Moos as cues: One-to-one biases in a non-linguistic and non-communicative domain

Kyle MacDonald[1], Ricardo A. H. Bion[1], & Anne Fernald[1]

[1] Stanford University

Author Note

Correspondence concerning this article should be addressed to Kyle MacDonald, 450 Serra Mall, Stanford, CA 94306. E-mail: kylem4@stanford.edu

Abstract

When hearing a novel name, children tend to select a novel object rather than a familiar one, a bias known as disambiguation. This bias is often assumed to reflect children's expectations about the nature of words or expectations about the communicative intention of speakers. This work investigated whether similar biases emerge in a domain that is non-linguistic and non-communicative for children, but in which strong regularities can be found: the vocalizations that animals produce. Using online processing measures, we first show that two-year-olds can identify familiar animals based on their vocalizations, though not as fast as they are to identify these same animals when hearing their names. We then show that children look at an unfamiliar animal when hearing a novel animal vocalization or novel animal name, being equally fast in both conditions. In a follow-up experiment, we replicate the key finding that children look at a novel animal when hearing a novel animal vocalization, but show that these biases do not necessarily lead to learning. We characterize disambiguation biases as resulting from domain-general processing mechanisms, rather than from lexical or communicative constraints.

Keywords: disambiguation, mutual exclusivity, environmental sounds, retention, word learning

Word count: 7923

Moos as cues: One-to-one biases in a non-linguistic and non-communicative domain

Introduction

When children encounter people and animals in daily life, they experience them through multiple sensory modalities simultaneously. When playing with pets, for example, the child learns to associate the physical features and actions of dogs and cows with the barking or mooing sounds typically produced by these animals. Eventually, the child will also link words to each type of animal, learning that they are associated with the object names dog or cow, as well as with onomatopoeic words resembling their characteristic vocalizations, such as woof-woof or moo. Thus, just as familiar object labels are consistently associated with particular kinds of animate and inanimate objects, animal vocalizations and onomatopoeic words for vocalizations also provide consistent associations between auditory stimuli and different types of animals. Here, we compare young children's efficiency in using these different sounds as cues to identifying particular animals. Also, we investigate whether children can use disambiguation strategies, which have frequently been characterized as pragmatic or lexically-specific (Bloom, 2002; Diesendruck & Markson, 2001; Markman, 1991), to infer which of two animals is associated with a novel animal vocalization.

The question of whether words are a special kind of stimulus is not new. Several studies have found advantages for speech sounds over tones in object individuation and categorization in young infants (Fulkerson & Waxman, 2007; Xu, 2002). Focusing on associations between objects and sounds, objects and tones, or objects and gestures, several studies found that younger infants accept several different forms as potential object labels but that older infants are more discriminating and favor words (Namy & Waxman, 1998; Woodward & Hoyne, 1999). Another line of research found that infants prefer to hear spoken words over some non-linguistic analogs (Vouloumanos & Werker, 2004, 2007a, 2007b) and that the neonate brain responds differently to speech as compared to backward speech (Pena et al., 2003). These studies found advantages for speech over non-linguistic analogs in categorization, individuation, crossmodal association, and speech preferences. However, they

all focused on arbitrary, non-linguistic cues that are not consistently associated with objects in children's everyday environments.

Other research has approached the question of whether speech is special from a different perspective, comparing how people process spoken words as compared to non-arbitrary environmental sounds, such as animal vocalizations (e.g., cat meowing) or the sounds produced by inanimate objects (e.g., car starting). Studies with adults have found similarities and differences in both behavioral and neural responses to cross-modal semantic associations between words and environmental sounds. For example, in a picture detection task, Chen and Spence (2011) found a facilitation effect for environmental sounds but not for words when the onset of the auditory stimulus preceded the image by approximately 350 ms. In a follow-up study, Chen and Spence (2013) presented environmental sounds and words across a wider range of time intervals before the image onset. They found that both naturalistic sounds and spoken words resulted in cross-modal priming, but that the effect of spoken words required more time between the auditory and visual stimuli as compared to the naturalistic sounds. They argue that these data are consistent with a differential processing account: that the recognition of environmental sounds is faster because words must also be processed at a lexical stage before accessing semantic representations, whereas environmental sounds activate semantic representations directly.

In contrast, other studies have found an advantage for the processing of lexical items compared to environmental sounds. For example, in a sound-to-picture matching task, Lupyan and Thompson-Schill (2012) showed that words ("cat") lead to faster and more accurate object recognition as compared to either nonverbal cues (the sound of a cat meowing) or lexicalized versions of the nonverbal cue (the word "meowing"). Recent work by Edmiston and Lupyan (2015) followed up on this result by manipulating the congruency between the environmental sounds and their corresponding images within the same basic-level category (e.g., pairing the sound of an acoustic guitar with an image of either an acoustic or an electric guitar). Adults were faster to identify a congruent sound-image pair,

suggesting that the environmental sound carried additional information about the specific type of object generating that sound. Critically, adults were fastest identify the images after hearing their labels. Edmiston and Lupyan (2015) argue that this lexical processing advantage is driven by the specificity of the conceptual representations that labels evoke. That is, words function as "unmotivated cues" that are decoupled from surface-level features of a particular instantiation of a category and are therefore better able to evoke abstract category representations that are more useful in distinguishing between categories.

In an ERP study with adults, Cummings et al. (2006) found that largely overlapping neural networks processed verbal and non-verbal meaningful sounds. In another study focusing on three different sound types that varied in arbitrariness, Hashimoto et al. (2006) found different neural mechanisms for the processing of animal names and vocalizations, with onomatopoeic words activating both areas. Because research on environmental sounds is relatively new, it is hard to reconcile these somewhat discrepant findings. Variations in tasks or timing of stimuli could influence results, and different theoretical commitments can lead to different interpretations. For example, environmental sounds are often treated as encompassing both the sounds of living and human-made objects (e.g., cow mooing, bell ringing) despite evidence that these sounds are treated differently by the adult brain (Murray, Camen, Andino, Bovet, & Clarke, 2006). Nevertheless, this line of research provides promising new ways to examine the question of whether language emerges from the interaction of domain-general cognitive processes or domain-specific mechanism (Bates, MacWhinney, & others, 1989). For example, recent research comparing the processing of speech and non-speech sounds is leading to new insights relevant to autism, developmental language impairment, and cochlear implants (Cummings & Ceponiene, 2010; McCleery et al., 2010).

From a developmental perspective, it is also important to understand how children process words and non-arbitrary non-linguistic sounds. However, few studies have examined this question. Using a preferential-looking paradigm, Cummings, Saygin, Bates, and Dick

(2009) found that 15- and 25-mo-olds can use words and environmental sounds to guide their attention to familiar objects, improving as they get older. Vouloumanos, Druhen, Hauser, and Huizink (2009) found that 5-month-olds can match some animals to the vocalizations they produce. And studies with children with autism and developmental language impairment found more severe deficits for the processing of words than environmental sounds (Cummings & Ceponiene, 2010; McCleery et al., 2010).

In the work reported there, we build on these earlier studies using the looking-while-listening paradigm, which has been widely used to assess real-time interpretation of spoken words by infants and young children (Fernald, Zangl, Portillo, & Marchman, 2008). One major goal of this research is to investigate how children process different types of auditory stimuli that are consistently associated with familiar animals but vary in level of arbitrariness. First, we ask whether 32-month-olds can use onomatopoeic sounds (e.g., bow-wow), and animal vocalizations (e.g., dog barking) to identify familiar animals, as well as familiar animal names (e.g., dog). By using real-time processing measures, we can determine whether these three sounds are equally effective as acoustic cues in guiding children's attention to a particular animal in the visual scene. The use of looking to visual stimuli, rather than object-choice responses, reduces the task demands of procedures requiring more complex responses such as reaching or pointing and yield continuous rather than categorical measures of attention on every trial, capturing differences in processing that might not be detected by offline tasks.

A second major goal of this research is to investigate how young children learn to link non-linguistic sounds to animate objects. Do two-year-olds make similar inferences when mapping a novel word and a novel vocalization to an unfamiliar animal? Typically, word learning is portrayed as an intractable challenge, while associating animals with the sounds they produce might appear trivial. The acoustic structure of vocalizations is influenced by the size and shape of the vocal tract and other physical features, linking sounds to their source in a non-arbitrary way. And the fact that many animal vocalizations are accompanied

by synchronous physical movements might provide children with additional non-arbitrary cues to the source of the sound. Even in the absence of additional visual cues, it is often possible to pinpoint the source of a sound with reasonable accuracy. In contrast, because the acoustic structure of a word is in most cases arbitrary concerning potential referents, and it is produced by a speaker and not by the object itself, learning to associate speech sounds with objects is often characterized as a complex problem of induction (Markman, 1991).

To solve the word-learning puzzle, children are said to use constraints on the possible meanings of words. The most widely studied of these constraints is that each object must have only one name (Markman, 1991). Evidence for this default assumption comes from disambiguation tasks in which children hear a novel label in the presence of a novel object and one or more familiar objects. In these situations, children tend to select the novel object as the referent for the novel word, presumably because the familiar objects already have names associated with them. The debate about the origins, scope, and generality of the Mutual Exclusivity (ME) constraint has focused on whether this response bias provides evidence for a lexical constraint, or whether it results instead from inferences about speakers' communicative intent. Lexical accounts characterize ME as a "domain specific mechanism specific to word leaning" (Marchena, Eigsti, Worek, Ono, & Snedeker, 2011), which "predicts disambiguation only within the domain of word learning (i.e., it is domain-specific)" (Scofield & Behrend, 2007). Pragmatic accounts propose that ME extends to communicative acts more broadly, reflecting assumptions that speakers are cooperative and should use conventional names to refer to familiar objects (Bloom, 2002; Clark, 19 A third possibility is that the bias toward one-to-one mappings reflects general tendencies to find simple regularities in complex domains, a perspective embraced by recent computational approaches to word learning (Frank, Goodman, & Tenenbaum, 2009; McMurray, Horst, & Samuelson, 2012; Regier, 2003).

Thus, lexical and pragmatic accounts of the scope of the ME constraint pred hat one-to-one biases are either unique to word learning or that they generalize to

communicative acts more broadly, while domain-general accounts predict that they would apply to any domain in which consistent one-to-one mappings are observed. To explore the possibility that one-to-one biases in sound-object mappings are not limited to interpreting communicative acts, we investigated whether children would show responses comparable to the mutual exclusivity bias in a domain that is neither linguistic nor communicative, but in which consistent associations are observed between objects and auditory cues.

Experiment 1

Experiment 1 asks two questions. First, can 32-month-olds use familiar animal names (e.g., dog), onomatopoeic words (e.g., bow-wow), and animal vocalizations (e.g., dog barking) to identify familiar animals? These three sound types differ in arbitrariness along a continuum, with speech as the most arbitrary and vocalizations as the least arbitrary cue. We ask whether these sounds are equally effective as acoustic cues in guiding children's attention to animals in a visual scene. Children will hear a sound cue while looking at images of two familiar animals, one that matches and one that does not match the cue, and we will compare children's looking to the matching animal when hearing the target sound.

One of three patterns of results is most likely to emerge: The first is that children are faster to identify animal names than onomatopoeic words, and faster to identify onomatopoeic sounds than animal vocalizations. This pattern of results could be predicted by computational models that use frequency as a crucial determinant of speed of processing (e.g., McMurray et al. (2012)). That is, children in urban environments are more likely to hear the names of animals than their vocalizations, resulting in more practice interpreting speech (high SES children might hear thousands of words daily, but not nearly as many animal sounds). This pattern of results could also be predicted by developmental accounts privilege language cues early development – either for getting children's attention (Vouloumanos & Werker, 2004, 2007a, 2007b) or for the fact that words refer to objects directly (Waxman & Gelman, 2009) and might evoke more category-diagnostic features

compared to environmental sounds (Edmiston & Lupyan, 2015).

The second possible pattern of results is that children are faster to identify animal vocalizations than onomatopoeic sounds, and faster to identify onomatopoeic sounds than animal names. This pattern of results could be predicted by accounts that propose that non-arbitrary sounds link directly to semantic representations, while words first activate lexical representations before reaching semantics (Chen & Spence, 2011, 2013). The fact that children have experience interpreting environmental sounds (e.g., balls bouncing, things falling) before learning to interpret speech referentially could also predict an advantage for environmental sounds. A third possibility is that children are equally efficient in exploiting these three sound types to guide their attention to a familiar animal. This pattern of results would parallel that of previous studies that showed little difference in the processing of environmental sounds and words (Cummings et al., 2009).

Our second question is whether two-year-olds make similar inferences when mapping a novel name and a novel animal vocalization to an unfamiliar animal. In a mutual exclusivity task, children will hear a novel animal name or novel animal vocalization (instead of a familiar one), while looking at the picture of a familiar and a novel animal (instead of two familiar objects). We compare children's proportion of looking to the novel animal when hearing one of these two sound cues. Considering dozens of studies on children's disambiguation biases, we predict that children will look at a novel animal when hearing a novel name. Thus, one of two patterns of results is most likely to emerge: The first is that children look at a novel animal when hearing a novel name, but are show no looking preference when hearing a novel animal vocalization. This pattern of result would be compatible with lexical accounts that predict disambiguation only within the domain of word learning, or by pragmatic accounts that predict disambiguation only within communicative contexts. The second pattern of findings is that children look at a novel animal when hearing a novel name and when hearing a novel animal vocalization, with comparable performance across these two conditions. This pattern would be compatible with

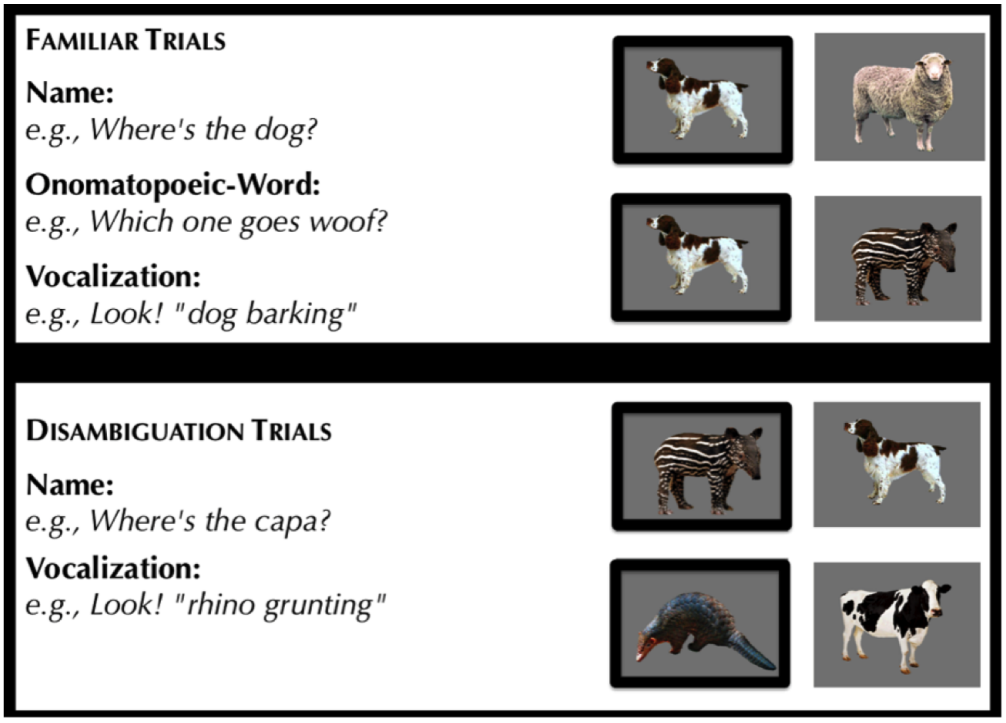

Figure 1. Trial types in Experiments 1 and 2 organized by type of cue: Familiar vs. Novel. The target animal for each trial type is on the left.

accounts that propose that disambiguation biases emerge from domain-general learning mechanisms that look for regularities in complex input.

Method

Participants. Participants were 23 32-month-old children (M=31.10; range = 30,32, 12 girls. All were reported by parents to be typically developing and from families where English was the dominant language. Two participants were excluded due to fussiness. Children were from mid/high-SES families.

Visual stimuli. The visual stimuli included pictures of four Familiar animals (horse, dog, cow, sheep) and two Novel animals (pangolin, tapir). According to parental report, the familiar animals were known by all children. Parents also reported that the novel animals were completely unfamiliar to the children. Each animal picture was centered on a grey background in a 640 x 480 pixel space

Auditory stimuli. The auditory stimuli consisted of sounds that were either Familiar or Novel to 32-month-olds. Figure 1 serves as a guide to the different sound types. The Familiar sounds were used in Familiar Trials, and consisted of one of three different sounds: names (horse, dog, cow and sheep), onomatopoeic words (neigh, woof-woof, moo and baa), and vocalizations (horse neighing, dog barking, cow mooing and sheep baaing). The Novel sounds were used in Disambiguation Trials, and consisted of one of two types of sounds: names (capa, nadu) and vocalizations (rhino grunting, gorilla snorting).

Trials in which the auditory cue was a familiar or novel animal name (e.g., Where's the dog?) or a familiar or novel lexical sound (Which one goes woof-woof?) began with a brief carrier frame. The duration of the target cue was 810 ms for lexical sounds and 750 ms for animal names. The intensity of the phrases was normalized using Praat speech analysis software (Boersma, 2002).

Trials with familiar or novel animal vocalizations began with a single word, used to draw children's attention (e.g., Look! "dog barking"). Familiar animal vocalizations were selected based on prototypicality. After selecting at least three vocalizations for each familiar animal, the authors voted on the one that we thought would be most easily recognized by children. Choosing the novel animal vocalizations was more challenging. A group of research assistants selected from different websites several vocalizations that they judged as unfamiliar. From these vocalizations, we selected two (i.e., rhino grunting and gorilla snorting) that we judged were equally likely to be produced by the 6 familiar and 2 novel animals based on the their size and vocal tract characteristics. These vocalizations were also maximally distinct from each other and from the familiar animal vocalizations and expected to be unfamiliar to children. We counterbalanced the vocalizations that were paired with the two novel animals, in order to control for the possibility that children judged one of the two novel animals as more likely to produce one of the novel vocalization. All children were reported by parents to have had no exposure to the novel animal's natural vocalizations. The duration of the target animal vocalizations was 2000 ms. More details about trial types

and conditions in Figure 1 will be given in the Procedure section.

Familiarization books. There were two main reasons for us to want to make sure that children knew the familiar onomatopoeic words and animal vocalizations before we administered our experiment. First, we wanted to make sure that any differences we observed in children's performance within Familiar-Animal Trials was due to processing speed and could not be explained by the fact that children were not familiar with one of the sound types. The looking-while-listening procedure has been shown to capture differences in processing efficiency even when words are considered "known" by offline reaching tasks or parental report (Fernald et al., 2008). Second, we wanted to make sure that children knew the pairings between familiar vocalizations and animals, a potential prerequisite for success on Disambiguation trials. Since we were working with children from mid/high-SES families growing up in an urban environment, we were particularly concerned that they would not be familiar with many animal vocalizations or onomatopoeic words. To ensure that all children had at least some experience with the familiar animals and familiar auditory cues used in our study, we gave two children's books to parents, both titled Sounds on the Farm, a week before their visit. Parents were instructed to share each book with their child for 5 to 10 min on at least three days prior to the experiment. The first book consisted of colorful pictures of each familiar animal and text designed to prompt parents to produce each animal's lexical sound (e.g., Wow, look at all those cows! This cow says moo, moo!). To give children exposure to the natural animal vocalizations, we used a Hear and There book, which contained buttons that children could press to hear the actual noise that each animal produces.

Procedure. Since we were interested in detecting differences in processing between sounds that we expected to be familiar to children, we choose to access speed and accuracy in identifying the correct target picture with the looking-while-listening (LWL) procedure (see Fernald, et al, 2008). Previous studies have shown that even when objects are reported by parents as familiar to their children, or when children are at ceiling in offline reaching

tasks, these real-time processing measures can capture meaningful differences in processing. These differences correlate to properties of the sound stimuli (e.g., word-frequency) and different aspects of the child's experience (e.g., their age, socioeconomic status, amount of parental talk). Looking-time measures have also been used in Disambiguation tasks with children from different ages, capturing differences in accuracy that relate to children's age and vocabulary size (Bion et al., 2013).

On each trial, a pair of pictures was presented on the screen for approximately 4 s, with the auditory stimuli starting after 2 s, followed by 1 s of silence. As seen in Figure 1, we have two main trials types, Familiar Trials and Disambiguation trials, paralleling our original two research questions on children's processing of familiar and novel auditory cues.

Within the Familiar Trials, we have three different sub-trials: name, onomatopoeic word, and vocalization. On 8 Name trials, each familiar animal served as the target twice and was paired once with another familiar animal and once with a novel animal. On 8 Onomatopoeic-word trials, each familiar animal served as the target twice. On 16 Vocalization trials, each familiar animal served as the target four times, paired twice with another familiar animal and twice with a novel animal. These three familiar sound types should allow us to answer our first research question, asking whether names, onomatopoeic words, or animal vocalizations, are equally effective as acoustic cues in guiding children's attention to animals in a visual scene.

Within the Disambiguation Trials, we have two different sub-trials: name, and vocalizations. On 6 Name trials, each novel animal was labeled three times with a novel animal name (i.e., capa, nadu), always paired with a familiar animal. On 8 Vocalization trials, each novel animal vocalization served as the target four times and was paired with each familiar animal once. These two sound types should allow us to answer our second research question, asking whether two-year-olds make similar inferences when mapping a novel name and a novel animal vocalization to an unfamiliar animal.

These different trial types were administered in two different visits. The Familiar and

Disambiguation Trials with animal names and onomatopoeic words were administered during children's first visit. The Familiar and Disambiguation Trials with the animal vocalizations were administered during their second visit. We administered the animal vocalizations on the second visit to allow children to become familiar with the procedure and to give parents additional time to use the familiarization books with the vocalizations with their children. During each visit, five Filler trials were interspersed throughout to add variety and maintain children's attention. Pairings of the novel animal and name, and side of presentation of target animals, were counterbalanced across participants. Caregivers wore darkened sunglasses so that they could not see the pictures and influence infants' looking throughout the 5-min procedure.

Measures of processing efficiency. Participants' eye movements were video-recorded and coded with a precision of 33 ms by observers who were blind to trial type. Inter- and intra-observer reliability checks were conducted for all coders. For 25% of the subjects, two measures of inter-observer reliability were assessed. The first was the proportion of frames (33-ms units) on each trial on which two coders agreed. In this case, agreement was 98%. However, because this analysis included many frames on which the child was maintaining fixation on one picture, we also calculated a more stringent test of reliability. This second measure focused only on shifts in gaze, ignoring steady-state fixations in each trial on which agreement was inevitably high. By this more conservative measure, coders agreed within one frame on 94% of all shifts.

Accuracy: On those trials in which the infant was fixating a picture at the onset of the speech stimulus, accuracy was computed by dividing the time looking to the target object by the time looking to both target and distracter, from 300 to 2500 ms from the onset of the target word. Accuracy before 300 ms was not included because shifts to the target occurring in this window had presumably been initiated before the onset of the noun. This analyses window was chosen because of the longer duration of the animal vocalizations (2 s.) and because of the introduction of novel auditory cues. A single analyses window was used

for all trial types for consistency. Mean accuracy was then computed for each participant on each trial type.

Reaction time: We calculated reaction time (RT) on those trials on which participants were looking at the distractor animal at the beginning of the sound. RT on each trial was the latency of the first shift to the correct animal within a 300- to 1,800-ms window from sound onset, as typically done in studies using this procedure (Fernald et al., 2008).

Results and discussion

Familiar Trials: Using familiar animal names, onomatopeic sounds, and animal vocalization to identify familiar animals:. Our first question is whether a familiar animal name, onomatopoeic word, and animal vocalization are equally effective in guiding children's attention to an animal in the visual scene. Figure 2A shows children's looking behavior over time on the LWL procedure (Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998). In order to capture children's speed of processing, we show children's responses on trials in which they start looking at the wrong animal. From sound onset onward, we show the mean proportion of trials in which children were looking at the correct picture, every 33ms, with different lines representing children's responses on Name trials (black), Onomatopoeic-word trials (light grey), and Vocalization trials (dark grey). The y-axis shows the mean proportion of trials on which children were looking at the correct animal. The x-axis represents time from sound onset in milliseconds. Around 750ms from sound onset there is already a substantial difference in the mean proportion of trials in which children are looking at the correct animal depending on whether they heard an animal name or an animal vocalization. Children are looking at the correct animal in a greater proportion of trials when they hear a familiar animal name, as compared to when they hear an animal vocalization, with performance on trials with onomatopoeic words falling between the two.

To quantify these differences, we fit Bayesian mixed-effects regression models using the

`rstanarm` (Gabry & Goodrich, 2016) package in R (3.4.1, R Core Team, 2017)[1]. The mixed-effects approach allowed us to model the nested structure in our data by including random intercepts for each participant and item and a random slope for each item. We used Bayesian methods in order to quantify support in favor of null hypotheses of interest – in this case, the absence of a difference in real-time processing across the different familiar cue types. To communicate the uncertainty in our estimates, we report the 95% Highest Density Interval (HDI) around the point estimates of the group means and the difference in means. The HDI provides a range of plausible parameter values given the data and the model. All analysis code can be found in the online repository for this project: https://github.com/kemacdonald/anime.

We computed reaction time (RT) as the mean time it took them to shift to the correct picture on trials in which they were looking at the wrong picture at sound onset for the three cue types. To make RTs more suitable for modeling on a linear scale, we analyzed responses in log space using a logistic transformation, with the final model was specified as: $log(RT) \sim cue\_type + (1 + sub\_id \mid item)$.

Figure 2 shows the data distribution for each participant's RT, the estimates of condition means, and the full posterior distribution of condition differences across the different cue types. Children were faster to identify the target animal while hearing its name (M_{name} = 876.35 ms), as compared to its onomatopeic animal sound ($M_{onomatopeia}$ = 649.70 ms), and its vocalization ($M_{vocalization}$ = 720.38 ms). The difference between RTs for the name and onomatopeic animal sounds was -70.67 ms with a HDI from -212.86 ms to 68.88 ms. While the null value of zero difference falls within the 95% HDI, 83.50% of the credible values fall below the null, providing some evidence for faster processing of animal names. The average difference in children's RT between the name and vocalization trials was -226.65 ms with a HDI from -361.37 ms to -85.23 ms and 99.95% of the credible values falling below

[1]We, furthermore, used the R-packages *here* (0.1, Müller, 2017), *knitr* (1.20, Xie, 2015), *papaja* (0.1.0.9492, Aust & Barth, 2017), and *tidyverse* (1.2.1, Wickham, 2017).

zero, providing strong evidence that children processed names more efficiently compared to vocalizations. Finally, the average difference in children's RT between the onomatopeic sounds and vocalization trials was -155.98 ms with a HDI from -301.18 ms to -5.77 ms and 97.95% of the credible values falling below zero.

Together, the RT modeling results provide strong evidence that children processed animal names around 227 ms faster than animal vocalizations, with almost all of the estimates of the plausible RT differences falling below the null value of zero. There was slightly weaker evidence that children processed animal names more efficiently compared to onomatopeic animal sounds but strong evidence of faster processing of onomatopeic animal sounds compared to animal vocalizations. In sum, there was evidence of a graded effect of cue type on RTs with names being faster than onomatopeic animal sounds, which were faster than animal vocalizations.

Children's reaction times showed a difference in speed of processing for names, onomatopoeic words, and vocalization. Next, we estimated children's attention to the target image over the course of the trial to ask whether the three trial types were equally effective cues to guide their attention to a familiar animal. We computed accuracy over a window from 300 to 2500 ms after the onset of the cue. The upper left panel of Figure 3A shows children's proportion looking to the target for each trial type. Each point represents a single participant's Accuracy, the grey line shows the full distribution of the data, and the horizontal line shows the median value. The orange points represent the most likely estimate for the mean proportion looking with the error bars showing the 95% HDI. Visual inspection of the plot suggests two things: (1) children reliably looked to the correct animal after hearing each of the three familiar cues and (2) children's overall looking behavior was strikingly similar across conditions.

Next, we quantified the strength of evidence for the absence of any condition differences and for the difference from random responding. We estimated the mean proportion looking for each trial type using a Bayesian linear mixed-effects model with the same specifications

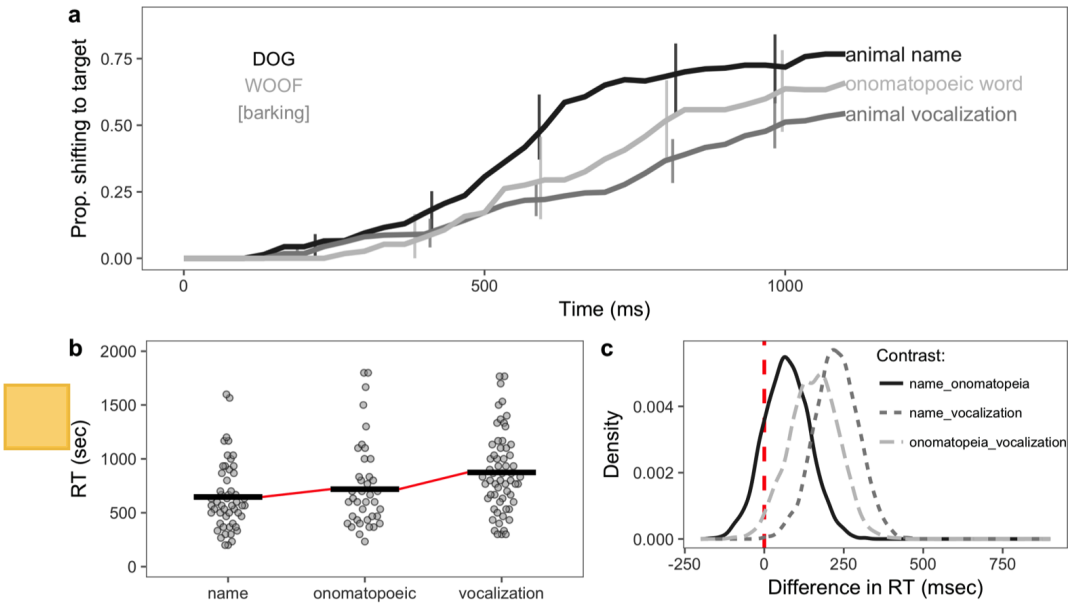

Figure 2. Reaction Time (RT) results for Experiment 1. Panel A shows the timecourse of children's looking to the target animal after hearing a familiar animal name (black), onomatopoeic word (light grey), or animal vocalization (dark grey). Panel B shows the distribution of RT data across conditions. Each grey point shows a RT for a single trial. The black bar represent the most likely estimate of the condition means. The red lines connect the condition means to illustrate shifts in the RT distributions. Panel C shows the posterior distribution of credible RT differences across conditions. Color and linetype represent the contrast of interest and the red vertical dashed line represents the null value of zero condition difference. All error bars represent 95% Highest Density Intervals.

as the RT model described above. We transformed the proprotion looking scores using the empirical logit, with the final model as: $logit(accuracy) \sim cue\_type + (1 + sub\_id \mid item)$. The orange points in Figure 3A show mean Accuracy for each familiar cue type (M_{name} = 0.66, $M_{onomatopeia}$ = 0.68, and ($M_{vocalization}$ = 0.67). Children's looking to the target image was reliably different from a model of random looking behavior across all conditions, with the null value of 0.5 falling well outside of the range of plausible values (see the difference between the horizontal dashed line and error bars in Figure 3A).

Moreover, the three cues types were equally effective in guiding children's attention to

the target animal over the course of the trial as shown by the high overlap in the posterior distributions of Accuracy and that the null value of zero difference fell within the HDI for all group comparisons (name vs. onomatopeia: $\beta_{diff} = 0.03$, 95% HDI from -0.07 to 0.11; name vs. vocalization: $\beta_{diff} = 0.01$, 95% HDI from -0.05 to 0.08; onomatopeia vs. vocalization: $\beta_{diff} = -0.01$, 95% HDI from -0.09 to 0.07). These results provide evidence that the animal vocalizations were equally effective at directing overall looking behavior to identify familiar animals as their names or onomatopeic sounds if children have enough time to process the cue.

Disambiguation Trials: Using novel animal names and animal vocalization to disambiguate novel animals. Our next question is whether children would orient to a novel animal after hearing a novel animal name or vocalization, thus showing evidence of one-to-one biases for the vocalizations that animals produce. We focus on children's accuracy, comparing their proportion of looking to the novel animal against chance performance and across cue types. The orange points in Figure 3A show Accuracy group means for each novel cue type ($M_{name} = 0.69$, $M_{vocalization} = 0.70$). Children's looking to the target image was reliably different from a model of random looking behavior across both cue types, with the null value of 0.5 falling well outside of the range of plausible values.

Moreover, the novel animal name and animal vocalizations were equally effective in guiding children's attention to the target animal over the course of the trial (name vs. onomatopeia: $\beta_{diff} = 0.01$, 95% HDI from -0.05 to 0.08). This result provides strong evidence that the animal vocalizations were equally effective at directing overall looking behavior to identify familiar animals as their names or onomatopeic sounds if children have enough time to process the cue.

Therefore, children seem to have one-to-one biases for the vocalizations that animals produce already at 32 months of age, the earliest age at which the disambiguation effect has been observed in a domain other than word learning. There were no significant differences between children's reaction time for novel animal names or vocalizations.

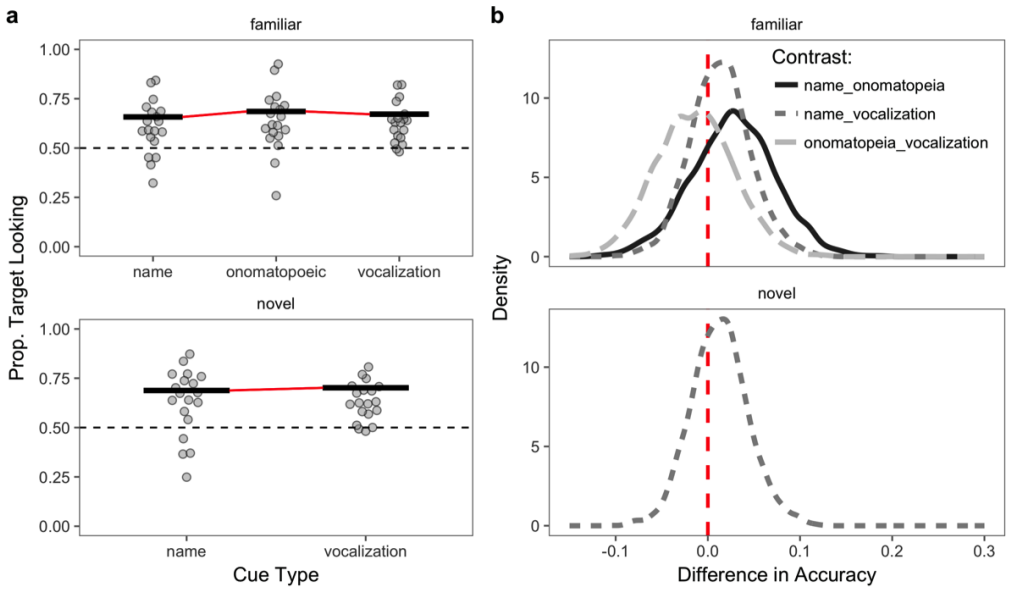

Figure 3. Accuracy results for Experiment 1 for familiar (upper panels) and novel (lower panels) trials. Panel A shows the data distribution and model estimates for Accuracy of children's looking behavior. Panel B shows the full posterior distribution over model estimates of differences in accuracy across conditions. The vertical dashed line represents the null model of zero difference in Accuracy. All other plotting conventions are the same as in Figure 2.

Experiment 2

One issue that has received much attention in recent years concerns the relation between children's referent selection and retention abilities. While earlier studies tended to conflate disambiguation strategies and children's word learning, more recent studies suggest that these two abilities should not be conflated (Bion, Borovsky, & Fernald, 2013; Horst & Samuelson, 2008)

Horst and Samuelson (2008) examined both referent selection and retention in four experiments with 2-year-olds. When children were shown a novel object among familiar objects, they selected the novel object when hearing a novel label, as found in previous studies. But surprisingly, on retention trials 5 min later, these children showed no evidence of remembering the names of the novel objects they had previously identified. Using a

looking-time task, Bion et al. (2013) replicated these findings in a study with 18-, 24-, and 30-month-old infants using looking time measures of performance.

Experiment 2 asks whether children can retain the link created through disambiguation between a novel animal and a novel animal vocalization. We aim to replicate the findings from Experiment 1, showing that children can identify familiar animals based on the vocalizations they produce and use novel vocalizations to disambiguate novel animals. We predict that children will succeed in disambiguation trials, but will show little evidence of retention on subsequent disambiguation trials, paralleling the findings of earlier studies with linguistic stimuli.

Method

Participants. Participants were 23 31-month-old children (M = 31.10 months; range = 30 - 32), 12 girls. All children were typically developing and from families where English was the dominant language.

Visual stimuli. The visual stimuli were the same as in Experiment 1, except for the novel animals (aardvark and capybara), which replaced the novel animals (pangolin and tapir) used in Experiment 1 (see animals in Figure 4). We decided to change the novel animals in order to confirm that our results were not restricted to the particular stimuli set in Experiment 1. All children were reported by parents to have had no exposure to the novel animals.

Auditory stimuli. The auditory stimuli consisted of the same familiar and novel animal vocalizations as in Experiment 1.

Familiarization books. As in Experiment 1, we sent home a children's book to ensure that all participants had at least some exposure to the familiar animals and auditory cues. Since, in Experiment 2, we were interested in the natural animal vocalizations and not the names/lexical sounds, only the Hear and There Sounds on the Farm book was used. Instructions given to the parents were the same as in Experiment 1, and the book was sent

Figure 4. Trial types in Experiments 4 organized by type of trial. Children hear familiar and novel vocalizations. The target animal for each trial type is on the left.

home a week before the visit.

Procedure. Experiment 2 consisted of one visit. Each child saw 30 trials, consisting of three trial types (Figure 4). The 16 Familiar trials and 8 Disambiguation trials were identical in structure to the Vocalization trials Experiment 1. In addition, on 6 Retention trials, the two novel animals were presented side by side, with each serving as the target three times. The same coding and speed/accuracy measures were used as in Experiment 1.

Results and discussion

Retention of the link between a novel animal and a novel vocalization:. Figure 5A shows children's proportion looking to the target animal after hearing a familiar or a novel animal vocalization over the same analysis window used in Experiment 1 (300 to 2500 ms after the onset of the vocalization). Visual inspection of the figure suggests that children successfully oriented to the target image after hearing both familiar and novel animal vocalizations. The orange points show mean proportion target looking for familiar ($M_{familiar}$ = 0.66, disambiguation ($M_{disambiguation}$ = 0.64, and retention trials ($M_{retention}$ = 0.50).

Children's looking to the target image was reliably different from random-looking

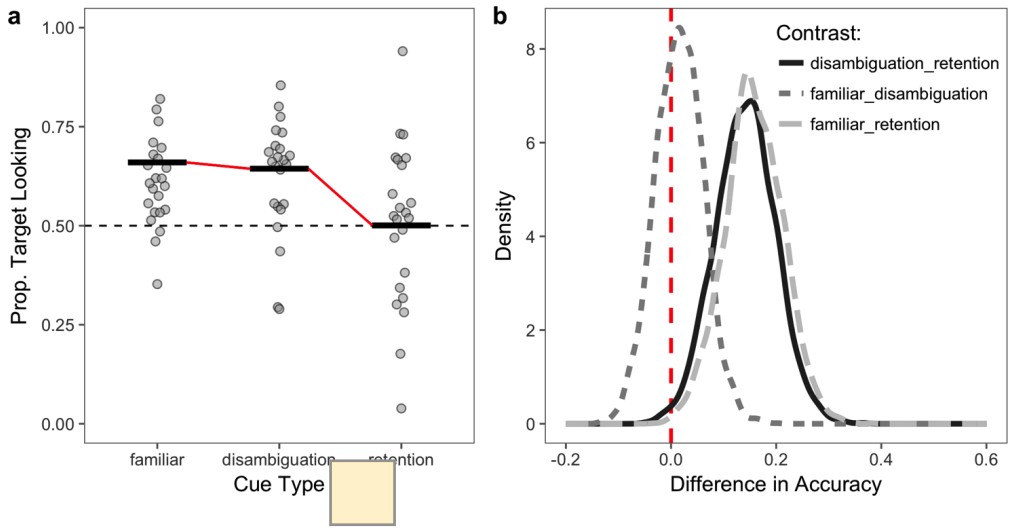

Figure 5. Accuracy of responses to familiar and novel animal vocalizations in Experiment 2. Panel A shows the data distribution alongside the model estimates of mean Accuracy across the different trial types. Panel B shows the full posterior distribution over model estimates of differences in accuracy across trial types. The vertical dashed line represents the null model of zero difference. All other plotting conventions are the same as in Figures 1 and 2.

behavior for both familiar and disambiguation trials, with the null value of 0.50 falling well outside of the range of plausible values. Moreover, the novel animal vocalizations and the familiar animal vocalizations were equally effective in guiding children's attention to the target animal over the course of the trial (familiar vs. disambiguation: $\beta_{diff} = 0.02$, 95% HDI from -0.08 to 0.11).

In contrast to children's success on familiar trials and disambiguation trials, they did not show evidence of retaining the link between the novel animal and the novel animal vocalization, with the null value of 0.5 proportion looking falling well within the range of plausible estimates. Moreover, there was strong evidence that children were less accurate on retention trials compared to both disambiguation trials ($\beta_{diff} = 0.14$, 95% HDI from 0.03 to 0.25) and familiar trials ($\beta_{diff} = 0.16$, 95% HDI from 0.05 to 0.27).

Three findings emerged from the accuracy analysis: First, children oriented to a familiar animal after hearing a familiar animal vocalization. Second, children oriented to a

novel animal after hearing a novel animal vocalization. These two results are an internal replication of the key findings from Experiment 1 in a new sample. In addition, we found that children did not show evidence of retaining the link between a novel animal vocalization and a novel animal. These results and their implications are discussed in more detail in the following section.

General Discussion

Three main findings emerged from this work. The first finding was that 30-month-olds responded fastest to a familiar animal name and slowest to a familiar animal vocalization, with onomatopoeic sounds falling somewhere in between. Children could, however, identify the familiar animals after hearing any of on these three sound types. The second finding was that children showed disambiguation biases for the types of vocalizations that animals produce, similar to their biases in word learning. The third finding was that these biases do not necessarily lead to learning, as children were not successful in retaining the link between novel animals and their vocalizations. This lack of retention parallels the findings of recent word-learning studies (see McMurray, Horst, Toscano, and Samuelson (2009)) and emphasizes the theoretical importance of disentangling processes of disambiguating reference, an in-the-moment phenomenon, and word learning, which occurs over a longer timescale.

In our study, we found a processing speed advantage for words over other meaningful sounds. Some theories of language development argue that words are unique stimuli because they refer to objects in the world (Waxman & Gelman, 2009), while other theories argue that words are special because they activate conceptual information more quickly, accurately, and in a more categorical way than nonverbal sounds (Edmiston & Lupyan, 2015). It is also possible that words and nonverbal sounds might be processed by different brain regions, with words being accessed more rapidly. A second explanation for the advantage for words might be differences in sheer frequency in the input. At least in our sample, it is safe to assume that children have heard the word cow many more times than they have heard an actual cow

mooing. Frequency effects have been robustly demonstrated in the processing of words, with adults being faster to recognize words that they hear more frequently (Dahan, Magnuson, & Tanenhaus, 2001). A final explanation is that words are very effective at presenting a lot of information in a short period. When children see a simple visual world consisting of a dog and a sheep, the first phoneme of the target word is already sufficient to determine the animal that is likely to be talked about next.

Much less is known about children's and adults' processing of onomatopoeic sounds. Hashimoto et al. (2006) compared brain responses to nouns, animal sounds, and onomatopoetic sounds, and found that onomatopoeic sounds were processed by extensive brain regions involved in the processing of both verbal and nonverbal sounds. Cummings et al. (2009) argues that onomatopoeic sounds might provide young children with information about intermodal associations, bridging their understanding of non-arbitrary environmental sounds and arbitrary word-object associations. Fernald and Morikawa (1993) reported that 52% of Japanese mothers used onomatopoeic sounds to label target objects, while only 1 in 30 American mothers did so. While our results do not speak directly to these theoretical issues, they do suggest that onomatopoeic sounds function like words in that they are capable of activating conceptual representations that drive children's visual attention to seek the physical referent of the sound. However, there was some evidence that children processed onomatopeic sounds less efficiently compared to words in our task.

Our second finding was that children looked at a novel animal when hearing a novel animal vocalization, with accuracy comparable to their disambiguation of novel animal names. Bloom (2002) outlines three different theories that could explain children's disambiguation biases: These biases could be a specifically lexical phenomenon that applies only to words (lexical account), a product of children's theory of mind restricted to communicative situations (pragmatic account), or a special case of a general principle of learning that exaggerates regularities across domains (domain-general account). By using animal sounds, this study provides an important data point since the theoretical accounts

make different predictions for children's looking behavior in response to a stimulus that is non-linguistic and non-communicative.

Previous studies contrasted lexical-specific and pragmatic accounts. For example, diesendruck2001children found that children expect speakers to use consistent facts to refer to objects, and they select a novel object when hearing a novel fact. Recent studies suggest that different strategies might be used to make inferences about speakers' communicative intent and the meaning of a novel word. Autistic children who are known to have pragmatic deficits show disambiguation biases and select a novel word when hearing a novel object (Preissler & Carey, 2005). Moreover, disambiguation biases for words are correlated with vocabulary, and disambiguation biases for facts are correlated with social-pragmatic skills (Marchena et al., 2011). These findings suggest that disambiguation biases for words might not be motivated uniquely by pragmatic inferences, but they do not provide evidence for or against domain-general accounts of the disambiguation bias.

Relatively few studies have looked at disambiguation biases in non-linguistic domains. moher2010one showed that three-year-olds link different voices to unique faces, showing that one-to-one biases might extend to other communicative domains. However, Yoshida, Rhemtulla, and Vouloumanos (2012) found that adults in a statistical learning task were less likely to show evidence of using disambiguation biases to learn nonspeech sounds even though this behavior would have facilitated task performance. These results suggest that at some point in development disambiguation constraints may operate more strongly over speech compared to nonspeech sounds. Critically, these results do not provide evidence against a domain-general account since participants had no reason to expect that the mapping between random non-linguistic sounds and objects should be mutually exclusive. It could be that similar learning strategies might be applied to non-linguistic sounds when they become meaningfully related to objects in the environment or relevant for communication with other people.

The work reported here demonstrates that young children do show evidence of

disambiguation biases in a non-linguistic and non-communicative domain. These findings support predictions made by domain-general accounts that explain disambiguation biases as the byproduct of a system that attempts to find regularities in complex learning tasks that involve consistent mappings. Previous Connectionist and Bayesian models of word learning showed that disambiguation biases emerge as children are exposed to consistent mappings between words and objects, without built-in constraints on the types of meanings words could have (Frank et al., 2009; McMurray et al., 2012; Regier, 2003). In principle, these same biases could emerge if these models attempted to map animal vocalizations to animals and there were consistent co-occurrences present in the learning environment.

One open question is whether a single disambiguation mechanism can account for the diverse set of contexts in which children show disambiguation behaviors. Here, we found that children could disambiguate stimuli other than words and facts, suggesting at least the existence of a domain-general mechanism that leads to disambiguation. A preference for parsimony suggests that we should favor of a single mechanism. But as pointed out by recent computational work, it is possible that different mechanisms jointly contribute to disambiguation behavi xplaining findings across different populations and contexts (Lewis & Frank, 2013). Thus, it is possible that the same behavior - selecting a novel object when hearing a novel auditory stimulus - might result from different computational mechanisms or motivations depending on the task at hand, children's age, or the particular stimuli set and assumptions about the people involved in the interaction. That is, children could use a domain-general mechanism to learn about novel animal vocalizations, and they could use a lexical and pragmatic constraint to learn about novel words.

Our third finding was that one-to-one biases for animal vocalizations do not necessarily lead to retention of the link between a novel animal and a novel vocalization. This finding dovetails with recent cross-situational models of early word learning that emphasize the separate processes of figuring out reference in the moment and learning word-object labels over time (McMurray et al., 2012). McMurray and colleagues propose that referent-selection

requires that children give their best guess about a new word's meaning in a specific ambiguous situation, but that learning operates over a much longer timescale, requiring multiple exposures to build up stable word-object links. Although disambiguation can be viewed as the product of learning that has occurred up to that point, for younger children it does not necessarily result in learning. These claims are also supported by evidence from studies on early word learning using online and offline measures of retention (Bion et al., 2013).

These results also add to a recent body of work that encourages us to think differently about disambiguation biases. This work has emphasized the role of experience, showing that the tendency to select a novel object when hearing a novel word is not robustly present across populations. For example, bilingual children, children from lower socioeconomic status, children who receive less language input, and children with less structured vocabularies or smaller vocabulary sizes, take longer to show evidence of disambiguation biases (Bion et al., 2013; Yurovsky, Bion, Smith, & Fernald, 2012). Other studies have problematized the relation between disambiguation biases and word learning, showing that success in referent selection does not necessarily mean that the link between the novel word and novel object will be retained (Bion et al., 2013; Horst, McMurray, & Samuelson, 2006; McMurray et al., 2012). And the current study adds to recent studies taking a fresh look at an old question: the scope of disambiguation biases (Suanda & Namy, 2012, 2013). Taken together, these findings suggest that it is important to consider variability in the emergence, function, and scope of any language learning processes that might be characterized as universal based on studies using a particular population or a specific stimulus type.

Finally, our results emphasize that children's learning about objects in their environment involves more than learning their names. Before object names are learned, sounds and actions might form the basis on which objects are conceptualized. For example, children might see barking as a defining feature of dogs and may say bow-wow in response to the picture of a dog, even before they learn the animal name (Nelson, 1974). Learning the

meaning of an object, therefore, requires learning several cross-modal associations, including learning the object's texture, smell, as well as its sounds and names. Children do not have explicit constraints that freshly baked cookies should have only one smell. Yet, they might recognize and get excited about the familiar smell coming from the kitchen and might assume their mothers are baking something new when smelling something unfamiliar.

Conclusions

Children use different types of knowledge to make sense of a constantly changing world. They might identify animal vocalizations based on the shape of the vocal tract of the animal, its location and size, and their previous knowledge about animal vocalizations. Importantly, these cues normally converge in helping children identify an animal in the environment. The same is true for their identification of referents for words. Children can identify the referent for a word based on semantics (Goodman, McDonough, & Brown, 1998), cross-situational statistics (Smith & Yu, 2008), syntax (Brown, 1957), and pragmatic and social cues (Baldwin, 1993), and disambiguation biases [markman1991whole]. As children grow older, these different sources of information provide converging evidence that a novel word should refer to a novel object. Children can rely on their knowledge about the world, speakers, and on their previous experiences with words to figure out what speakers are talking about – a task we continue to do throughout our lives when learning new words and interpreting complex sentences.

References

Aust, F., & Barth, M. (2017). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(02), 395–418.

Bates, E., MacWhinney, B., & others. (1989). Functionalism and the competition model. *The Crosslinguistic Study of Sentence Processing*, *3*, 73–112.

Bion, R. A., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30months. *Cognition*, *126*(1), 39–53.

Bloom, P. (2002). *How children learn the meaning of words*. The MIT Press.

Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, *55*(1), 1.

Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1554.

Chen, Y.-C., & Spence, C. (2013). The time-course of the cross-modal semantic modulation of visual picture processing by naturalistic sounds and spoken words. *Multisensory Research*, *26*(4), 371–386.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*(2), 417–431.

Cummings, A., & Ceponiene, R. (2010). Verbal and nonverbal semantic processing in children with developmental language impairment. *Neuropsychologia*, *48*(1), 77–85.

Cummings, A., Ceponiene, R., Koyama, A., Saygin, A., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, *1115*(1), 92–107.

Cummings, A., Saygin, A. P., Bates, E., & Dick, F. (2009). Infants' recognition of meaningful

verbal and nonverbal sounds. *Language Learning and Development*, *5*(3), 172–190.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*(5), 630–641.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in japanese and american mothers' speech to infants. *Child Development*, *64*(3), 637–656.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, *44*, 97.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*(5), 578–585.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218–228.

Gabry, J., & Goodrich, B. (2016). Rstanarm: Bayesian applied regression modeling via stan. R package version 2.10. 0.

Goodman, J. C., McDonough, L., & Brown, N. B. (1998). The role of semantic context and memory in the acquisition of novel nouns. *Child Development*, *69*(5), 1330–1344.

Hashimoto, T., Usui, N., Taira, M., Nose, I., Haji, T., & Kojima, S. (2006). The neural mechanism associated with the processing of onomatopoeic sounds. *Neuroimage*,

31(4), 1762–1770.

Horst, J. S., McMurray, B., & Samuelson, L. K. (2006). Online processing is essential for learning: Understanding fast mapping and word learning in a dynamic connectionist architecture. In *Proceedings of the cognitive science society* (Vol. 28).

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157.

Lewis, M., & Frank, M. (2013). An integrated model of concept learning and word-concept mapping. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170.

Marchena, A. de, Eigsti, I.-M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*(1), 96–113.

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. *Perspectives on Language and Thought: Interrelations in Development*, 72–106.

McCleery, J. P., Ceponiene, R., Burner, K. M., Townsend, J., Kinnear, M., & Schreibman, L. (2010). Neural correlates of verbal and nonverbal semantic integration in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(3), 277–286.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831.

McMurray, B., Horst, J. S., Toscano, J. C., & Samuelson, L. K. (2009). Integrating connectionist learning and dynamical systems processing: Case studies in speech and

lexical development. *Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered*, 218–249.

Murray, M. M., Camen, C., Andino, S. L. G., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *Journal of Neuroscience*, *26*(4), 1293–1302.

Müller, K. (2017). *Here: A simpler way to find your files*. Retrieved from https://CRAN.R-project.org/package=here

Namy, L. L., & Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, *69*(2), 295–308.

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, *81*(4), 267.

Pena, M., Maki, A., Kovac, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702–11705.

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition*, *97*(1), B13–B23.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Regier, T. (2003). Emergent constraints on word-learning: A computational perspective. *Trends in Cognitive Sciences*, *7*(6), 263–268.

Scofield, J., & Behrend, D. A. (2007). Two-year-olds differentially disambiguate novel words and facts. *Journal of Child Language*, *34*(4), 875–889.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*(3), 545–559.

Suanda, S. H., & Namy, L. L. (2013). Young word learners' interpretations of words and

symbolic gestures within the context of ambiguous reference. *Child Development*, *84*(1), 143–153.

Vouloumanos, A., Druhen, M. J., Hauser, M. D., & Huizink, A. T. (2009). Five-month-old infants' identification of the sources of vocalizations. *Proceedings of the National Academy of Sciences*, *106*(44), 18867–18872.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276.

Vouloumanos, A., & Werker, J. F. (2007a). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164.

Vouloumanos, A., & Werker, J. F. (2007b). Why voice melody alone cannot explain neonates' preference for speech. *Developmental Science*, *10*(2), 169.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263.

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. Retrieved from https://CRAN.R-project.org/package=tidyverse

Woodward, A., & Hoyne, K. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, *70*(1), 65–77.

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.name/knitr/

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250.

Yoshida, K., Rhemtulla, M., & Vouloumanos, A. (2012). Exclusion constraints facilitate statistical word learning. *Cognitive Science*, *36*(5), 933–947.

Yurovsky, D., Bion, R. A., Smith, L. B., & Fernald, A. (2012). Mutual exclusivity and vocabulary structure. *N. Miyake, D. Peebles, & RP Cooper (Eds.)*, 1197–1202.

Figure captions

Figure 1. Trial types in Experiments 1 and 2 organized by type of cue: Familiar vs. Novel. The target animal for each trial type is on the left.

Figure 2. Reaction Time (RT) results for Experiment 1. Panel A shows the timecourse of children's looking to the target animal after hearing a familiar animal name (black), onomatopoeic word (light grey), or animal vocalization (dark grey). Panel B shows the distribution of RT data across conditions. Each grey point shows a RT for a single trial. The black bar represent the most likely estimate of the condition means. The red lines connect the condition means to illustrate shifts in the RT distributions. Panel C shows the posterior distribution of credible RT differences across conditions. Color and linetype represent the contrast of interest and the red vertical dashed line represents the null value of zero condition difference. All error bars represent 95% Highest Density Intervals.

Figure 3. Accuracy results for Experiment 1 for familiar (upper panels) and novel (lower panels) trials. Panel A shows the data distribution and model estimates for Accuracy of children's looking behavior. Panel B shows the full posterior distribution over model estimates of differences in accuracy across conditions. The vertical dashed line represents the null model of zero difference in Accuracy. All other plotting conventions are the same as in Figure 2.

Figure 4. Trial types in Experiments 4 organized by type of trial. Children hear familiar and novel vocalizations. The target animal for each trial type is on the left.

Figure 5. Accuracy of responses to familiar and novel animal vocalizations in Experiment 2. Panel A shows the data distribution alongside the model estimates of mean Accuracy across the different trial types. Panel B shows the full posterior distribution over model estimates of differences in accuracy across trial types. The vertical dashed line represents the null model of zero difference. All other plotting conventions are the same as in Figures 1 and 2.