Pressure to communicate across knowledge asymmetries leads to pedagogical-like language input

Anonymous CogSci submission

Abstract

Child-directed speech might support learning in part due to communicative pressure: parents must use language that their children understand. We present longitudinal corpus data of parent-child interaction to demonstrate that parents provide information-rich referential communication (using both gesture and speech during the same reference) more for infrequent referents and when their children are younger. We then present a Mechanical Turk study to experimentally validate this idea, asking Turkers to communicate with a listener who a novel language less well. Participants could communicate in 3 ways: pointing (expensive but unambiguous), labelling (cheap but knowledge-dependent), or both. They won points only for communicating successfully; using pointing and labelling together was costly, but could teach, allowing cheaper communication on later trials. Participants modulated their communicative behavior in response to their own and their partner's knowledge and teaching emerged when the speaker had substantially more knowledge of the lexicon. Speaker behavior in this paradigm can be explained by a rational planning model where speakers attempt to maximize total expected utility over the course of the game. While language is more than reference games, this work validates the hypothesis that communicative pressure alone can lead to supportive language input.

Keywords: Language; Learning; Child-Directed Speech; POMDP.

Introduction

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?

One possibility is that children's caregivers deserve most of the credit; that the language parents produce to their children is optimized for teaching. Although there is some evidence that aspects of child-directed speech support learning, other aspects–even in the same subproblem, e.g. phoneme discrimination–appear to make learning more difficult (Eaves Jr, Feldman, Griffiths, & Shafto, 2016; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). In general, parents rarely explicitly correct their children, and children are resistant to the rare explicit language correction they do get (Newport, Gleitman, & Gleitman, 1977). Thus while parents may occasionally offer a supervisory signal, the bulk of the evidence suggests that parental supervision is unlikely to explain rapid early language acquisition

Alternatively, even the youngest infants may already come to language acquisition with a precocious ability to learn the latent structure of language from the statistical properties of speech in their ambient environment (Saffran & 2003, 2003; L. B. Smith & Yu, 2008). While a number of experiments clearly demonstrate the early availability of such mechanisms, there is reason to be suspicious about just how precocious they are early in development. For example, infants' ability to track the co-occurrence information connecting words to their referents appears to be highly constrained by both their developing memory and attention systems (L. B. Smith & Yu, 2013; Vlach & Johnson, 2013). Further, computational models of these processes show that the rate of acquisition is highly sensitive to variation in environmental statistics (Blythe, Smith, & Smith, 2010; Vogt, 2012). Thus precocious unsupervised statistical learning also appears to fall short of an explanation for rapid early language learning.

We explore the consequences of a third possibility: The language that children hear is neither designed for pedagogy, nor is it random; it is designed for communication (Brown, 1977). We take as the caregiver's goal the desire to communicate with the child, not about language itself, but instead about the world in front of them. To succeed, the caregiver must produce the kinds of communicative signals that the child can understand, and thus might tune the complexity of their speech not for the sake of learning itself, but as a byproduct of in-the-moment pressure to communicate successfully (Yurovsky, 2017).

To examine this hypothesis, we first analyze parent communicative behavior in a longitudinal corpus of parent-child interaction in the home (Goldin-Meadow et al., 2014). We investigate the extent to which parents tune their communicative behavior across their child's development to align to their child's developing linguistic knowledge, as demonstrated at other levels of language analysis (e.g., sytnax, see Yurovsky, Doyle, & Frank, 2016). Specifically, we look at *communicative modality* to analyze parent's use of multi-modal reference– using *both* gesture and a verbal label to indicate the same referent, in the same instance. These instances might be particularly powerful learning opportunities for young children because they provide a verbal label in the presence of a highly disambuiguting gestural cue (e.g., pointing), which might greatly reduce referential uncertainty

(gesture-speech citation)**.**

We then turn to the emergence of structured input in a simple model system: an iterated reference game in which two players earn points for communicating successfully with each other. Modeled after our corpus data, particpants are asked to make choices about which communicative strategy to use (akin to modality choice). In an experiment on Mechanical Turk using this model system, we experimentally induce structured language input from a pressure to communicate. We then show that human behavior can be explained by a rational planning model that seeks to optimize its total expected utility over the course of the game.

Corpus Analysis

To begin with, we investigate linguistic input in naturalistic, parent-child corpus data. We focus on parent referential communication to examine how parents might be aligning to their children. The degree of parental alignment should be sensitive to the child's age, such that parents will be more likely to provide richer communicative information when their child is younger, when she has less linguistic knowedge, than as she gets older (Yurovsky et al., 2016). Parental alignment should also be stronger for infrequent objects, where again children are likely to have less knowledge of the relevant label. We analyze the production of multi-modal reference for the same referent, in the same instance, which could reflect alignment if parents are senstively using this information-rich cue for young children and infrequent objects.

Methods

We used come from the Language Development Project, a large-scale, longitudinal corpus of parent child-interaction in the home with families who are representative of the Chicago community in socio-economic, racial, and gender diversity (Goldin-Meadow et al., 2014). These data are drawn from a sample of 10 families from the greater corpus. Recordings were 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.

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. An independent coder analyzed each of these communicative instances and identified each time a concrete noun was referenced using speech (in specific noun form), gesture (only deictic gestures were coded for ease of coding and interpretation– e.g., pointing) or both simultaneously.

Results

These corpus data were analyzed using a mixed effects regression to predict parent use of multi-modal reference for a given referent. Random effects of subject and referent were included in the model. Our key predictors were child age and logged referent frequency (i.e. how often a given object was referred to overall across our data).

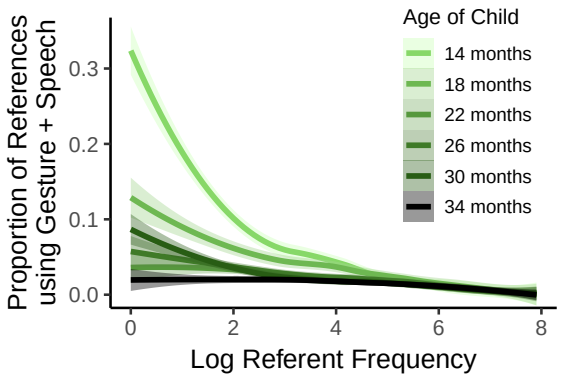

Figure 1: Proportion of parent multi-modal referential talk across development. The log of a referent's frequency is given on the x-axis, with less frequent items closer to zero.

We find a signficant negative effect of child age on multi-modal reference, such that parents are signficantly less likely to provide the gesture-speech cue as their child gets older (B < -0.04, $p < 0.0001$). Examing referents, we find a signficant negative effect of referent frequency on multi-modal reference as well, such that parents are signficantly less likely to provide the gesture-speech cue for frequent referents than infrequent ones (B < -0.13, $p < 0.0001$). Thus, in these data, we see early evidence that parents are providing richer, structured input about rarer things in the world for their younger children.

Discussion

We have shown that parents provided more multi-modal reference for less familiar objects and when their child is younger. These longitudinal corpus findings are consistent with an account of parental alignment: parents are sensitive to their child's developing linguisitc knowledge and adjust their communicative behavior accordingly. These alignment data could be argued to derive from a pedagoical pressure to teach younger children, however we will argue that these data could be explained by a simpler, potentially-selfish pressure: to communicate successfully. The distinction between these pressures is difficult to draw from naturalistic data, so we next developed a paradigm to try to experimentally induce richly-structured, aligned input from a pressure to communicate in the moment.

Experimental Framework

We developed a simple reference game in which participants would be motivated to communicate successfully on a trial-by-trial basis. By manipulating the relative costs of the communicative methods and examining the role of asymmetric knowledge states, we aim to experimentally determine the circumstances under which richly-structured input emerges,

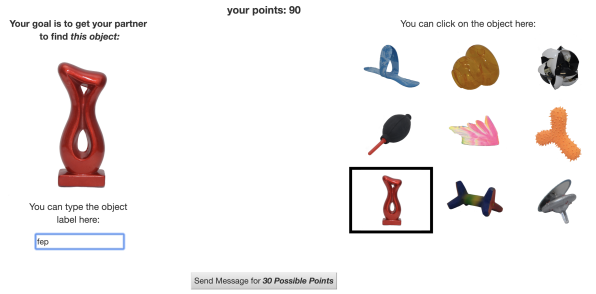

Figure 2: Screenshot of speaker view during gameplay.

without an explicit pedagogical goal. In all conditions, paritcpants are placed in the role of speaker and asked to communicate with a computerized listner whose responses were programmed to be contingent on speaker behavior.

Method

Participants 480 participants were recruited though Amazon Mechanical Turk and received a $1 payment for their participation. Data from 51 participants were excluded from subsequent analysis for failing the critical manipulation check and a further 28 for producing pseudo-English labels (e.g., 'pricklyyone'). The analyses reported exclude the data from thos participants, but all analyses were also conducted without excluding any participants and all patterns hold ($ps < 0.01$).

Design and Procedure 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, two times, or one time. Thus, participants had a total of 20 training trials. Participants were then given a simple recall task to establish their baseline knowledge of the novel lexicon (pretest).

Prior to beginning the game, participants are told how much exposure their partner has had to the lexicon and also that they will be asked to discuss each object three times. As a simple manipulation check, participants are then asked to report their partner's level of exposure, and are corrected if they answer wrongly. Then during gameplay, speakers saw a target object in addition to an array of all nine objects (see Figure 2 for the speaker's perspective). Speakers had the option of either directly click on the target object in the array (gesture)- a higher cost cue but without ambiguity- or typing a label for the object (speech)- a lower cost cue but contingent on the listener's shared linguistic knowledge. After sending the message, speakers are shown which object the listener selected.

Speakers could win up to 100 points per trial if the listener correctly selected the target referent based on their message. We manipulated the relative utilities of each of the strategies between-subjects across two conditions: 'Talk is Cheap' and 'Talk is Less Cheap.' In the 'Talk is Cheap' condition, sending a message by gesturing cost 70 points while sending a label cost 0 points, yielding 30 points and 100 points respectively if the listner selected the target object. In the 'Talk is Less Cheap' condition speakers were charged 50 points for gesturing and 20 points for labeling, yielding up to 50 points and 80 points respectively. If the listener failed to identify the target object, the speaker nevertheless paid the relevant cost for that message in that condition. As a result of this manipulation, there was a higher relative expected utilty for labeling in the 'Talk is Cheap' conditon than the 'Talk is Less Cheap' condition.

Critically, particpants were told about a third type of possible message using both gesture and speech within a single trial to effectively teach the listener an object-label mapping. This action directly mirrors the multi-modal reference behavior from our corpus data– it presents the listener with an information-rich, potentially pedagogical learning moment. In order to produce this teaching behavior, speakers had to pay the cost of producing both cues (i.e. both gesture and speech). Note that, in all utility conditions, teaching nets participants 30 points.

Incorporating teaching means that our speaker must also reason about their interlocutor's knowledge state more explicitly, in order to make rational decisions about what to teach and when. To address this added dimension, we also manipulated participants' expectations about their partner's knowledge across 3 conditions. Participants were told that their partner had either no experience with the lexicon, had the same experience as the speaker, or had twice the experience of the speaker.

Listeners were programmed with starting knowledge states initialized accordingly. Listeners with no exposure began the game with knowledge of 0 object-label pairs. Listeners with the same exposure of the speaker began with knowledge of five object-label pairs (3 high frequency, 1 mid frequency, 1 low frequency), based the average retention rates found previously. Lastly, the listener with twice as much exposure as the speaker began with knowledge of all nine object-label pairs. If the speaker produced a label, the listener was programmed to consult their own knowledge of the lexicon and check for similar labels (selecting a known label with a levenshtein edit distance of two or fewer from the speaker's production), or select among unknown objects if no similar labels are found. Listners could integrate new words into their knowledge of the lexicon if taught.

Crossing our two between-subjects manipulations, we had 6 conditions (2 utility condition: 'Talk is Cheap' and 'Talk is Less Cheap'; and 3 levels of partner's exposure: none, same, double), with 80 participants in each condition. We expected to find results that mirrored our corpus findings such that there would be higher rates of teaching behavior when there was an asymmetry in knowledge where the speaker knew more (none conditions) comapred with when there was equal knowledge (same conditions) or when the listener was more familiar with the languauge (double conditions). If partic-

pants are motivated to communciate efficiently, we expected that particpants would also be sensitive to our utility manipulation, such that rates of labeling and teaching would be higher in the 'Talk is Cheap' conditions than the other conditions.

Results

As expected, participants were sensitive to both the exposure rate and the relative utilities of the communication strategies. As an initial check of our knowledge manipulation, a logistic regression indicated that the more exposures to a given object-label pair during training, the more likely participants were to recall that label at pretest (**B = XYZ, p < XYZ**). On average, participants knew at least 6 of the 9 words in the lexicon (mean = 6.28, sd = 2.26).

Gesture-Speech Tradeoff. To determine how gesture and speech are trading off across conditions, we looked at a mixed effects logistic regression to predict whether speakers chose to produce a label during a given trial as a function of the exposure rate, object instance in the game (first, second, or third), utility manipulation, and partner manipulation. A random subjects effects term was included in the model. There was a significant effect of exposure rate such that the more exposures to a particular object-label pair during training, the more likely a speaker was to produce a label (B = 0.59, $p < 0.0001$). Compared with the first instance of an object, speakers were significantly more likely to produce a label on the second appearance (B = 0.2, $p < 0.001$) or third instance of a given object (B = 0.46, $p < 0.001$). Participants also modulated their communicative behavior on the basis of the utility manipulation and our partner exposure manipulation. Speakers in the Talk is Cheap condition produced significantly more labels than participants in the Talk is Less Cheap condition (B = -0.84, $p < 0.001$). Speakers did more labeling with more knowledable partners; compared with the listener with no exposure, there were significantly higher rates of labeling in the same exposure (B = 1.74, $p < 0.0001$) and double exposure conditions (B = 3.13, $p < 0.001$).

Thus, participants are sensitive to our manipulations, altering their choices about how to communicate with their partner on the basis of the degree of training, and the imposed utilities. Figure 3 illustrates the gesture-speech tradeoff pattern in the double exposure condition (where the listener has more exposure than the speaker, as there was minimal teaching in that condition and thus the speech-gesture trade-off is most interpretable). The effects on rates of gesture mirror those found for speaking and are thus not included for brevity ($ps < 0.01$). Note that these effects cannot be explained by the degree of participant knowledge; all patterns above hold when looking *only* at words known by the speaker at pretest ($ps < 0.01$).

Emergence of Teaching. In line with our hypotheses, a mixed effects logistic regression predicting whether or not teaching occurred on a given trial revealed that teaching rates

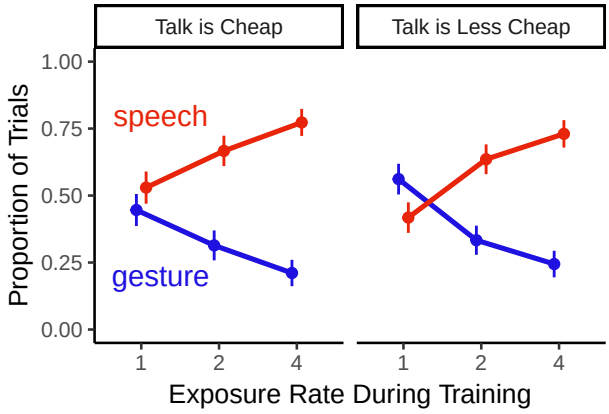

Figure 3: Speakers' communicative method choice as a function of exposure and the utility manipulation. Note that these data are taken only from the condition Double Exposure condition (where listeners had more exposure to the lexicon than speakers). Rates of teaching were minimal and are not plotted here.

across conditions depend on all of the same factors that predict speech and gesture. There was a significant effect of initial exposure to the mapping on the rates of teaching, such that more exposures to a word predicted higher rates of teaching behavior (B = 0.07, $p < 0.01$). There was also a significant effect of the utility manipulation such that being in the Talk is Cheap condition predicted higher rates of teaching than being in the Talk is Less Cheap condition (B = -0.96, $p < 0.01$), a rational response considering teaching allows one to use a less costly strategy in the future and that strategy is especially superior in the Talk is Cheap condition.

Critically, we found an effect for partner's exposure for teaching as well: participants were significantly more likely to teach a partner with no prior exposure to the language than a partner with the same amount of exposure as the speaker (B = -1.63, $p < 0.001$) or double their exposure (B = -3.51, $p < 0.001$).

The planned utility of teaching comes from using another, cheaper strategy (speech) on later trials, thus the expected utility of teaching should decrease when there are fewer subsequent trials for that object, predictng that teaching rates should drop dramatically across trials for a given object. Compared with the first trial for an object, speakers were significantly less likely to teach on the second trial (B = -0.84, $p < 0.001$) or third trial (B = -1.67, $p < 0.001$).

Discussion

In line with our predictions, the data from our paradigm corroborate our findings from the corpus analysis, demonstrating that pedagogical-like 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

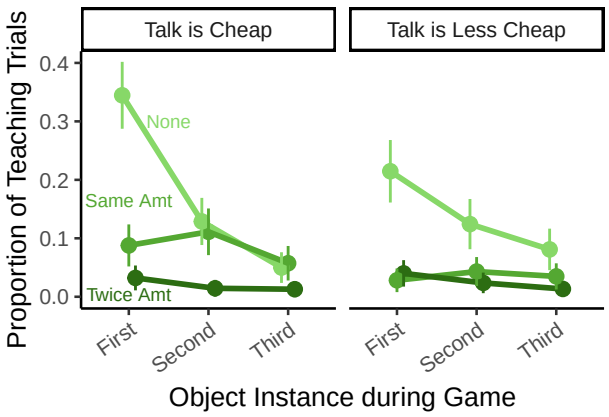

Figure 4: Rates of teaching across the partner exposure manipulation and appearances taken from speaker behavior during gameplay. Both the Talk is Cheap and Talk is Less Cheap condition yielded 30 points for correctly teaching a label, but the relative benefits of later labeling differed.

proof of concept that the sturctured and tuned language input we see in those data could arise from a pressure to communicate. The paradigm's clear, quantitative predictions also allow us to build a formal model to predict our empirical results.

Model: Communication as planning under uncertainty

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.

We take as inspiration the idea that communication is a kind of action–e.g. talking is a speech act (Austin, 1975). Consequently, we can understand the choice of *which communicative act* a speaker should take as a question of which act would maximize their utility: achieving successful communication while minimizing their cost (M. C. Frank & Goodman, 2012) **(for some reason this keeps rendering as "M C Frank & Goodman", instead of just "Frank & Goodman". In this game, speakers can take three actions: talking, pointing, or teaching. In this reference games, these Utilities (U) are given directly by the rules. Because communication is a repeated game, people should take actions that maximize their Expected Utility (EU) over the course of not just this act, but all future communicative acts with the same conversational partner. We can think of communication, then as a case of recursive planning. However, people do not have perfect knowledge of each-other's vocabularies (v). Instead, they only have uncertain beliefs (b) about these vocabularies that combine their expectations about what kinds of words people with as much linguistic experience as their partner are likely to know with their observations of their partner's behavior in past communicative interactions. This makes communication a kind of planning under uncertainty well modeled as a Partially Observable Markov Decision Process (POMDP, Kaelbling, Littman, & Cassandra, 1998).

Optimal planning in a POMDP involves a cycle of four phases: (1) Plan, (2) Act, (3) Observe, (4) Update beliefs. 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 people care about success now compared to success in the future. In our simulations, we set $\gamma = .5$ in line with prior work. 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 ($\mathbb{E}_{v\sim b}$).

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

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 (M. C. 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. For the results reported here, we set $\alpha = 2$ based on hand-tuning, but other values produce similar results.

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

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

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

The critical feature of a repeated communication game is that people can change their partner's vocabulary. In teaching, people pay the cost of both talking and pointing together, but can leverage their partner's new knowledge on future trials. **Note here that teaching has an upfront cost and the only benefit to be gained comes from using less costly communication modes later. There is no pedagocical goal– the**

model treats speakers as selfish agents aiming to maximize their own utilitie by communicating sucessfully. We assume for simplicity that learning is approximated by a simple Binomial learning model. If someone encounters a word w in an unambiguous context (e.g. teaching), they add it to their vocabulary with probability p. We also assume that over the course of this short game that people do not forget–words that enter the vocabulary never leave, and that no learning happens by inference from mutual exclusivity.

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

The final detail is to specify how people estimate their partner's learning rate (p) 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 p is the rate that maximizes the probability of them having learned their initial vocabularies from the trials they observed. People can then expect their partner to have a similar p (per the "like me" hypothesis, Meltzoff, 2005). Having an estimate of their partner's p, they can estimate their vocabulary by simulating their learning from the number of learning trials we told them their partner had before the start of the game.

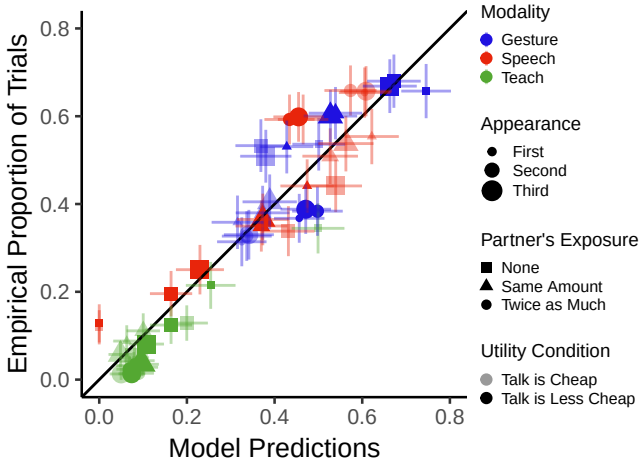

Figure 5: Fit between model predictions and empirical data.

```
    Pearson's product-moment correlation

data:  model_fit$mean_emp and model_fit$mean_pred_plot
t = 20, df = 50, p-value <2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.90 0.97
sample estimates:
 cor
0.94
```

Model Results

The fit between our model's predictions and our empirical data from our reference game study on Amazon Turk can be seen in Figure 5. The model outputs trial-level action predictions (e.g. "speak") for every speaker in our empirical data. These model outputs were aggragated across the same factors as the empirical data: modality, appearance, partner's exposure, and utility condition. We see a signficant correlation of our model predictions and our empirical data (*r = 0.94, p<0.0001*). Our model provides a strong fit for these data, supporting our conclusion that richly-structured language input could emerge from in-the-moment pressure to communicate, without a goal to teach.

General Discussion

We showed that people tune their communicative choices to varying cost and reward structures, and also critically to their partner's linguistic knowledge–teaching when partners are unlikely to know language and many more rounds remain. These data are consistent with the patterns shown in our corpus analysis of parent referential communication and demonstrate that such pedagogically-supportive input could arise from purely selfish motives to maximize the utility of communicating successfully while minimizing the cost of communication. We take these results as a proof of concept that both the features of child-directed speech that support learning as well as those that inhibit it may arise from a single unifying goal: The desire to communicative efficiently.

Of course, this work is limited in a number of important ways. While our task allows us to distill key pressures and manipulations, there are signficant divergences from the corpus data-collecting environment as a result. For one, there are fewer and less dynamic, reciporical interactions between interlocutors and no time-constraints are imposed on communication. Also, parents are likely aware of not just how much language their children know, but *what* their children know (at least at a lexical level). Critical next steps in this line of work will return back to the level of parent-child interaction.

Our framework is not specific to any particular language phenonmenon, though we have focused on gesture-speech coocurrence here. Given the right data or paradigm, our account should hold equally well when explaining how other information-rich language input might arise in naturalistic communication. Using corpus and experimental data, this work validates the hypothesis that communicative pressure alone can lead to supportive language input, which should have important downstream consequences for language learning.

References

Austin, J. L. (1975). *How to do things with words* (Vol. 88).

Oxford university press.

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.

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

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.

Frank, M. C., & Goodman, N. D. (2012). Predicting Pragmatic Reasoning in Language Games. *Science*, *336*(6084), 998–998.

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.

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*, 99–134.

Luce, R. D. (1959). Individual choice behavior.

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.

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.

Meltzoff, A. N. (2005). Imitation and other minds: The "like me" hypothesis. *Perspectives on Imitation: From Neuroscience to Social Science*, *2*, 55–77.

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 noneffects of maternal speech style. In C. A. Ferguson (Ed.), *Talking to children language input and interaction* (pp. 109–149). Cambridge University Press.

Saffran, J. R., & 2003. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110–114.

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 . . .* , *9*, 25–49.

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.

Yurovsky, D. (2017). A communicative approach to early word learning. *New Ideas in Psychology*, 1–7.

Yurovsky, D., Doyle, G., & Frank, M. C. (2016). Linguistic input is tuned to childrens developmental level. In *Proceedings of the annual meeting of the cognitive science society* (pp. 2093–2098).