Gen	M = 0.94, SD = 0.16	M = 0.32, SD = 0.40	t(27) = 7.04, p < 0.001

References

Batchelder, E. O. (2002). Bootstrapping the lexicon: A computational model of infant speech segmentation. *Cognition, 83*, 167–206.

Benavides-Varela, S., & Mehler, J. (2015). Verbal positional memory in 7-month-olds. *Child Development, 86*, 209–223.

Benitez, V. L., Yurovsky, D., & Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *Journal of Memory and Language, 90*, 31–48.

Bergelson, E. (2017). SEEDLingS 6 month. In *Databrary*. https://doi.org/10.17910/B7.330.

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences, 114*, 12916–12921.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*, 3253–3258.

Bion, R. 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*, 39–53.

Blaser, E., & Kaldy, Z. (2020). Infants get five stars on iconic memory tests: A partial-report test of 6-month-old infants' iconic memory capacity. *Psychological Science, 21*, 1643–1645.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. I. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*, 977–985.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69–B77.

Chemla, E., Mintz, T. H., Bernal, S., & Christophe, A. (2009). Categorizing words using "frequent frames": What cross-linguistic analyses reveal about distributional acquisition strategies. *Developmental Science, 12*, 396–406.

Chen, C., Gershkoff-Stowe, L., Wu, C., Cheung, H., & Yu, C. (2017). Tracking multiple statistics: Simultaneous learning of object names and categories in English and mandarin speakers. *Cognitive Science, 41*, 1485–1509.

Chen, C., Zhang, Y., & Yu, C. (2018). Learning object names at different hierarchical levels using cross-situational statistics. *Cognitive Science, 42*, 591–605.

Clerkin, E. M., Hart, E., Regh, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B, 372*, 20160055.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language and Speech, 49*, 137–174.

Gerken, L. A., Landau, B., & Remez, R. E. (1990). Function morphemes in young children's speech perception and production. *Developmental Psychology, 26*, 204–216.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition, 73*, 135–176.

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

Graham, S. A., & Poulin-Dubois, D. (1999). Infants' reliance on shape to generalize novel labels to animate and inanimate objects. *Journal of Child Language, 26*, 295–320.

Halberda, J. (2003). The development of a word-learning strategy. *Cognition, 87*, B23–B34.

Hochmann, J.-R. (2013). Word frequency, function words, and the second gavagai problem. *Cognition, 128*, 13–25.

Hochmann, J. R., Endress, A. D., & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition, 115*, 444–457.

Hudson Kam, C., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology, 59*, 30–66.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151–195.

Kaldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition, 97*, 153–177.

Kandhadai, P., Hall, D. G., & Werker, J. F. (2017). Second label learning in bilingual and monolingual infants. *Developmental Science, 20*, Article e12429.

MacNamara, J. (1972). Cognitive basis of language learning in infants. *Psychological Review, 79*, 1–13.

Marino, C., Bernard, C., & Gervain, J. (2020). Word frequency is a cue to lexical category for 8-month-old infants. *Current Biology, 30*, 1380–1386.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science, 14*, 57–77.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101–B111.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317*, 631.

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*, 9014–9019.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*, 678–686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91–117.

Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24-and 36-month-olds. *Developmental Psychology, 41*, 17–29.

Mintz, T. H., Wang, F. H., & Li, J. (2014). Word categorization from distributional information: Frames confer more than the sum of their (bigram) parts. *Cognitive Psychology, 75*, 1–27.

Monaghan, P., Mattock, K., Davies, R. A. I., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science, 39*, 1099–1112.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11–28.

Padmapriya, K., Hall, D. G., & Werker, J. F. (2017). Second label learning in bilingual and monolingual infants. *Developmental Science, 20*, e12429.

Piantadosi, S. T., Palmeri, H., & Aslin, R. N. (2018). Limits on composition of conceptual operations in 9-month-olds. *Infancy, 23*, 310–324.

Quine, W. V. O. (1960). *Word and object (studies in communication)*. New York and London: Technology Press of MIT.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development, 74*, 1807–1822.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences, 112*, 12663–12668.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy, 10*, 187–198.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11–B21.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences, 22*, 325–336.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent mapping. *Trends in Cognitive Sciences, 18*, 251–258.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*, 1558–1568.

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

Stager, C., & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388*, 381–382.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*, 86–132.

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent-child interactions. *Cognition, 148*, 117–135.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*, 1611–1617.

Waxman, S. R., & Booth, A. E. (2001). Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. *Cognitive Psychology, 43*, 217–242.

Wilcox, T. (1999). Object individuation: infants' use of shape, size, pattern, and color. *Cognition, 72*, 126–166.

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. (2012). Modeling cross-situational word–referent learning: Prior questions. *Psychological Review, 119*, 21–39.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145*, 53–62.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science, 16*, 959–966.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science, 37*, 891–921.

Zhang, Y., Chen, C., & Yu, C. (2019). Mechanisms of cross-situational learning: Behavioral and computational evidence. *Advances in Child Development and Behavior, 56*, 37–63.