- Communicative pressure on caregivers leads to language input that supports
- children's word learning
- Ben Morris¹ & Daniel Yurovsky²
 - ¹ University of Chicago
- ² Carnegie Mellon University

2

6 Abstract

Children do not learn language from passive observation of the world, but from interaction with caregivers motivated to communicate with them. Does this communicative pressure on caregivers lead them to structure children's linguistic input in a way that facilitates learning? As a case study, we analyze a canonical pedagogical behavior in caregivers: use 10 of ostensive labeling to teach children new words. Using a multi-method approach, we 11 show that this pedagogically supportive behavior can emerge from communicative pressure 12 alone, without any explicit pedagogical goals. First, in a corpus study, we show that 13 caregivers communicate by ostensive labeling precisely when children are likely to learn a new word. In an iterated reference game, we experimentally show that this strategy can 15 arise from pressure to communicate successfully with a less knowledgeable partner. Then, we show that speaker behavior in our experiment can be explained by a rational communication model that includes planning, without any explicit teaching goal. Finally, 18 in a series of simulations, we explore the language learning consequences of having a 19 communicatively-motivated caregiver. We show that under many parameterizations, simple 20 learning mechanisms interacting with a communicatively-motivated partner outperform 21 more powerful learning mechanisms. This perspective offers a first step toward a unifying, 22 formal account of both how linguistic input is structured and how children learn from it. 23 Keywords: communication; child-directed speech; language learning; computational 24 modeling

Word count: X

Communicative pressure on caregivers leads to language input that supports children's word learning

One of the most striking aspects of children's language learning is just how quickly
they master the complex system of their natural language (Bloom, 2000). In just a few
short years, children go from complete ignorance to conversational fluency in a way that is
the envy of second-language learners attempting the same feat later in life (Newport,
1990). What accounts for this remarkable transition?

Distributional learning presents a unifying account of early language learning:

Infants come to language acquisition with a powerful ability to learn the latent structure of
language from the statistical properties of speech in their ambient environment (Saffran,
2003). Distributional learning mechanisms can be seen in accounts across levels of language,
including phonemic discrimination (Maye et al., 2002), word segmentation (Saffran, 2003),
learning the meanings of both nouns (Smith & Yu, 2008) and verbs (Scott & Fischer,
2012), learning the meanings of words at multiple semantic levels (Xu & Tenenbaum,
2007), and perhaps even the grammatical categories to which a word belongs (Mintz,
2003). A number of experiments have confirmed both the early availability of distributional
learning mechanisms, and their potential utility across these diverse language phenomena
(DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Estes et al., 2007; Gomez & Gerken,
1999; Maye et al., 2002; Saffran et al., 1996; Smith & Yu, 2008; Xu & Tenenbaum, 2007).

However, there is reason to be suspicious about just how precocious statistical learning abilities are in early development. Although these abilities are available early, they are highly constrained by limits on other developing cognitive capacities. For example, infants' ability to track the co-occurrence information connecting words to their referents is constrained significantly by their developing memory and attention systems (Smith & Yu, 2013; Vlach & Johnson, 2013). In addition, computational models of statistical learning show that the rate of acquisition is highly sensitive to variation in environmental statistics.

For example, the Zipfian distribution of word frequencies and word meanings in the real
world can lead to very long learning times for rare words (Vogt, 2012). Further, a great
deal of empirical work demonstrates that cross-situational learning even in adults drops off
rapidly when participants are asked to track more referents, and also when the number of
intervening trials is increased–features likely typical of the naturalistic learning
environment (e.g., Yurovsky & Frank, 2015). Thus, precocious unsupervised statistical
learning appears to fall short of a complete explanation for rapid early language learning.

Even relatively constrained statistical learning could be rescued, however, if
caregivers structured their language in a way that simplified children's learning problem.

Caregivers do adjust many aspects of their communication when conversing with young

caregivers structured their language in a way that simplified children's learning problem.

Caregivers do adjust many aspects of their communication when conversing with young

children, in both their language (infant directed speech, Snow, 1977) and actions

(motionese, Brand et al., 2002)-and such modifications have been shown to yield learning

benefits across a number of language phenomena. For example, in phoneme learning,

infant-directed speech provides examples that seem to facilitate the acquisition of

phonemic categories (Eaves Jr et al., 2016). In word segmentation tasks, infant-directed

speech facilitates infant learning more than matched adult-directed speech (Thiessen et al.,
2005). In word learning scenarios, caregivers produce more speech during episodes of joint

attention with young infants, which uniquely predicts later vocabulary (Tomasello &

Farrar, 1986). Child-directed speech even seems to support learning at multiple levels in

parallel-e.g., simultaneous speech segmentation and word learning (Yurovsky et al., 2012).

For each of these language problems faced by the developing learner, caregiver speech

exhibits structure that seems uniquely beneficial for learning.

Under distributional learning accounts, the existence of this kind of structure is a
theory-external feature of the world that does not have an independently motivated
explanation. Such accounts view the explaining the process that generates structure in the
language environment as a problem separate from explaining language learning. However,
across a number of language phenomena, the language environment is not merely

supportive, but seems calibrated to children's changing learning mechanisms (Yurovsky, 2018). For example, across development, caregivers engage in more multimodal naming of 81 novel objects than familiar objects, and rely on this temporal synchrony between verbal labels and gesture most with young children (Gogate et al., 2000). The prevalence of such 83 synchrony in child-directed speech parallels infant learning mechanisms: young infants appear to rely more on synchrony as a cue for word learning than older infants, and 85 language input mirrors this developmental shift (Gogate et al., 2000). Beyond age-related changes, caregiver speech may also support learning through more local calibration to a child's knowledge. Caregivers have been shown to provide more language about referents that are unknown to their child, and adapt their language in-the-moment to the knowledge their child displays during a referential communication game (Leung et al., 2021). The calibration of caregivers' production to the child's learning and knowledge suggests a co-evolution such that these processes should not be considered in isolation.

What then gives rise to the structure in early language input that mirrors children's 93 learning mechanisms? Because of widespread agreement that caregivers' speech is not usually motivated by explicit pedagogical goals (Newport et al., 1977), the calibration of speech to learning mechanisms seems a happy accident; caregivers' speech just happens to be calibrated to children's learning needs. If caregivers' speech was pedagogically-motivated, extant formal frameworks of teaching could be used to derive predictions and expectations (e.g., Shafto et al., 2014). This framework-in which caregivers choose what they say in order to maximize their children's learning-has seen some success 100 in explaining children's developing phoneme discrimination (Eaves Jr et al., 2016). Even 101 here, however, the statistical structure of phonemes in caregivers' input also contains 102 features that are not well explained by a pedagogical goal (McMurray et al., 2013). 103

Instead, the recent outpouring of work exploring optimal communication (the Rational Speech Act model, see Frank & Goodman, 2012) provides a different framework for understanding caregivers' production. Under optimal communication accounts, speakers

and listeners engage in recursive reasoning to produce and interpret speech cues by making 107 inferences over one another's intentions (Frank & Goodman, 2012). These accounts have 108 made room for advances in our understanding of a range of language phenomena previously 109 uncaptured by formal modeling, most notably a range of pragmatic inferences (e.g., Frank 110 & Goodman, 2012; Bohn & Frank, 2019; Goodman & Frank, 2016). In this work, we 111 consider the communicative structure that emerges from an optimal communication system 112 across a series of interactions where one partner has immature linguistic knowledge. This 113 perspective offers the first steps toward a unifying account of both the child's learning and 114 the caregiver's production: Both are driven by a pressure to communicate successfully. 115

Early, influential functionalist accounts of language learning focused on the 116 importance of communicative goals (e.g., Brown, 1977). Recent work has demonstrated 117 that many structural aspects of natural language may have arisen from pressure to 118 communicate efficiently (Gibson et al., 2019). Our goal in this work is to formalize the 119 intuitions in these accounts in a computational model of language input, and to test this 120 model against experimental data. We take as the caregiver's goal the desire to 121 communicate with the child, not about language itself, but instead about the world in front 122 of them. To succeed, the caregiver must produce the kinds of communicative signals that the child can understand and respond contingently, potentially leading caregivers to tune the complexity of their speech as a byproduct of this in-the-moment pressure to communicate successfully (Yurovsky, 2018). 126

To examine this hypothesis, we draw on evidence from naturalistic data, a reference game experiment, a formal model, and learning simulations. We focus on ostensive labeling (i.e. using both gesture and speech in the same referential expression) as a case-study phenomenon of information-rich structure in the language learning environment. We first analyze naturalistic caregiver communicative behavior in a longitudinal corpus of parent-child interaction in the home (Goldin-Meadow et al., 2014). We investigate the extent to which caregivers tune their ostensive labeling across their child's development to

align to their child's developing linguistic knowledge (Yurovsky et al., 2016).

We then experimentally induce this form of structured language input in a simple 135 model system: an iterated reference game in which two players earn points for 136 communicating successfully with each other. Modeled after our corpus data, participants 137 are asked to make choices about which communicative strategy to use (akin to modality 138 choice). In an experiment on Mechanical Turk using this model system, we show that 139 pedagogically-supportive input can arise from a pressure to communicate. We then show 140 that participants' behavior in our game conforms to a model of communication as rational 141 planning: People seek to maximize their communicative success while minimizing their 142 communicative cost over expected future interactions. Finally, we demonstrate potential 143 benefits for the learner through a series of simulations to show that communicative 144 pressure on caregivers' speech facilitates learning. Under a variety of parameter settings, simple learners interacting with communicative partners outperform more complex 146 statistical learners.

In sum, our goal in this work is to argue that the fundamental unit of analysis for 148 understanding children's language learning is not the child alone, but rather the 149 caregiver-child dyad. To provide converging evidence for this claim, we use apply a 150 multi-method approach to a case study of one pedagogically-supportive behavior. We use 151 observational data to understand the contexts in which ostensive labeling happens, and 152 then experimentally manipulate these contexts in a model system in order to demonstrate that the observed relationships are causal. We develop a formal model that quantitatively accounts for this experimental data, and then simulate the long-term learning outcomes of 155 this model for children's language learning. Together, these different methods show that 156 rapid language learning can emerge from even highly-constrained child learners working 157 together with communicative-motivated caregivers. 158

Corpus Analysis

We first investigate parents' use of ostensive labeling in referential communication 160 in a longitudinal corpus of parent-child interaction. Ostensive labeling—the behavior 161 pointing to and labeling an object with its name-is a powerful source of information for 162 word learning because it reduces the ambiguity about the possible meanings of the word 163 the child is hearing. The word could of course still have many possible meanings—it could 164 refer to a part of the object, or something about it's state for instance, but it is unlikely to 165 refer to one of the other objects in the room (Quine, 1960). Prior work has shown that 166 parents tend to use ostensive labeling when referring to objects with basic level words (e.g. "chair", "dog", Callanan, 1985, @ninio1980), and that ostensive labeling is a particularly powerful source of data for young children (e.g., Baldwin, 2000; Gogate et al., 2000). We 169 take the ostensive labeling with multi-modal cues to be a case-study phenomenon of 170 pedagogically supportive language input. While our account should hold for other language 171 phenomena, by focusing on one phenomenon we attempt to specify the dynamics involved 172 in the production of such input. 173

In this analysis of naturalistic communication, we examine the prevalence of ostensive labeling in children's language environment at different ages. We find that this pedagogically-supportive form of input shows a key hallmark of adaptive tuning: caregivers using this information-rich cue more for young children and infrequent objects. Thus, parents' production of ostensive labeling is tuned to children's developing linguistic knowledge (Yurovsky et al., 2016).

$^{_{ m 80}}$ Methods

174

175

177

178

179

159

We used data from the Language Development Project—a large-scale, longitudinal corpus of naturalistic parent child-interaction in the home (Goldin-Meadow et al., 2014).

The Language Development Project corpus contains transcription of all speech and

communicative gestures produced by children and their caregivers over the course of the
90-minute home recordings. We coded each of these communicative instances to identify
each time a concrete noun was referenced using speech, gesture, or both in the same
referential expression (so called ostensive labeling). In these analyses, we focus on
caregivers' productions of ostensive labeling in the form of a multi-modal reference.

189 Participants

The Language Development Project aimed to recruit a sample of families who are representative of the Chicago community in socio-economic and racial diversity (Goldin-Meadow et al., 2014). These data are drawn from a subsample of 10 families from the larger corpus. Our subsample contains data taken in the home every 4-months from when the child was 14-months-old until they were 34-months-old, resulting in 6 timepoints (missing one family at the 30-month timepoint). Recordings were 90 minute sessions, and participants were given no instructions.

Of the ten target children, five were girls, three were Black and two were
Mixed-Race. Families spanned a broad range of incomes, with two families earning \$15,000
to \$34,999 and 1 family earning greater than \$100,000. The median family income was
\$50,000 to \$74,999.

o Procedure

From the extant transcription and gesture coding, we specifically coded all concrete noun referents produced in either the spoken or gestural modality (or both). Spoken reference was coded only when a specific noun form was used (e.g., "ball"), to exclude pronouns and anaphoric usages (e.g., "it"). Gesture reference was coded only for deictic gestures (e.g., pointing to or holding an object up for view) to minimize ambiguity in determining the intended referent. In order to fairly compare rates of communication across modalities, we need to examine concepts that can be referred to in either gesture or

speech (or both) with similar ease. Because abstract entities are difficult to gesture about using deictic gestures, we coded only references to concrete nouns.

211 Reliability

To establish the reliability of the referent coding, 25% of the transcripts were double-coded. Inter-rater reliability was sufficiently high (Cohen's $\kappa = 0.76$).

Disagreements in coding decisions were discussed and resolved by hand.

To ensure that our each referent could potentially be referred to in gesture or 215 speech, we focused on concrete nouns. We further wanted to ensure that the referents were 216 physically present in the scene (and thus available for deictic gestures). Using the 217 transcripts, a human rater judged whether the referent was likely to be present, primarily 218 relying on discourse context (e.g., a referent was coded as present if the deictic gesture is 219 used or used at another timepoint for the reference, or if the utterance included 220 demonstratives such as "This is an X"). A full description of the coding criteria can be 221 found in the Supporting Materials. 222

To ensure our transcript-based coding of referent presence was sufficiently accurate, a subset of the transcripts (5%) were directly compared to corresponding video data observation. Reliability across the video data and the transcript coding was sufficiently high ($\kappa = 0.72$). Based on transcript coding of all the referential communication about concrete nouns, 90% of the references were judged to be about referents that were likely present. All references are available in the open access dataset for further analysis.

229 Results

Corpus data were analyzed using a mixed effects regression to predict parents' use
of ostensive labeling for a given referent. The model included fixed effects of age in months,
frequency of the referent, and the interaction between the two. The model included a

random intercept and random slope of frequency by subject and a random intercept for 233 each unique referent. Frequency and age were both log-scaled and then centered both 234 because age and frequency tend to have log-linear effects and to help with model 235 convergence. The model showed that parents use ostensive labeling less with older children 236 $(\beta = -0.84, p < .001)$ and less for more frequent referents $(\beta = -0.09, p = .045)$. In 237 addition, the interaction between the two was significant, indicating that for parents 238 ostensively label more for younger children when referents are infrequent ($\beta = 0.18$, p =239 .001). Thus, in these data, we evidence that parents provide more pedagogically-supportive 240 input about rarer things in the world for their younger children (Figure 1). 241

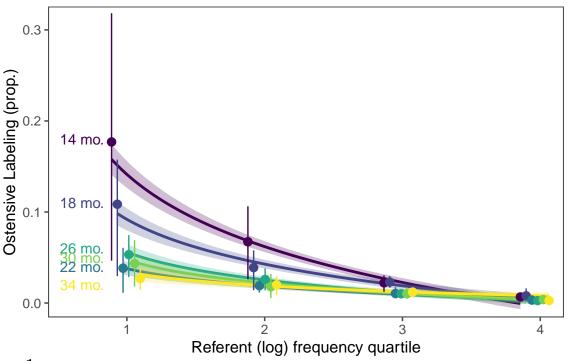


Figure 1

Parents' rate of ostensive labelling via multi-modal reference for referents of varying frequency. Log frequences were binned into quartiles for visualization. Points show empirical frequencies along with 95% confidence intervals computed by non-parametric bootstrapping. Lines show estimates from a logisitic regression along with ribbons indicating one standard error. Parents used ostensive labeling more for younger children and infrequent referents.

Discussion

Caregivers are not indiscriminate in their use of ostensive labeling; in these data,
they provided more of this support when their child was younger and when discussing less
familiar objects. These longitudinal corpus findings are consistent with an account of
caregiver alignment: parents are sensitive to their child's linguistic knowledge and adjust
their communication accordingly (Yurovsky et al., 2016).

This caregiver alignment is straightforwardly predicted from a dyadic perspective 248 that privileges communicative goals, yet wholly missing from learner-centric accounts of language learning (Yurovsky, 2018). While some early accounts of language learning 250 suggested a important role for caregiver tuning of linguistic input (e.g., Snow, 1972), these 251 accounts fell out of favor in light of influential investigations that did not find tuning of 252 syntactic structure (Newport et al., 1977). However, this work, along with other recent 253 papers, shows that caregivers may be well calibrated to their children's semantic knowledge 254 and this tuning may be quite relevant for language problems like word learning where 255 communicative success is central (e.g., Yurovsky et al., 2016, @leung2021). 256

While language input that is tuned to the child's linguistic competence could undoubtedly aid in language learning, the presence of such input does not necessarily imply pedagogical goals. These data could potentially be explained by a simpler, less altruistic goal: to communicate successfully. Because we do not have access to parents' goals, we these data alone cannot distinguish between these competing accounts not can they determine whether word frequency and children's linguistic knowledge are causally related to the use of ostensive labeling. To demonstrate that communicative pressure alone is sufficient to explain these patterns in ostensive labeling, we next developed an experimental paradigm to study these and other factors in a controllable model system.

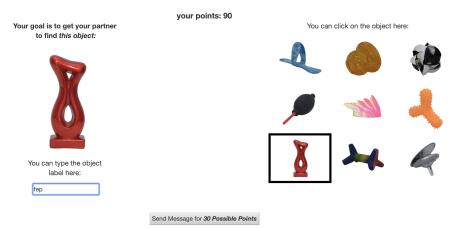


Figure 2
Screenshot showing the participant view during gameplay.

266 Experiment

To study the emergence of pedagogically supportive input from communicative pressure, we developed a simple reference game in which participants would be motivated to communicate successfully. After giving people varying amounts of training on novel names for nine novel objects, we asked them to play a communicative game in which they were given one of the objects as their referential goal, and they were rewarded if their partner successfully selected this referent from among the set of competitors (Figure 2). Although we told participants that their partner would be played by another human, all partners were actually pre-programmed bots.

Participants could choose to refer either using the novel labels they had been exposed to, or they could use a deictic gesture (i.e. point) to indicate the referent to their partner. The point was unambiguous, and thus would always succeed. However, in order for language to be effective, the participant and their partner would have to know the correct novel label for the referent.

In choosing a communicative signal, participants should try to maximize their likelihood of success while minimizing their communicative cost (see e.g. Frank &

Goodman, 2012). This cost should be related to the energy required to produce the signal.

Though it need not be the case for all gestures and contexts, our task compares simple

lexical labeling and unambiguous deictic gestures, which are likely are slower and more

effortful to produce (e.g., see Yurovsky et al., 2018). We thus assumed that pointing should

be more costly than speaking. Nonetheless, because we do not have a way of estimating

the costs of pointing and speaking from the observational data, we manipulated them

experimentally across conditions to understand how they impact peoples' behavior.

Critically, participants were told that they would play this game repeatedly with their partner. In these repeated interactions, participants are able to learn about an interlocutor and potentially influence their learning. Thus, there is a third type of signal participants could send: using both pointing and speech within a single trial to effectively teach the listener through ostensive labeling. This strategy necessitates making inferences about their partner's knowledge state, so we induced knowledge asymmetries between the participant and their partner. To do so, we manipulated how much training they thought their partner had received.

Our communicative game was designed to reward in-the-moment communication,
and thus teaching required the participant pay a high cost upfront. However, rational
communicators may understand that if one is accounting for future trials, paying the cost
upfront to teach their partner allows them to use a less costly message strategy on
subsequent trials (namely, speech). Manipulating the partner's knowledge and the utility of
communicative strategies, we aimed to experimentally determine the circumstances under
which richly-structured input emerges, without an explicit pedagogical goal.

While our reference game setting has limited ecological validity, this setup allows us
to explicitly manipulate the crucial features of the communicative setting (e.g.,
communicative cost, strategy, and partner knowledge). In this controlled task, we can look
for the emergence of structure that parallels the naturalistic input described in our corpus

evidence, while also experimentally testing for possible drivers of such structure. This
experimental setup further allows us to straightforwardly test and compare key predictions
using a formal model to explain participant behavior.

Method

In this experiment, participants were recruited to play our reference game via
Amazon Mechanical Turk, an online platform that allows workers to complete surveys and
short tasks for payment. In this study, all participants were placed in the role of speaker
and listener responses were programmed.

316 Transparency and Openness

This sample size, experimental design, and analysis plan were pre-registered at https://osf.io/63qdg. Sample size was determined based on prior pilot experiments. All data, analysis code, and research materials are available at https://osf.io/d9gkw/. Data were analyzed using R version 4.1.1 (R Core Team, 2021). Models were estimated using version 1.1-27.1 of the lme4 package (Bates et al., 2015).

Participants

480 participants were recruited though Amazon Mechanical Turk and received \$1 for their participation. Data from 51 participants were excluded from subsequent analysis for failing the critical manipulation check (accurately reporting their partners knowledge prior to gameplay) and a further 28 for producing pseudo-English labels (e.g., "pricklyyone"). The analyses reported here exclude the data from those participants, but all analyses were also conducted without excluding any participants and all patterns hold (ps < 0.05).

$_{29}$ $Design\ and\ Procedure$

Participants were told they would be introduced to novel object-label pairs and then
asked to play a communication game with a partner wherein they would have to refer to a
particular target object. Participants were exposed to nine novel objects, each with a
randomly assigned pseudo-word label. We manipulated the exposure rate within-subjects:
during training participants saw three of the nine object-label mappings four times, three
of them two times, and three of them just one time, yielding a total of 21 training trials.

Participants were then given a simple recall task to establish their knowledge of the novel
lexicon (pretest).

During gameplay, participants saw the target object in addition to an array of all
nine objects. Participants had the option of either directly selecting the target object from
the array (pointing)—a higher cost, but unambiguous cue—or typing a label for the object
(speech)—a lower cost cue contingent on their partner's knowledge. After sending the
message, participants were shown which object their partner selected.

We also manipulated participants' expectations about their partner's knowledge to
explore the role of knowledge asymmetries. Prior to beginning the game, participants were
told how much exposure their partner had to the lexicon. Across three between-subjects
conditions, participants were told that their partner had either no experience with the
lexicon, had the same experience as them, or had twice their experience. As a
manipulation check, participants were then asked to report their partner's level of
exposure, and were corrected if they answered incorrectly. Participants were then told that
they would be asked to refer to each object three times during the game.

Partners were programmed with starting knowledge states initialized according to
the partner knowledge condition. Partners with no exposure began the game with
knowledge of 0 object-label pairs. Partners with the same exposure as the participant
began with knowledge of five object-label pairs (three high-frequency, on mid-frequency,

one low-frequency), based on the learning we observed from participants in a pilot
experiment. Lastly, partners with twice as much exposure as the participant began with
knowledge of all nine object-label pairs.

To simulate knowledgeable behavior, when the participant typed an object label, 358 the partner was programmed to consult their own knowledge. Messages were evaluated by 359 taking the Levenshtein distance (LD) between the typed label and each possible label in the 360 partner's vocabulary. Partners then selected the candidate with the smallest edit distance 361 (e.g., if a participant typed the message "tomi", the programmed partner would select the 362 referent corresponding to "toma", provided toma was found in its vocabulary). If the 363 participant's message was more than two edits away from all of the words in the partner's 364 vocabulary, the partner selected an object whose label they did not know. If the participant 365 clicked on an object (pointing), the partner was programmed to always select that referent. 366

Participants could win up to 100 points per trial if their partner correctly selected 367 the target referent based on their message. If the partner failed to identify the target 368 object, participants received no points. We manipulated the relative utility of the speech 369 cue between subjects across two conditions: Higher Speech Efficiency and Lower Speech 370 Efficiency. In the Higher Speech Efficiency condition, participants received 30 points for 371 gesturing and 100 points for labeling, and thus speech had very little cost relative to pointing and participants should be highly incentivized to speak. In the Lower Speech Efficiency condition, participants received 50 points for gesturing and 80 points for labeling, and thus gesturing is still costly relative to speech, but the difference between 375 them is smaller lowering the incentivize to speak. 376

Participants were told about a third type of possible message: using both pointing
and speech within a single trial to effectively teach their partner an object-label mapping.
This action directly mirrors the ostensive labeling behavior parents produced in the corpus
data—it yields an information-rich, pedagogically-supportive learning moment. In order to

produce this teaching behavior, participants had to pay the cost of producing both cues
(i.e. both pointing and speech). Note that, in all utility conditions, teaching yielded
participants 30 points (compared with the much more beneficial strategy of speaking which
yielded 100 points or 80 points across our two utility manipulations). Partners were
programmed to integrate new taught words into their knowledge of the lexicon, and check
those taught labels on subsequent trials when evaluating participants' messages.

Crossing our 2 between-subjects manipulations yielded 6 conditions (2 utility 387 manipulations: Higher Speech Efficiency and Lower Speech Efficiency; and 3 levels of 388 partner's exposure: None, Same, Twice), with 80 participants in each condition. We 389 expected to find results that mirrored our corpus findings such that rates of teaching would 390 be higher when there was an asymmetry in knowledge where the participant knew more 391 (None manipulation) compared with when there was equal knowledge (Same manipulation) 392 or when the partner was more familiar with the language (Twice manipulation). We 393 expected that participants would also be sensitive to our utility manipulation, such that 394 rates of labeling and teaching would be higher in the Higher Speech Efficiency conditions than the other conditions.

Results

In each trial, participants could choose one of 3 communicative strategies: pointing, speech, or teaching. We expected participants to flexibly use communicative strategies in response to their relative utilities, their partner's knowledge of the lexicon, and participants' own lexical knowledge. To test our predictions about each communicative behavior (pointing, speech, and teaching), we fit separate logistic mixed effects models for each behavior, reported below. It should be noted that these three behaviors are mutually exhaustive. First, we report how well participants learned our novel lexicon during training.

Learning

As an initial check of our exposure manipulation, we first fit a logistic regression 406 predicting accuracy at test from a fixed effect of exposure rate and random intercepts and 407 slopes of exposure rate by participant as well as random intercepts by item. We found a 408 reliable effect of exposure rate, indicating that participants were better able to learn items 409 that appeared more frequently in training ($\beta = 1.08$, p < .001, see Figure 3). On average, 410 participants knew approximately 6 of the 9 words in the lexicon (M(sd) = 6.28 (2.26)). An 411 analysis of variance confirmed that learning did not differ systematically across participants 412 by partner's exposure, utility manipulation, or their interaction (ps > 0.05). 413

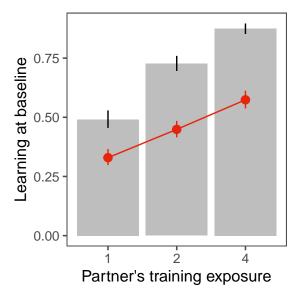


Figure 3

Participants' performance on the baseline recall task for the lexicon, as function of amount of exposure during training (grey bars). The red line shows the proposition of trials during gameplay in which participants used the learned labels, excluding teaching behaviors. Error bars show 95% confidence intervals computed by non-parametric bootstrapping.

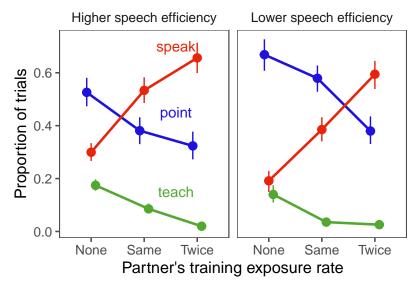


Figure 4

Participants' communicative method choice as a function of exposure and the utility manipulation. Error bars indicate 95% confidence intervals computed by non-parameteric bootstrapping

Pointing

424

425

426

When should we expect participants to rely on pointing? Pointing has the highest 415 utility for words you failed to learn during training, words you think your partner is 416 unlikely to know (i.e., for lower partner knowledge conditions), and when the utility scheme 417 is relatively biased toward pointing (i.e., the Lower Speech Efficiency condition). To test 418 these predictions, we ran a mixed effects logistic regression to predict whether participants 419 chose to point during a given trial as a function of the target object's exposure rate during 420 training, object instance in the game (first, second, or third), utility manipulation, and 421 partner manipulation. Random effects terms for subject and object were included in the 422 model. 423

Consistent with our predictions, exposure rate during training was a significant negative predictor of pointing during the game, such that participants were less likely to rely on pointing for well trained (and thus well learned) objects ($\beta = -0.50$, p < .001).

Additionally, participants were significantly more likely to point in the Lower Speech 427 Efficiency condition where pointing is relatively less costly, compared with the Higher 428 Speech Efficiency condition ($\beta = 1.20, p < .001$; see Figure 4). We also found a significant 429 negative effect of partner's knowledge, such that participants pointed more for partners 430 with less knowledge of the lexicon ($\beta = -0.81$, p < .001). 431

Note that these effects cannot be explained by solely participants' knowledge; all 432 patterns above hold when looking only at words known by the participant at pretest (ps <433 0.01). Further, these patterns mirror previous corpus analyses demonstrating parents' use 434 of pointing in naturalistic parental communicative behaviors, and parents likely have 435 lexical knowledge of even the least frequent referent (see Yurovsky et al., 2018). 436

Speech437

438

452

When should we expect participants to use speech? Speech has the highest utility for words you learned during training, words you think your partner is likely to know (i.e., 439 for higher partner knowledge conditions), and when utility scheme is relatively biased 440 toward speech (i.e., the Higher Speech Efficiency condition). To test these predictions, we 441 ran a mixed effects logistic regression to predict whether participants chose to speak during 442 a given trial as a function of the target object's exposure rate during training, object 443 instance in the game (first, second, or third), utility manipulation, and partner 444 manipulation. Random effects terms for subjects and object were included in the model. Consistent with our predictions, speech seemed to largely trade off with gesture. 446 Exposure rate during training was a significant positive predictor of speaking during the game, such that participants were more likely to use speech for well trained (and thus well learned) objects ($\beta = 0.35, p < .001$). Additionally, participants were significantly less likely to speak in the Lower Speech Efficiency condition where speech is relatively more 450 costly, compared with the Higher Speech Efficiency condition ($\beta = -0.87$, p = .001). We 451 also found a significant positive effect of partner's knowledge, such that participants used

speech more for partners with more knowledge of the lexicon ($\beta = 1.95$, p < .001). Unlike for gesture, there was a significant effect of object instance in the game (i.e., first, second, or third trial with this target object) on the rate of speaking, such that later trials were more likely to elicit speech ($\beta = 0.72$, p < .001). This effect of order likely stems from a trade-off with the effects we see in teaching (described below); after a participant teaches a word on the first or second trial, the utility of speech is much higher on subsequent trials.

459 Emergence of Teaching.

Thus far, we have focused on relatively straightforward scenarios to demonstrate
that a pressure to communicate successfully in the moment can lead participants to trade
off between gesture and speech sensibly. Next, we turn to the emergence of teaching
behavior.

When should we expect participants to teach? Teaching has the highest utility for 464 words you learned during training, words you think your partner is unlikely to know (i.e., 465 for lower partner knowledge conditions), and when utility scheme is relatively biased 466 toward speech (i.e., the Higher Speech Efficiency condition). In this utility scheme, it is 467 more valuable to pay the cost of teaching early because of the increased benefit of using 468 speech later. To test these predictions, we ran a mixed effects logistic regression to predict 469 whether participants chose to teach during a given trial as a function of the target object's 470 exposure rate during training, object instance in the game (first, second, or third), utility 471 manipulation, and partner manipulation. Random effects terms for subjects and object 472 were included in the model.

Consistent with our predictions, rates of teaching were higher for more highly trained words, less knowledgeable partners, and when speech had the highest utility. Exposure rate during training was a significant positive predictor of teaching during the game, such that participants were more likely to teach for well trained (and thus well learned) objects ($\beta = 0.14$, p = .001). While costly in the moment, teaching can be a

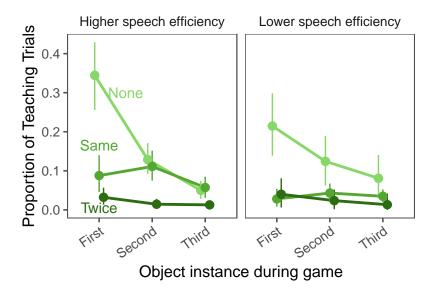


Figure 5

Rates of teaching across the six utility and partner knowledge conditions as a function of how many times the current target referent object had previously been the target. Error bars show 95% confidence intervals computed by non-parametric bootstrapping.

beneficial strategy in our reference game because it subsequently allows for lower cost 479 strategy (i.e. speaking), thus when speaking has a lower cost, participants should be more 480 incentivized to teach. Indeed, participants were significantly less likely to teach in the 481 Lower Speech Efficiency condition where speech is relatively more costly, compared with 482 the Higher Speech Efficiency condition ($\beta = -0.96$, p = .001). We also found a significant 483 negative effect of partner's knowledge, such that participants taught more with partners 484 that had less knowledge of the lexicon ($\beta = -2.23$, p < .001). There was also a significant 485 effect of object instance in the game (i.e., whether this is the first, second, or third trial 486 with this target object) on the rate of teaching. The planned utility of teaching comes from using another, cheaper strategy (speech) on later trials, thus the expected utility of 488 teaching should decrease when there are fewer subsequent trials for that object, predicting that teaching rates should drop dramatically across trials for a given object. Participants 490 were significantly less likely to teach on the later appearances of the target object ($\beta =$ 491 -1.09, p < .001). 492

93 Discussion

502

503

504

505

506

517

As predicted, the results of this experiment corroborate our findings from the corpus analysis, demonstrating that pedagogically supportive behavior emerges despite the initial cost when there is an asymmetry in knowledge and when speech is less costly than other modes of communication. While this paradigm has stripped away much of the interactive environment of the naturalistic corpus data, it provides important proof of concept that the structured and tuned language input we see in those data could arise from a pressure to communicate. The paradigm's clear, quantitative trends also allow us to build a formal model to predict our empirical results.

The results from this experiment are qualitatively consistent with a model in which participants make their communicative choices to maximize their expected utility from the reference game. We next formalize this model to determine if these results are predicted quantitatively as well.

Model: Communication as planning

In order to model when people should speak, point, or teach, we begin from the 507 problem of what goal people are trying to solve (Marr, 1982). Following a long history of 508 work in philosophy of language, we take the goal of communication to be causing an action 509 in the world by transmitting some piece of information to one's conversational partner 510 (e.g., Wittgenstein, 1953; Austin, 1975). If people are near-optimal communicators, they 511 should choose communicative signals that maximize the probability of being understood 512 while minimizing the cost of producing the signal (Clark, 1996; Grice, 1975). In the special 513 case of reference, solving this problem amounts to producing the least costly signal that 514 correctly specifies one's intended target referent in such a way that one's conversational 515 partner can select it from the set of alternative referents. 516

Recently, Frank and Goodman (2012) developed the Rational Speech Act

framework—a formal instantiation of these ideas. In this model, speakers choose from a set of potential referential expressions in accordance to a utility function that maximizes the 519 probability that a listener will correctly infer their intended meaning while minimizing the 520 number of words produced. This framework has been successfully applied to a variety of 521 linguistic phenomena such as scalar implicature, conventional pact formation, and 522 production and interpretation of hyperbole (Goodman & Frank, 2016; see also related work 523 from Franke, 2013). These models leverage recursive reasoning-speakers reasoning about 524 listeners who are reasoning about speakers—in order to capture cases in which the literal 525 meaning and the intended meaning of sentences diverge. 526

To date, this framework has been applied primarily in cases where both 527 communicative partners share the same linguistic repertoire, and thus communicators know 528 their probability of communicating successfully having chosen a particular signal. This is a 520 reasonable assumption for pairs of adults in contexts with shared common ground. But 530 what if partners do not share the same linguistic repertoire, and in fact do not know the 531 places where their knowledge diverges? In this case, communicators must solve two 532 problems jointly: (1) Figure out what their communicative partner knows, and (2) produce 533 the best communicative signal they can given their estimates of their partner's knowledge. If communicative partners interact repeatedly, these problems become deeply intertwined: 535 Communicators can learn about each-other's knowledge by observing whether their attempts to communicate succeed. For instance, if a communicator produces a word that 537 they believe identifies their intended referent, but their partner fails to select that referent, 538 the communicator can infer that their partner must not share their understanding of that 539 word. They might then choose not to use language to refer to this object in the future, but 540 choose to point to it instead. 541

Critically, communicators can also change each-other's knowledge. When a
communicator both points to an object and produces a linguistic label, they are in effect
teaching their partner the word that they use to refer to this object. While this behavior is

costly in the moment, and no more referentially effective than pointing alone, it can lead to more efficient communication in the future-instead of pointing to this referent forever 546 more, communicators can now use the linguistic label they both know that they share. 547 This behavior naturally emerges from a conception of communication as planning: 548 Communicators' goal is to choose a communicative signal today that will lead to efficient 549 communication not just in the present moment, but in future communications as well. If 550 they are likely to need to refer to this object, it is worth it to be inefficient in this one 551 exchange in order to be more efficient future. In this way, pedagogically supportive 552 behavior can emerge naturally from a model with no separate pedagogical goal. In the 553 following section, we present a formal instantiation of this intuitive description of 554 communication as planning and show that it accounts for the behavior we observed in our 555 experiments.

Alternatively, pedagogically-supportive input could emerge from an explicit 557 pedagogical goal. Shafto et al. (2014) have developed an framework of rational pedagogy 558 built on the same recursive reasoning principles as in the Rational Speech Act Framework: 550 Teachers aim to teach a concept by choosing a set of examples that would maximize 560 learning for students who reason about the teachers choices as attempting to maximize 561 their learning. Rafferty et al. (2016) et al. expanded this framework to sequential teaching, in which teachers use students in order to infer what they have learned and choose the 563 subsequent example. In this case, teaching can be seen as a kind of planning where 564 teachers should choose a series of examples that will maximize students learning but can 565 change plans if an example they thought would be too hard turns out too easy-or vice-versa. In the case of our reference game, this model is indistinguishable from a communicator who seeks to maximize communicative success but is indifferent to 568 communicative cost. A cost-indifferent model makes poor predictions about caregivers' 569 behavior in our corpus, and also adults' behavior in our experiments, but we return to it in 570 the subsequent section to consider how differences in caregivers' goals and differences in

children's learning contribute to changes in the rate of language acquisition.

73 Formal Model

596

We take as inspiration the idea that communication is a kind of action—e.g., talking 574 is a speech act (Austin, 1975). Consequently, we can understand the choice of which 575 communicative act a speaker should take as a question of which act would maximize their 576 utility: achieving successful communication while minimizing their cost (Frank & Goodman, 2012). In this game, speakers can take three actions: talking, pointing, or teaching. The Utilities (U) are given directly by the rules of this game. Because communication is a repeated game, people should take actions that maximize their 580 Expected Utility (EU) not just for the current round, but for all future communicative 581 acts with the same conversational partner. We can think of communication, then as a case 582 of recursive planning. However, people do not have perfect knowledge of each-other's 583 vocabularies (v). Instead, they only have uncertain beliefs (b) about these vocabularies 584 that combine their expectations about what kinds of words people with as much linguistic 585 experience as their partner are likely to know with their observations of their partner's 586 behavior in past communicative interactions. This makes communication a kind of 587 planning under uncertainty well modeled as a Partially Observable Markov Decision 588 Process (POMDP, Kaelbling et al., 1998). 589

Optimal planning in a Partially Observable Markov Decision Process involves a cycle of three phases: (1) Plan, (2) Act, and (3) Update beliefs. On each trial of the referential game, the model first makes a plan–reasoning about which action it should take on this trial and on subsequent trials to come. To help build intuition for the model, we first describe how it might operate in a series of example trials. We then give a formal description of the model's learning and behavior.

At the start of each trial, the model makes a plan consider the actions it will take

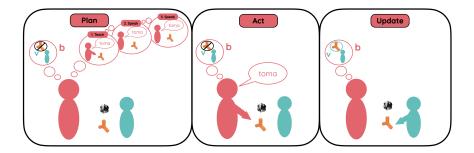


Figure 6

A schematic of the plan-act-update cycle of communication as a partially observable Markov decision process (POMDP). The red figure corresponds to caregivers in the world, or participants in the experiment. The blue figure corresponds to children or experimental partners. The caregiver is using their beliefs (b) about their child's true vocabulary (v) to plan their next communicative act. Teaching by ostensive labeling is costly in the moment, but has utility because it enables a plan to speak in the future.

on this trial and future trials. If the model speaks-producing the label for the target 597 object, two outcomes are possible. First, its partner could select the correct referent. This 598 would give maximal utility on the current trial, and also cause the model to know that its 590 partner knows the right label for this object. In that case, the model could speak again on 600 future trials and expect to succeed and be rewarded. On the other hand, its partner could 601 select the wrong referent. This could happen either because the model itself does not know 602 the right label, or because its partner does not. In that case, the model would get low 603 utility on this round, and would likely teach or point for this referent on subsequent trials. 604 Alternatively the model could point. This would give some utility on the current trial, and 605 would leave the model in the same state of uncertainty about whether its partner knows the correct label for the referent on subsequent trials. Finally the model could teach. This would lead to very little utility on the current trial, but would cause the model to know that it can speak to refer to this referent on future trials (as in Figure 6). Reasoning 609 forward about however many trials are left to play for this referent, the model makes a 610 plan with the appropriate number of steps. At the start of the game, the model knows it 611

will play three times for each referent so it might make a plan like {speak, speak, speak}

After formulating this plan, the model will take the first action in the plan sequence

(e.g. speak). It will then observe its partner's behavior. In this case, suppose that its

partner selects the incorrect referent. The model will then update its beliefs—its partner

must not know the correct label for the target object. The next time it needs to

communicate about the same object, it will be very unlikely to plan to speak, even though

its previous plan was to do so. This is because the model's belief about the world has

changed and now it will be more likely to {point, point} or to {teach, speak}.

We next formally specify each step in the cycle and finally define how people form initial beliefs about their partner's language. All code for implementing the model is available on the Open Science Foundation project page associated with this paper.

Plan

620

621

622

When people plan, they compute the expected utility of each possible action (a) by combining the expected utility of that action now with the Discounted Expected Utility they will get in all future actions. The amount of discounting (γ) reflects how much people care about success now compared to success in the future. Because utilities depend on the communicative partner's vocabulary, people should integrate over all possible vocabularies in proportion to the probability that their belief assigns to that vocabulary $(\mathbb{E}_{v\sim b})$.

$$EU\left[a|\,b\right] = \mathbb{E}_{v \sim b}\left(U(a|v) + \gamma \,\mathbb{E}_{v',o',a'}\,\left(EU\left[a'|b'\right]\right)\right)$$

630 Act

Next, people take an action as a function of its expected utility. Following other models in the Rational Speech Act framework, we use the Luce Choice Axiom, in which each choice is taken in probability proportional to its exponentiated utility (Frank & Goodman, 2012; Luce, 1959). This choice rule has a single parameter α that controls the

noise in this choice—as α approaches 0, choice is random and as α approaches infinity, choice is optimal.

$$P(a|b) \propto \alpha e^{EU[a|b]}$$

637 Update beliefs

After taking an action, people observe (o) their partner's choice—sometimes they 638 correctly select the intended object, and sometimes they do not. People then update their 639 beliefs about the partner's vocabulary based on this observation. For simplicity, we assume 640 that people think their partner should always select the correct target if they point to it, or 641 if they teach, and similarly should always select the correct target if they produce its label 642 and the label is in their partner's vocabulary. Otherwise, they assume that their partner 643 will select the wrong object. People could of course have more complex inferential rules, 644 e.g., assuming that if their partner does know a word they will choose among the set of objects whose labels they do not know (mutual exclusivity, Markman & Wachtel, 1988). 646 Empirically, however, our simple model appears to accord well with people's behavior.

$$b'(v') \propto P(o|v', a) \sum_{v \in V} P(v'|v, a) b(v)$$

The critical feature of a repeated communication game is that people can change 648 their partner's vocabulary. In teaching, people pay the cost of both talking and pointing 649 together, but can leverage their partner's new knowledge on future trials. Note here that 650 teaching has an upfront cost and the only benefit to be gained comes from using less costly communication modes later. There is no pedagogical goal—the model treats speakers as selfish agents aiming to maximize their own utilities by communicating successfully. We 653 assume for simplicity that teaching is always successful in this very short game, that 654 communicative partners do not forget words once they have learned them, and that no 655 learning happens by inference from mutual exclusivity. 656

$$P(v'|v,a) = \begin{cases} 1 & \text{if } v_w \in v\&v' \mid a = \text{point+talk} \\ 0 & \text{otherwise} \end{cases}$$

$_{657}$ $Initial \ Beliefs$

The final detail is to specify how people estimate their partner's learning rate (p)658 and initial vocabulary (v). We propose that people begin by estimating their own learning rate by reasoning about the words they learned at the start of the task: Their learning rate 660 (p) is the rate that maximizes the probability of them having learned their initial 661 vocabularies from the trials they observed. People can then expect their partner to have a 662 similar p (per the "like me" hypothesis, Meltzoff, 2005). Having an estimate of their 663 partner's p, they can estimate their vocabulary by simulating their learning from the 664 amount of prior exposure to language their partner had before the game. In our 665 experiments, we explicitly manipulated this expectation by telling participants how much 666 exposure their partner had relative to their own exposure. 667

668 Method

We implemented the planning model using WebPPL—a programming language
designed for specifying probabilistic models (Goodman & Stuhlmüller, 2014). We began
with the POMDP specification developed by Evans et al. (2017). To derive predictions
from the model, we exposed it to the same trial-by-trial stimuli as the participants in our
experiment, and used the probabilistic equations defined above to determine the likelihood
of choosing each behavior (i.e., speak, point, or teach) on every trial. Separate predictions
were made for each trial for each participant on the basis of all of the information available
to each participant at that point in time (e.g., how many words they had learned, their
partner's observed behavior previously, etc).

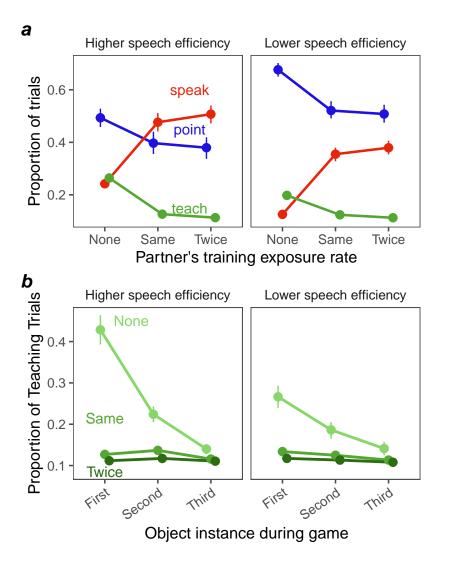
The model's behavior is contingent on two parameters-discounting (γ) , and its 678 rationality (α) . In order to determine the values of these parameters that best characterize 679 human participants, we used empirical Bayesian inference to estimate the posterior means 680 of both. Using posterior mean estimates rather than the maximum likelihood estimates 681 naturally penalizes models for their ability to predict patterns of data that were not 682 observed, applying a kind of Bayesian Occam's razor (MacKay, 1992). Because of we found 683 substantial variability in the best parameter estimates across individual participants, we 684 estimated parameters hierarchically, with group-level hyper-parameters forming the priors 685 for individual participants' parameters. This hierarchical estimation process achieves the 686 same partial pooling as as subject-level random effects in mixed-effects models, giving 687 estimates of the group-level parameters (Gelman & Hill, 2006). Details of the estimation 688 procedure can be found in the Supplemental Materials.

690 Model Results

In line with previous work on rational speech act models, and decision making, we expected rationality (α) to be around 1 or 2 (Frank & Goodman, 2012, 2014). We estimated the posterior mean rationality (α) to be 1.33 with a 95% credible interval of [1.24, 1.42]. We did not have strong expectations for the value of the discounting parameter (γ), but estimated it to be 0.42 [0.39, 0.44], suggesting that on average participants weighed the next occurrence of a referent as slightly less than half as important as the current occurrence.

To derive predictions from the model, we ran 100 simulations of the model's choices participant-by-participant and trial-by-trial using our posterior estimates of the hyper-parameters α and γ . Because we did not use our participant-level parameter estimates, this underestimates the correlations between model predictions and empirical data (as it ignores variability across participants). Instead, it reflects the model's best predictions about the results of a replication of our experiment, where individual

Figure 7



(a) Model prediction choice of communicative method choice as a function of exposure and the utility manipulation. (b) Model predicted probability of teaching by Partner's language knowledge and exposure rate.

participants' parameters will not be known apriori. Figure 7a shows the predictions from 704 the model in analogous format to the empirical data in Figure 4. The model correctly 705 captures the qualitative trends in participants' behavior: It speaks more and points less in 706 the Higher speech efficiency condition. Figure 7b shows the model's predicted teaching 707 behavior in detail in an analogous format to the empirical data in Figure 5. The model 708 again captures the qualitative trends apparent in participants' behavior. The model 709 teaches less knowledgeable partners, especially those who it believes have no language 710 knowledge at all. The model teaches more when speech is relatively more efficient, and 711 thus the future utility of teach a partner is higher. And finally the model teaches most on 712 the first occurrence of each object, and becomes less likely to teach on future occurrences 713 when (1) partners should be more likely to know object labels, and (2) the expected future 714 rewards of teaching are smaller. 715

To estimate the quantitative fit between model predictions and empirical data, we compute the Pearson correlation between the model's probability of using each action and participants' probability of using that same action as a function of appearance, condition, and partner's exposure. Across experimental manipulations, the model's predictions were highly correlated with participant behavior $(r = 0.92 \ [0.86, 0.95], t(52) = 16.67, p < .001;$ Figure 8).

Finally, we compare this model to two simpler alternative models: (1) A no-cost 722 model in which people are indifferent to the costs of communication, and (2) a myopic 723 model in which people do not plan for future interactions, and instead only care about the 724 utility of their communicative choices on the immediate communicative event. We 725 estimated parameters for these two simpler models using the same procedure as the full 726 model: We first fit individual participant-level parameters and then estimated the posterior mean parameters for the population of participants. To compare these reduced models to 728 our full model, we computed the log likelihood of observing the experimental data if 729 participants behaved according to each of the three models. These likelihoods combine 730

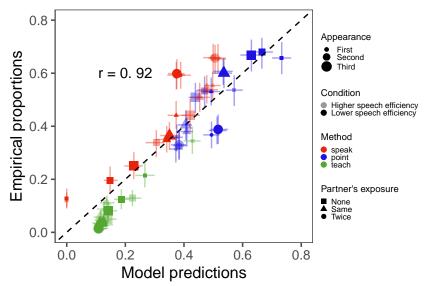


Figure 8

Fit between model predictions and empirical data.

both the probability of observing the empirical the data under the model, and the 731 probability of the model parameters under the model priors. This prior probability 732 implements a kind of Bayesian Occam's razor, penalizing the two models which involve 733 planning and thus fit a discounting parameter (full, no-cost) relative to the no-planning 734 model which has only a rationality parameter. For the full model, the average likelihood across 100 runs of the model was -15,771.93. By comparison, the likelihoods for the no-cost 736 model and myopic model were -21,016.96 and -17,257.91 respectively. Thus, the probability 737 of observing the empirical data was thousands of times more likely under the full model 738 than either of the simpler alternatives. 739

740 Discussion

In both qualitative and quantitative analyses, participants' behavior in our
communication task was well explained by a model of communication as rational planning
under uncertainty. The key intuition formalized by this model is that the value of a
communicative acts derives from (1) the immediate effect on resolving the current

754

755

756

757

758

759

communicative need, and (2) the potential benefit of the act for communication with this conversational partner in the future. Crucially, this model is able to predict a putatively 746 altruistic behavior-teaching by ostensive labeling-without any altruistic goals at all. 747 Because ostensive labeling can increase the efficiency of future communication, it can be 748 beneficial even under a purely self-interested utility function. What's more, the model 749 correctly predicts the circumstances under which participants will engage in teaching 750 behavior: early interactions with linguistically naive communicative partners in 751 circumstances where language is a relatively efficient communicative modality. 752

Importantly, this model does not rule out the possibility that participants in our 753 experiment—and more broadly people in the real world—may teach because of other more altruistic mechanisms or pressures. The model simply shows that appealing to such mechanisms is not necessary to explain the ostensive labeling observed in caregivers' conversations with their children, and by extension other behaviors that may at first blush appear to be pedagogically motivated. By the same logic, the model predicts that there should be other pedagogically supportive behaviors in the interactions between caregivers and their children, and likely in the interactions between any two communicative partners 760 who have some expectation that they will communicate again in the future. This framework thus provides a potential explanation for the occurrence of these behaviors and a framework for understanding their impact on language learning.

Of course, not all potentially pedagogically-supportive behaviors will yield an 764 immediate or future communicative benefit. For instance, correcting children's syntactic 765 errors could be helpful for their language development, but unless it resolves a communicative ambiguity, it will have little impact on communicative success. Our framework would predict that these behaviors should be rare, and indeed such behaviors appear to be generally absent in children's input (Marcus, 1993). We return this issue at 769 greater length in the General Discussion. Before turning to that, however, we first consider 770 the consequences of this model of communication for children's language. In the next 771

784

785

786

787

788

789

790

791

792

795

796

section, we use simulation methods to ask how caregivers' communicative motivation may impact their children's learning, and how this impact changes as a function of the complexity of the world and the efficacy of children's learning mechanisms.

Consequences for Learning

In the model and experiments above, we asked whether the pressure to communicate successfully with a linguistically-naive partner would lead to pedagogically supportive input. These results confirmed its sufficiency: As long as linguistic communication is less costly than deictic gesture, people should be motivated to teach in order to reduce future communicative costs. Further, the strength of this motivation is modulated by predictable factors (speakers' linguistic knowledge, listeners' linguistic knowledge, relative cost of speech and pointing, learning rate, etc.), this modulation is quantitatively well predicted by a rational model of planning under uncertainty about a partner's vocabulary.

In this final section, we take up the consequences of communicatively-motivated linguistic input for a child learning language. To do this, we adapt a framework used by Blythe et al. (2010) to estimate the learning times for an idealized child learning language under different models of the child and their careviver We derive estimates by simulating exposure to successive communicative events, and measuring the probability that successful learning happens after each event. The question of how different caregiver goals during caregiver-child impact children's learning can then be formalized as a question of how much more quickly learning happens under one simulation of child and caregiver model than another.

We consider children's word learning in three different simulated language environments:

1. Teaching - The caregiver's goal in each interaction is to maximize their child's learning (by teaching on every trial). This goal is equivalent to a model in which the

goal is to maximize communicative success without minimizing communicative cost.

In this case, we model the child's learning with the same simple binomial model we
proposed for participants in the experiment. After every teaching event, if the child
has not yet learned the label, they do so with a probability defined by their learning
rate.

- 2. Communication The caregiver's goal in each interaction with their child is to 802 maximize their communicative success while minimizing their communicative cost. 803 This the model described in the Model section above. In this case, we model the 804 child's learning with the same simple binomial model whenever the caregiver produces a teaching event. If the caregiver instead points or speaks, the child cannot 806 learn at all. We make this assumption for simplicity—a fairer simulation would also 807 include the child's learning from co-occurrence statistics of speaking events as in the 808 model above. However, simultaneously simulating the child's learning from 809 ambiguous input and the caregivers' reasoning about their child's likelihood of having 810 learned from ambiguous input is computationally challenging. This model thus 811 underestimates the rate at which children would learn from communication and 812 should be considered a lower bound. 813
- 3. Talking The caregiver's goal in each interaction is to refer to their intended referent 814 so that a knowledgeable listener would understand them, without accounting for the 815 child's language knowledge. This goal is equivalent to minimizing communicative cost 816 without maximizing communicative success. Because the caregiver never teaches 817 under this model, the child must learn from the co-occurrence statistics of the words 818 they hear and objects they see. That is, the child needs to solve the cross-situational 819 learning problem (Yu & Smith, 2007). In this case, the child's learning is affected not 820 only by their own ability to remember words and track statistics, but also by the 821 ambiguity of the learning environment (i.e. how many objects are around when they 822

hear a word).

Formalizing these models allows us to ask three questions: (1) What is the lower bound on time to learn if caregivers are motivated to teach and always engage in ostensive labeling? (2) If caregivers have a less altruistic goal—communication—how much longer would it take to learn? (3) If the child instead had to rely on learning from co-occurrence statistics, how powerful a statistical learner would a child have to be in order to match the rate of learning of a very simple learner in the context of goal-motivated caregivers?

One important point to note is that we are modeling the learning of a single word 830 rather than the entirety of a multi-word lexicon (as in Blythe et al., 2010). Although 831 learning times for each word could be independent, an important feature of many models of word learning is that they are not (Frank et al., 2009; Yu, 2008; Yurovsky et al., 2014; although c.f. McMurray, 2007). Indeed, positive synergies across words are predicted by 834 the majority of models and the impact of these synergies can be quite large under some 835 assumptions about the frequency with which different words are encountered (Reisenauer 836 et al., 2013). We assume independence primarily for pragmatic reasons here—it makes the 837 simulations significantly more tractable (although it is also what our experimental 838 participants appear to assume about learners). Nonetheless, it is an important issue for 839 future consideration. Of course, synergies that support learning under a cross-situational 840 scheme must also support learning from communicators and teachers (Frank et al., 2009; 841 Markman & Wachtel, 1988; Yurovsky et al., 2013). Thus, the ordering across conditions 842 should remain unchanged. However, the magnitude of the difference across conditions 843 could potentially increase or decrease.

45 Method

In each of the sections below, we describe the join models of caregivers'
communication and children's learning that predict learning times under each of the three

848 models of caregivers' goals.

Teaching.

Because the teaching caregiver is indifferent to communicative cost, they in ostensive labeling (pointing + speaking) on each communicative event. Consequently, learning on each trial occurs with a probability that depends entirely on the learner's learning rate $(P_k = p)$. Because we assume that the learner does not forget, the probability that a learner has failed to successfully learn after n trials is equal to the probability that they have failed to learn on each of n successive independent trials (The probability of zero successes on n trials of a Binomial random variable with parameter p). The probability of learning after n trials is thus:

$$P_k(n) = 1 - (1 - p)^n$$

The expected probability of learning after n trials was thus defined analytically and required no simulation. For comparison to the other models, we computed P_k for values of p that ranged from .1 to 1 in increments of .1.

861 Communication.

To test learner under the communication model, we implemented the same model described in the Model section. However, because our interest was in understanding the relationship between parameter values and learning outcomes rather than inferring the parameters that best describe people's behavior, we made a few simplifying assumptions to allow many runs of the model to complete in a more practical amount of time. First, in the full model above, speakers begin by inferring their own learning parameters (p_s) from their observations of their own learning, and subsequently use their maximum likelihood estimate as a stand-in for their child's learning parameter (p_l) . Because this estimate will

converge to the true value in expectation, we omit these steps and simply stipulate that the speaker correctly estimates the listener's learning parameter.

Second, unless the speaker knows a priori how many times they will need to refer to 872 a particular referent, the planning process is an infinite recursion. However, each future 873 step in the plan is less impactful than the previous step (because of exponential 874 discounting). This infinite process is in practice well approximated by a relatively small 875 number of recursive steps. In our explorations we found that predictions made from models 876 which planned over three future events were indistinguishable from models that planned over four or more, so we simulated three steps of recursion¹. Finally, to increase the speed 878 of the simulations we re-implemented them in the R programming language. All other aspects of the model were identical. 880

In our simulations, we varied the children's learning rate (p) from .1 to 1 in steps of .1 as in the Teaching simulation, caregivers' future-weighting (γ) from .1 to 1 in steps of .1, the caregivers' rationality (α) from .5 to 3 in steps of .5, and considered three values each of the cost of speaking (S=(0,10,20)) and pointing (P=(50,60,70)). The utility of communicating successfully was always 100.

6 Talking.

When caregivers are are producing a label for one of the objects in the environment,
but the child does not know which one, learning requires tracking how often that words
occurs with each potential object. This learning problem has been studied extensively in
the language acquisition literature under the guise of "cross-situational learning" (Yu &
Smith, 2007). Models of cross-situational learning have taken a variety of forms in order to
instantiate different theoretical positions about the mechanisms involved in learning, most

¹ It is an interesting empirical question to determine how the level of depth to which that people plan in this and similar games (see e.g. bounded rationality in Simon, 1991; resource-rationality in Griffiths et al., 2015). This future work is outside the scope of the current project.

centrally whether learning is hypothesis-driven or associative (see Yu & Smith, 2012 for a 893 review). In our analyses, we do not attempt to distinguish among these classes of models 894 as has been done in other learning-time simulations (e.g. Smith et al., 2011). Because of its 895 natural alignment with the learning models we use in the Teaching and Communication 896 simulations, we implemented a simple positive hypothesis testing model. Our choice to 897 focus on hypothesis testing rather than other learning frameworks is purely a pragmatic 898 choice—the learning parameter p in this model maps cleanly onto the learning parameter in 890 our other models. We encourage other researchers to adapt the code we have provided to 900 estimate the long-term learning for other models. 901

In this model, learners begin with no hypotheses and add new ones to their store as 902 they encounter data. Upon first encountering a word and a set of objects, the model 903 encodes up to h hypothesized word-object pairs each with probability p. On subsequent 904 trials, the model checks whether any of the existing hypotheses are consistent with the 905 current data, and prunes any that are not. If no current hypotheses are consistent, it adds 906 up to h new hypotheses each with probability p. The model has converged when it has 907 pruned all but the one correct hypothesis for the meaning of a word. This model is most 908 similar to the Propose but Verify model proposed in Trueswell et al. (2013), with the exception that it allows for multiple hypotheses. Because of the data generating process, 910 storing prior disconfirmed hypotheses (as in Stevens et al., 2017), or incrementing hypotheses consistent with some but not all of the data (as in Yu & Smith, 2012) has no 912 impact on learner and so we do not implement it here. We note also that, as described in 913 Yu and Smith (2012), hypothesis testing models can mimic the behavior of associative learning models given the right parameter settings (Townsend, 1990). 915

In contrast to the Teaching and Communication simulations, the behavior of the
Talking model depends on which particular non-target objects are present on each naming
event. We thus began each simulation by generating a corpus of 100 naming events. On
each event, we sampled the correct target as well as (C-1) competitors from a total set of

M objects. We then simulated learning over this set of events as described above, and recorded the first trial on which the learner converged (having only the single correct hypothesized mapping between the target word and target object). We repeated this process 1000 times for each simulated combination of M = (8, 16, 32, 64, 128) total objects, C = (1, 2, 4, 8) objects per trial, h = (1, 2, 3, 4) concurrent hypotheses, as the child's learning rate p varied from .1 to 1 in increments of .1.

26 Results

In order to understand how learning rates vary with model parameters, we first discuss the dependence of each of the three tested models on its parameters, and then discuss relationships between the models. For clarity of exposition, we analyze the number of events required for 75% of simulated learners to acquire the target word, and plot a representative subset of parameter values.

In addition the results reported here, we have made the full set of simulated results available in an interactive web application at dyurovsky.shinyapps.io/ref-sims. We encourage readers to fully explore the relationships among the models beyond the summary we provide.

${\it Teaching.}$

Because the Teaching model behaves identically on each trial regardless of the
learner, the rate of learning under this model depends entirely on the learner's learning
rate p. If the learning rate was high (e.g. .8), more than 75% of learners acquired the word
after a single learning instance. If the learning rate was medium, closer to the range we
estimated for adult learners (.6), more than 75% of learners acquired the word after only 2
instances. Finally, if the learning rate was very low (.2), the same threshold was reached
after 7 instances. Thus, the model is predictably sensitive to learning rate, but even very
slow learners are expected to acquire words after a small number of communicative events.

5 Communication

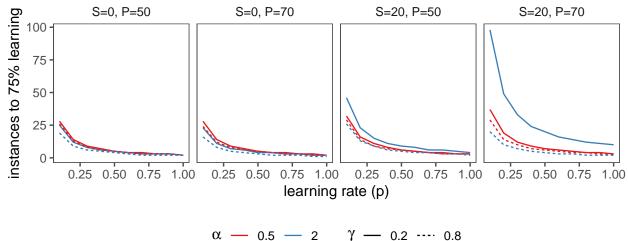


Figure 9

Number of exposures required for 75% of children to learn a word under the Communication model as parameters vary. Color shows rationality (α) , Linetype shows future weighting (γ) , facets indicate the cost of speaking (S) and pointing (P). The middle two facets corresponds to Higher Speech Efficiency and Lower Speech efficiency conditions of the experiment.

The Communication model's behavior depends on parameters of both the child 946 learner and the caregiver communicator. In general, parameters of both participants had 947 predictable effects on learning: Children learned faster when they had higher learning 948 rates, when caregivers were more rational, and when caregivers gave greater weight to the 949 future. Further, the effects of caregivers' parameters were more pronounced at the lowest 950 learning rates. However, as the cost of speaking increased relative to pointing, the effects of 951 caregivers' parameters changed. In particular, highly rational caregivers who heavily discounted the future lead to significantly slower learning. At these parameter settings, the 953 caregiver becomes very likely to point on any given trial in order to maximize the local 954 utility at the expense of discounted future utility gained from teaching. In addition, as the 955 cost of both modalities increases, the utility of communicating successfully (here defined as 956 100 points) becomes less motivating. Thus, caregivers become less discriminating among 957

their communicative choices. Figure 9 shows the number of trials required for 75% of learners to acquire a word as a function of parameters in the Communication model.

960 Talking.

Finally, when caregivers spoke on each trial and children had to learn from 961 cross-situational statistics, learning was controlled by the the child's learning rate, the 962 number of hypotheses the child could entertain, the number of objects per event, and to a 963 small extent the total vocabulary size. In general, children learned faster when they had a higher learning rate, and could entertain more hypotheses. Learning was also predictably slower when there were more objects on each event and thus ambiguity was higher. Finally, as the total vocabulary size increased, the rate of learning increased slightly, as it does with 967 human cross-situational learners (Yu & Smith, 2007). This counter-intuitive outcome 968 occurs because the rate of spurious co-occurrences, in which the target word consistently 969 co-occurs with an object that is not its referent, decreases as the set of potential foils 970 expands. The the effect of context size (C) and number of hypotheses can be seen along 971 with the learning rates of the other two models in Figure 10. 972

973 Comparing the Models

Because the real-world parameters appropriate for each model are difficult to
determine, we consider the relationship between the models over the range of their possible
parameters. Figure 10 shows the time for 75% of learners to acquire a word in each of the
three models. Across all possible child learning rates (p), the Teaching model lead to the
fastest learning as expected. We can treat this model as a lower bound how quickly
learning could possibly happen.

For the Communication model, we considered the range of all possible rates of learning that could unfold as the parameters of both child and caregiver varied. The range

was substantial. If caregivers weigh the future near equally to the present, and are highly 982 rational, the child's resultant rate of learning is nearly identical to the rate of learning 983 under the Teaching model: Children required 1.07 times as many learning instances under 984 the Communication model as the Teaching model when averaging over all child learning 985 rates. In contrast, if the caregiver weighs the future much less than the present, and is 986 relatively irrational about maximizing utility, the rate of learning can be quite slow—in the 987 worst case requiring children to have 24.30 as many learning instances as under the 988 Teaching model. Despite this bad worst case scenario, if caregivers' parameters are close to 980 the ones we estimated in our experiment, Communication would require only 1.75 as many 990 instances as Teaching if speech is high efficiency relative to pointing, and 3.12 as many 991 instances if speech is lower efficiency. 992

For the Talking model, we also observed a wide range of learning times as a function of both the ambiguity of the learning environment and the number of simultaneous hypotheses that the child can maintain. When the environment was unambiguous—only 2 objects were present at a time—and the child could encode both, learning under Talking took only 2.03 times as many instances as Teaching. In contrast, if ambiguity was high, and learners could only track a single hypothesis, learning was significantly slower under Talking than Teaching, (requiring 10.05 times as many instances).

Comparing Communication and Talking to each-other, we find that that Talking 1000 can lead to faster learning under some parameter settings. In particular, if events are low 1001 in ambiguity, or children can maintain a very large number of hypotheses about the 1002 meaning of a word relative the number of objects in each event, children can learn rapidly 1003 even if caregivers are just Talking. This learning can be faster than simpler child models 1004 learning from highly myopic or relatively irrational caregivers Communicating, especially if 1005 speech is high-cost. At medium levels of ambiguity, Communication and Talking are 1006 similar and their ordering depends on other parameters. At high levels of ambiguity 1007 Communication is the clear winner. 1008

Together, these results suggest that if the set of possible candidate referents is 1009 small, even simple cross-situational learners can cope just fine even if their caregiver is just 1010 Talking; they learn roughly two to three times more slowly than if their caregiver was 1011 Teaching them. However, if the set of possible referents is four, or, eight, or even more on 1012 average, cross-situational learners need to have very high bandwidth or their rates of 1013 learning will be an order of magnitude slower than if their caregiver were Teaching them. 1014 In these cases, even the simplest learner—who can encode a single hypothesis about the 1015 meaning of a word and gets no information from co-occurrence statistics—can learn quite 1016 rapidly if they are learning from a caregiver that Communicates with them. 1017

1018 Discussion

1019

1020

102

1022

1023

1024

Most of the language that children hear from their caregivers is unlikely to be designed to teach them language. However, the language that caregivers direct to them *is* designed to communicate successfully. Here we consider the learning consequences of these differences in design. How different are the learning consequences of language designed for teaching, language designed for communication, and ambient language not designed for the child at all?

If input is not communicatively motivated, the rate of learning depends entirely on 1025 what the learner brings to the table. In line with prior analyses of cross-situational 1026 learning, we find that learning can be quite rapid if environments are low in ambiguity or 1027 the learner has very high bandwidth for storing candidate hypotheses (Smith et al., 2011; 1028 Yu & Smith, 2012). However, the child's environment is neither guaranteed to be 1029 unambiguous nor are young children likely to have high bandwidth for statistical 1030 information (Medina et al., 2011; Vlach & Johnson, 2013; Woodard et al., 2016). In fact, 1031 when the set of candidate referents is small, it is quite likely to be small in part because 1032 caregivers have designed the context to support communication (Tomasello & Farrar, 1986). 1033

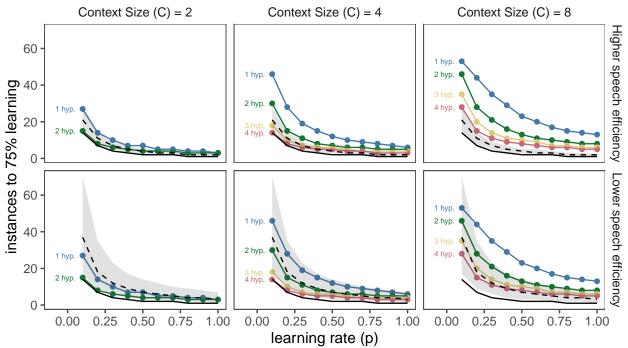


Figure 10

Comparing the number of exposures required for 75% of children to learn a word under all three models as parameters vary. Columns show variation in context size (C), a parameter of the Talking model. Rows show the two variations in the costs of Speech and Pointing for the Communication model used in our experiments. In each facet, the solid black line shows learning under the Teaching model, the light gray region shows an envelope of learning times corresponding to all variations in Communication model parameters, and the black dotted line shows learning time under the Communication model with parameters equal to the empirical estimates from experiments. Colored lines show learning times under the Talking model with varying numbers of hypotheses. Because there was little effect of the total number of objects (M) in the Talking model, all panels show results for 128 objects. Note that Communication model parameters vary across rows, while Talking model parameters vary across columns.

Learning from communication consistently outperforms learning from ambient 1034 language for all but the most precocious learners. If we take learning from teaching as an 1035 upperbound, we see that the rate of learning from communication is almost as fast under 1036 many possible parameter settings we explored. On average, across all possible parameter 1037 values, learning from communication is only 2.5 times slower than learning from teaching. 1038 Further, in this model, the learner gets no information from co-occurrence statistics at all. 1039 Combining learning from communication with low-bandwidth cross-situational learning 1040 could bring the expected rate of learning down to very close to learning from teaching 1041 (MacDonald et al., 2017). We thus might make significant progress on understanding how 1042 children learn language so quickly not just by studying children, but also by understanding 1043 how caregivers design the language they produce in order to support successful 1044 communication (Leung et al., 2021). 1045

General Discussion

Across naturalistic corpus data, experimental data, and model predictions and 1047 simulation, we see evidence that pressure to communicate successfully with a linguistically 1048 immature partner could fundamentally structure caregiver production and shape child 1040 learning. In our experiment, we showed that people tune their communicative choices to 1050 varying cost and reward structures, and also critically to their partner's linguistic 1051 knowledge- providing richer cues when partners are unlikely to know the language and 1052 many more rounds remain. These data are consistent with the patterns shown in our 1053 corpus analysis of caregiver referential communication and demonstrate that such 1054 pedagogically supportive input could arise from a motivation to maximize communicative 1055 success while minimizing communicative cost—no additional motivation to teach is 1056 necessary. In simulation, we demonstrate that simple learners whose caregivers want to 1057 communicate with them out-learn more powerful statistical learners whose caregivers do 1058 not have a communicative goal. 1059

Accounts of language learning often aim to explain its striking speed in light of the 1060 sheer complexity of the language learning problem itself. Many such accounts argue that 1061 simple (associative) learning mechanisms alone seem insufficient to explain the rapid 1062 growth of language skills and appeal instead to additional explanatory factors, such as the 1063 so-called language acquisition device, working memory limitations, word learning biases, 1064 and many more (e.g., Chomsky, 1965; Goldowsky & Newport, 1993; Markman, 1990). 1065 While some have argued for the simplifying role of language distributions (e.g., McMurray, 1066 2007), these accounts largely focus on learner-internal explanations. For example, Elman 1067 (1993) simulates language learning under two possible explanations to intractability of the 1068 language learning problem: one environmental, and one internal. He first demonstrates 1069 that learning is significantly improved if the language input data is given incrementally, 1070 rather than all-at-once. He then demonstrates that similar benefits can arise from learning 1071 under limited working memory, consistent with the "less-is-more" proposal (Elman, 1993; 1072 Goldowsky & Newport, 1993). Elman dismisses the first account arguing that ordered 1073 input is implausible, while shifts in cognitive maturation are well-documented in the 1074 learner; our account's emphasis on calibration to such learning mechanisms suggests the 1075 role of ordered or incremental input from the environment may be crucial. Our findings 1076 support the idea that rapid language learning may be facilitated by the *combination* of the 1077 learner's limited statistical learning skills combined with communicatively (but not 1078 pedagogically) motivated caregiver input. Such results emphasize the importance of a 1079 dyadic learning approach, whereby considering the joint contributions of learner and 1080 caregiver can yield new insights (Yurovsky, 2018). 1081

This account is consonant with work in other areas of development, such as recent demonstrations that the infant's visual learning environment has surprising consistency and incrementality, which could be a powerful tool for visual learning. Notably, research using head mounted cameras has found that infant's visual perspective privileges certain scenes and that these scenes change across development. In early infancy, the child's

1088

1089

1090

1091

1092

1093

1094

1095

1096

1097

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

1110

1111

1112

1113

egocentric visual environment is dominated by faces, but shifts across infancy to become more hand and hand-object oriented in later infancy (Fausey et al., 2016). This observed shift in environmental statistics mirrors learning problems solved by infants at those ages, namely face recognition and object-related goal attribution respectively. These changing environmental statistics have clear implications for learning and demonstrate that the environment itself is a key element to be captured by formal efforts to evaluate statistical learning (Smith et al., 2018). Frameworks of visual learning must incorporate both the relevant learning abilities and this motivated, contingent structure in the environment.

By analogy, the work we have presented here aims to draw a similar argument for the language environment, which is also demonstrably beneficial for learning and changes across development. In the case of language, the contingencies between learner and environment are even clearer than visual learning. Functional pressures to communicate and be understood make successful caregiver speech highly dependent on the learner. Any structure in the language environment that is continually adapting to changing learning mechanisms must come in large part from caregivers themselves. Thus, a comprehensive account of language learning that can successfully grapple with the infant curriculum must explain caregiver production as well as learning itself. Recent work has shown that many aspects of the the structure of natural languages, both in semantics and in syntax, can be explained by evolutionary optimization of language to be more efficient for communication (see Gibson et al., 2019 for a review). Further, people talking with other native language speakers make in-the-moment word choices that are well predicted by models of optimal communication (e.g., Mahowald et al., 2013). Here we extend these ideas to asymmetric communications between speakers and less knowledgeable listeners, and show that they can predict how caregivers modify their communicative acts when talking with their children.

Explaining caregiver modification is a necessary condition for building a complete theory of language learning, but modification is certainly not a sufficient condition for language learning. No matter how calibrated the language input, non-human primates are

unable to acquire language. Indeed, caregiver modification need not even be a necessary condition for language learning. Young children are able to learn novel words from (unmodified) overheard speech between adults (Foushee et al., 2016; although c.f. Shneidman & Goldin-Meadow, 2012). Our argument is that the rate and ultimate attainment of language learners will vary substantially as a function of caregiver modification, and that describing the cause of this variability is a necessary feature of models of language learning.

Our account aims to explain caregivers' production and child learning in the same system, putting these processes into explicit dialogue. While we have focused on ostensive labeling as a case-study phenomenon, our account should reasonably extend to the changing structure found in other aspects of child-directed speech. Some such phenomena will be easily accounted for; aspects of language that shape communicative efficiency should shift in predictable patterns across development. For example, the exaggerated pitch contours seen in infant-directed speech serve to draw infants' attention and facilitate phoneme learning. These language modifications are well-explained by our proposed framework, though incorporating them will likely require altering aspects of our account and decisions about which alterations are most appropriate. In the example of exaggerated pitch, one could expand the definition of communicative success to include the goal of maintaining attention, or accomplish the same goal by altering the cost structure to penalize loss of engagement. Thus, while this account should generalize to other modifications found in child-directed speech, such generalizations will likely require alterations to the extant structure of the framework.

Of course, not all aspects of language should be calibrated to the child's language development. Our account also provides an initial framework for explaining aspects of communication that would not be modified in child-directed speech: aspects of communication that minimally affect communicative efficiency. In other words, communication goals and learning goals are not always aligned. For example, young

children sometimes overregularize past and plural forms, producing incorrect forms such as 1141 "runned" or "foots" (rather than the irregular verb "ran" or irregular plural "feet", Marcus 1142 et al., 1992). Mastering the proper tense endings (i.e. the learning goal) might be aided by 1143 feedback from caregivers; however, adults rarely provide explicit corrective feedback for 1144 these errors (Marcus, 1993). This is perhaps because incorrect grammatical forms 1145 nonetheless successfully communicate their intended meaning, and thus do not prevent the 1146 successful completion of the communicative goal of language (Chouinard & Clark, 2003). 1147 The degree of alignment between communication and learning goals should predict the 1148 extent to which a linguistic phenomenon is modified in child-directed speech. 1149

Some aspects of caregiver production are unrepresented in our framework, such as 1150 aspects of production driven by speaker-side constraints. Furthermore, our account is 1151 formulated primarily around concrete noun learning and future work must address its 1152 viability in other aspects of language learning. We chose to focus on ostensive labeling as a 1153 case-study phenomenon because it is an undeniably information-rich cue for young 1154 language learners, however ostensive labeling varies substantially across socio-economic, 1155 linguistic, and cultural groups (Hoff, 2003). This is to be expected to the extent that 1156 caregiver-child interaction is driven by different goals (or goals given different weights) 1157 across these populations-variability in goals could give rise to variability in the degree of 1158 modification. Indeed, child-directed speech itself varies cross-linguistically, both in its 1159 features (Fernald et al., 1989) and quantity (e.g., Shneidman & Goldin-Meadow, 1160 2012)—although, there is some evidence that child-directed speech predicts learning even in 1161 cultures where it is qualitatively different and less prevalent than in American samples 1162 (Shneidman & Goldin-Meadow, 2012). Future work is needed to establish the 1163 generalizability of our account beyond the western samples studied here. 1164

We see this account as building on established, crucial statistical learning skills—distributional information writ large and (unmodified) language data from overheard speech are undoubtedly helpful for some learning problems (e.g., phoneme learning). There

1165

1166

1167

is likely large variability in the extent to which statistical learning skills drive learning for a 1168 given learning problem, which could derive from domain or cultural differences. 1169 Understanding generalizability of this sort and the limits of statistical learning will likely 1170 require a full account spanning both caregiver production and child learning. A full 1171 account that explains variability in modification across aspects of language will rely on a 1172 fully specified model of optimal communication. Such a model will allow us to determine 1173 both which structures are predictably unmodified, and which structures must be modified 1174 for other reasons. Nonetheless, this work is an important first step in validating the 1175 hypothesis that language input that is structured to support language learning could arise 1176 from a single unifying goal: The desire to communicate effectively. 1177

1178 Conclusion

1179

1180

1181

1182

1183

Building on early functional accounts of language learning, our perspective considers the caregiver-child dyad as the fundamental unit of analysis and emphasizes the importance of communicative success in shaping language input and language learning. We have developed an initial formal account for jointly considering caregiver productions and child language learning within the same system.

Such an account helps to explain caregivers' increased production of ostensive 1184 labeling in naturalistic communication for infrequent referents and when children are very 1185 young. In an experimental context, pressure to with a less knowledgeable partner results in 1186 analogous behavior from participants in an iterated reference game. A rational 1187 communication model that includes planning can explain these behaviors without an 1188 explicit teaching goal, and such a model produces learning outcomes that outperform all 1189 but the best cross-situational learners in simulations. In sum, this work demonstrates that 1190 the pressure to communicate successfully across knowledge asymmetries may help create a 1191 learning environment that fosters language learning. Rapid language learning in early 1192 childhood may be best explained by considering that the child's powerful, though not 1193

precocious, learning mechanisms are met with a language environment specifically designed to communicate successfully with them.

Acknowledgement

The authors are grateful to Madeline Meyers for her work coding referential communication in the corpus data, and to Mike Frank and Chen Yu for their feedback on the manuscript. This research was funded by James S. McDonnell Foundation Scholar Award in Understanding Human Cognition #220020506 to DY.

1201 References

- Austin, J. L. (1975). How to do things with words (Vol. 88). Oxford university press.
- Baldwin, D. (2000). Interpersonal understanding fuels knowledge acquisition. Current

 Directions in Psychological Science, 9, 40–45.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- https://doi.org/10.18637/jss.v067.i01
- Bloom, P. (2000). How children learn the meanings of words. MIT press: Cambridge, MA.
- Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, 34, 620–642.
- Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language.

 Annual Review of Developmental Psychology, 1, 223–249.
- Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for "motionese":
- Modifications in mothers' infant-directed action. Developmental Science, 5(1), 72–83.
- Brown, R. (1977). Introduction. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to*children: Language input and interaction. Cambridge, MA.: MIT Press.
- Callanan, M. A. (1985). How parents label objects for young children: The role of input in the acquisition of category hierarchies. *Child Development*, 508–523.
- 1220 Chomsky, N. (1965). Aspects of the theory of syntax. MA: MIT Press.
- Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, 30, 637–669.
- Clark, H. H. (1996). Using language. In *Journal of Linguistics* (pp. 167–222). Cambridge
 Univ Press.

- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, 208(4448), 1174–1176.
- DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development*, 9(2), 133–150.
- Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological Review*, 123(6), 758.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99.
- Estes, K. G., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science, 18(3), 254–260.
- Evans, O., Stuhlmüller, A., Salvatier, J., & Filan, D. (2017). Modeling Agents with

 Probabilistic Programs. http://agentmodels.org.
- Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, 152, 101–107.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & Fukui, I.

 (1989). A cross-language study of prosodic modifications in mothers' and fathers'

 speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501.
- Foushee, R., Griffiths, T. L., & Srinivasan, M. (2016). Lexical complexity of child-directed and overheard speech: Implications for learning.
- Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, 336, 998–998.
- Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, 75, 80–96.

- Frank, M. C., Goodman, N., & Tenenbaum, J. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20, 578–585.
- Franke, M. (2013). Game theoretic pragmatics. Philosophy Compass, 8(3), 269–284.
- Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge university press.
- Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy,
 R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*,
 23(5), 389–407.
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese:

 The role of temporal synchrony between verbal labels and gestures. *Child*Development, 71(4), 878–894.
- Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist*, 69(6), 588–599.
- Goldowsky, B. N., & Newport, E. L. (1993). Limitations on the acquisition of morphology:

 The less is more hypothesis. The Proceedings of the Twenty-Fourth Annual Child

 Language Research Forum, 124.
- Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109–135.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, 20(11), 818–829.
- Goodman, N. D., & Stuhlmüller, A. (2014). The Design and Implementation of

 Probabilistic Programming Languages. http://dippl.org.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and

- semantics: Vol. 3: Speech acts (pp. 41–58). New York, NY: Academic Press.
- Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources:
- Levels of analysis between the computational and the algorithmic. *Topics in*
- $Cognitive\ Science,\ 7(2),\ 217-229.$
- Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects
- early vocabulary development via maternal speech. Child Development, 74 (5),
- 1368-1378.
- Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in
- partially observable stochastic domains. Artificial Intelligence, 101, 99–134.
- Leung, A., Tunkel, A., & Yurovsky, D. (2021). Parents fine-tune their speech to children's
- vocabulary knowledge. Psychological Science, 32(7), 975–984.
- Luce, R. D. (1959). Individual choice behavior.
- MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the
- representations underlying cross-situational learning. Cognitive Psychology, 94,
- 1288 67-84.
- MacKay, D. J. (1992). Bayesian interpolation. Neural Computation, 4(3), 415–447.
- Mahowald, K., Fedorenko, E., Piantadosi, S. T., & Gibson, E. (2013). Info/information
- theory: Speakers choose shorter words in predictive contexts. Cognition, 126(2),
- 313–318.
- Marcus, G. F. (1993). Negative evidence in language acquisition. Cognition, 46(1), 53–85.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H.
- (1992). Overregularization in language acquisition. Monographs of the Society for
- 1296 Research in Child Development, i–178.
- Markman, E. M. (1990). Constraints children place on word meanings. Cognitive Science,
- 14(1), 57-77.

- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157.
- Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York, NY: W. H. Freeman.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3), B101–B111.
- McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, 317(5838), 631–631.
- McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence? *Cognition*, 129(2), 362–378.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019.
- Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. *Perspectives*on Imitation: From Neuroscience to Social Science, 2, 55–77.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. Cognition, 90(1), 91-117.
- Newport, E. L. (1990). Maturational constraints on language learning. Cognitive Science, 14(1), 11-28.
- Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, I'd rather do it myself:
- Some effects and non-effects of maternal speech style. In C. A. Ferguson (Ed.),
- Talking to children language input and interaction (pp. 109–149). Cambridge
- University Press.
- Ninio, A. (1980). Ostensive definition in vocabulary teaching. Journal of Child Language,

- 7(3), 565-573.
- Quine, W. V. O. (1960). Word and object. In *Cambridge, Mass*. Cambridge, Mass.: MIT

 Press.
- Rafferty, A. N., Brunskill, E., Griffiths, T. L., & Shafto, P. (2016). Faster teaching via pomdp planning. *Cognitive Science*, 40(6), 1290–1332.
- R Core Team. (2021). R: A language and environment for statistical computing. R

 Foundation for Statistical Computing. https://www.R-project.org/
- Reisenauer, R., Smith, K., & Blythe, R. A. (2013). Stochastic dynamics of lexicon learning in an uncertain and nonuniform world. *Physical Review Letters*, 110(25), 258701.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. Current

 Directions in Psychological Science, 12(4), 110–114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Scott, R. M., & Fischer, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, 122, 163–180.
- Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, 71, 55–89.
- Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: How important is directed speech? *Developmental Science*, 15(5), 659–673.
- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization*Science, 2(1), 125–134.
- Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, 35(3), 480–498.

- Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, 22(4), 325–336.
- Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558–1568.
- Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. Language Learning and Development, 9, 25–49.
- Snow, C. E. (1972). Mothers' speech to children learning language. *Child Development*, 43, 549–565.
- Snow, C. E. (1977). Mothers' speech research: From input to interaction. Talking to

 Children: Language Input and Acquisition, 3149.
- Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, 41, 638–676.
- Thiessen, E., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7, 53–71.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child*Development, 57, 1454–1463.
- Townsend, J. T. (1990). Serial vs. Parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished.

 Psychological Science, 1(1), 46–54.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify:

 Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1),

 126–156.
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, 127(3), 375–382.

- Vogt, P. (2012). Exploring the robustness of cross-situational learning under zipfian distributions. *Cognitive Science*, 36(4), 726–739.
- Wittgenstein, L. (1953). *Philosophical investigations*. Oxford, UK: Basil Blackwell.
- Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development*, 12(3), 252–261.
- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–272.
- Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. Language Learning and Development, 4(1), 32–62.
- 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. (2018). A communicative approach to early word learning. New Ideas in

 Psychology, 50, 73–79.
- Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to children's developmental level. *Proceedings of the Annual Meeting of the Cognitive Science*Society, 2093–2098.
- Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, 145, 53–62.
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, 21, 1–22.
- Yurovsky, D., Meyers, M., Burke, N., & Goldin-Meadow, S. (2018). Children gesture when speech is slow to come. In J. Z. Chuck Kalish Martina Rau & T. Rogers (Eds.),

- Proceedings of the 40th annual meeting of the cognitive science society (pp. 2765–2770).
- Yurovsky, D., Yu, C., & Smith, L. B. (2012). Statistical speech segmentation and word
 learning in parallel: Scaffolding from child-directed speech. Frontiers in Psychology,
 3, 374.
- Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, 37, 891–921.