Parents fine-tune their speech to children's vocabulary knowledge (SOM-R)

XXXXX, XXXXX, and XXXXX

S1 Estimating ages of acquisition for animal words

In designing the stimuli for our experiment, our goal was to use a set of target animals that varied in their average age of acquisition (AoA). To do this, we used two sources of information: (1) Concurrent parent-report estimates of children's vocabularies (Wordbank; Frank, Braginsky, Yurovsky, & Marchman, 2017), and (2) Retrospective self-report estimates from a large group of adults on Amazon Mechanical Turk (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012).

Wordbank is a large and growing repository of administrations of the MacArthur-Bates Communicative Development Inventory (CDI; Fenson et al., 2007)–a checklist of words and other items administered to parents in order to estimate their child's vocabulary. Because Wordbank contains a mixture of cross-sectional and longitudinal data, and we wanted to ensure independence of data across measurements, we used only the first administration for each American English-learning child in the database, yielding 4706 children. For each animal word, we fit a separate robust general linear model, estimating the proportion of children whose parents reported their producing the word from eight to 30 months (including data from both the Words and Gestures and Words and Sentences forms). Each word's normative age of acquisition was defined to be the first month of age at which 50% or more children were estimated to know the animal.

Because only a subset of the animals in the Rossion & Pourtois (2004) image set are included on the MacArthur-Bates Child Development Inventory, and thus available in Wordbank, we also used adult self-report norms from Kuperman et al. (2012) to derive estimates for the remaining animals. Typically, adult self-report estimates of age of acquisition are highly correlated with parent-report estimates, and they were for the 30 animals in both data sources (r = 0.8, t = 7.11, $p < .001$). However, self-report estimates were made on a 1-7 Likert scale rather than on the scale of months.

In order to estimate the ages of acquisition for animals missing from Wordbank, the fit a a general linear model estimating Wordbank age of acquisition from Kuperman et al. (2012) age of acquisition for all animals in both sets (`Wordbank ~ Kuperman + 1`). We then used this model to scale age of acquisitions for the 23 animals in the Kuperman et al. (2012) set missing from Wordbank. Table S1 shows the final estimated ages of acquisition for each animal in the Rossion & Pourtois (2004) set as estimated from Wordbank, Kuperman et al. (2012), and our regression model. For comparison, Figure S1 shows the proportion of parents of 2-2.5-year-olds in our study who reported that their child knew each of the tested animals. These proportions were highly correlated with the model-predicted ages of acquisition (r = -0.94, t = -10.3, $p < .001$).

Table S1: Estimated age of acqusition (AoA) for each animal in months.

animal	Wordbank	Kuperman	model estimate	AoA
alligator	26.00	57.36	23.76	26.00
ant	25.00	51.84	22.40	25.00
bear	20.00	42.96	20.20	20.00
bee	22.00	60.00	24.42	22.00
beetle		63.84	25.37	25.37
bird	18.00	42.24	20.02	18.00
butterfly	23.00	44.04	20.47	23.00
camel		61.32	24.74	24.74
cat	18.00	44.16	20.50	18.00
caterpillar		62.04	24.92	24.92
chicken	23.00	39.12	19.25	23.00
cow	20.00	47.28	21.27	20.00
deer	27.00	62.04	24.92	27.00
dog	15.00	33.60	17.88	15.00
donkey	29.00	72.00	27.39	29.00
duck	18.00	42.00	19.96	18.00
eagle		69.96	26.88	26.88
elephant	23.00	57.60	23.82	23.00
fish	19.00	48.60	21.60	19.00
fly		36.60	18.62	18.62
fox		60.21	24.47	24.47
frog	22.00	51.84	22.40	22.00
giraffe	25.00	60.00	24.42	25.00
goat		62.52	25.04	25.04
gorilla		68.88	26.62	26.62
grasshopper		69.36	26.73	26.73
horse	21.00	49.80	21.89	21.00
kangaroo		66.60	26.05	26.05
leopard		82.08	29.88	29.88
lion	23.00	53.04	22.69	23.00
lobster		89.28	31.67	31.67
monkey	22.00	50.52	22.07	22.00
mouse	23.00	59.28	24.24	23.00
ostrich		77.04	28.64	28.64
owl	24.00	74.52	28.01	24.00
peacock		69.16	26.69	26.69
penguin	27.00	68.16	26.44	27.00
pig	21.00	46.08	20.97	21.00
rabbit	21.00	47.28	21.27	21.00
raccoon		81.48	29.74	29.74
rhinoceros		72.00	27.39	27.39
rooster	28.00	76.92	28.61	28.00
seahorse		69.96	26.88	26.88
seal		65.04	25.67	25.67
sheep	23.00	51.00	22.19	23.00
skunk		63.84	25.37	25.37
snail		69.48	26.76	26.76
snake		61.20	24.71	24.71
spider		41.16	19.75	19.75
squirrel	25.00	53.28	22.75	25.00
swan		75.81	28.33	28.33
tiger	24.00	48.00	21.45	24.00
turtle	23.00	50.04	21.95	23.00
zebra	25.00	57.48	23.79	25.00

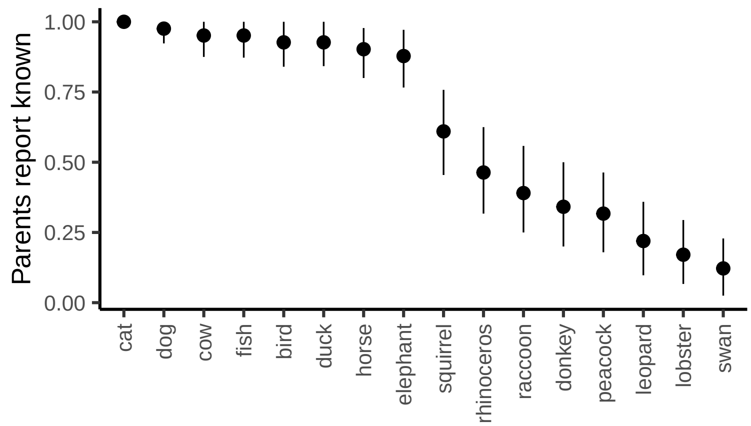

Figure S1: Proportion of parents who reported that their child knew the canonical word for each target animal. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

S2 Model Details

For readability, the main text includes only the key effects for each statistical model rather than a full specification. We include those here. In all cases, we began with the maximal model justified by the design. If this model did not converge, we removed effects iteratively beginning with interactions. We always prioritized random slopes of theoretical importance (e.g. random slopes of word knowledge for each participant) over control variables. Each model included at least a random intercept for each subject and item. Models were estimated using version 1.1-23 of the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015).

S2.1 Target animal difficulty

To validate that parents were more likely to say that their children knew early age of acquisition animals than late age of acquisition animals, we fit a mixed-effects model predicting parents' judgments from *a priori* early and late categories (see paper). The model specification and output are shown in Table S2.

Table S2: Late age of acquisition (AoA) animals were less likely to be known. Model was specified as `understands` $\sim$ `type + (1 | subj) + (1 | animal)`

term	estimate	std.error	z-value	p-value
intercept	5.00	0.74	6.75	$< .001$
type late	-6.48	0.84	-7.74	$< .001$

S2.2 Selection accuracy

To confirm that parent-child dyads communicated successfully in the reference game, we analyzed children's choices. We fit 2 models. First, we asked whether children selected the target animal on each trial above chance levels (33%). To do this we fit a mixed-effects model in which the only fixed effect was an intercept, and we used an offset of $log(\frac{1}{3})$ so that an intercept different from zero would indicate above chance performance. The results of this model are presented in Table S3. We then repeated the same analysis separately for animals that parents judged that their children knew, and animals that they judged that their children did not (Table S4).

Table S3: Overall accuracy on each trial. Model specified as `correct ∼ 1 + offset(log(1/3)) + (1 | subj) + (1 | animal)`

term	estimate	std.error	z-value	p-value
intercept	2.07	0.20	10.35	$< .001$

Table S4: Accuracy for known and unknown animals. Models specified as `correct ∼ 1 + offset(log(1/3)) + (1 | subj) + (1 | animal)` separately for known and unknown animals

term	estimate	std.error	z-value	p-value
known intercept	2.61	0.24	10.93	$< .001$
unknown intercept	1.23	0.15	8.35	$< .001$

After confirming that parents were communicating successfully overall, we asked what predicted children's success at picking the correct animal on each trial. We predicted success on each trial from the number of words in the child's vocabulary (as estimated by the pre-experiment survey), the (log) length of parents' referring expressions, whether parents believed their child knew the target animal, and the interaction between length and whether the animal was known. Children with larger vocabularies were more accurate, children were more accurate for known animals, and longer utterances lead to lower success for known animals and higher success for unknown animals (Table S5).

Table S5: Predicting accuracy on each trial from referring expressions. Model specified as `correct ∼ log(length) · unknown + scaled(vocab) + offset(log(1/3)) + (1 | subj) + (1 | animal)`

term	estimate	std.error	z-value	p-value
intercept	3.14	0.30	10.36	$< .001$
(log) length	-0.40	0.15	-2.69	.007
unknown animal	-1.86	0.41	-4.55	$< .001$
scaled(vocab size)	0.40	0.13	3.19	.001
(log) length · unknown animal	0.46	0.20	2.24	.025

S3 Tuning

Our key analyses concerned the relationship between the length of parents' referring expressions and their children's lexical knowledge. If parents tune the information in their utterances to children's language knowledge, they should produce longer referring expressions for unknown animals. To test this, we fit a model predicting the (log) length of parents' referential expressions on each trial from the child's vocabulary, the proportion of parents who reported that their child knew the target animal, whether the parent reported that their individual child knew the target animal, whether this was the first or second appearance of the target, and the interaction of appearance and the child's target animal knowledge. We found that both the proportion of all children who knew the animal and the parent's belief about their individual child's knowledge affected the length of parents' referring expressions. However, the effect of the individual child's knowledge was reduced on the second appearance of each animal (Table S6).

Table S6: Predicting length of referring expressions. Model specified as `log(length)` $\sim$ `appearance` $\cdot$ `known + prop. know + scaled(vocab) + (known | subj) + (appearance | animal)`

term	estimate	std.error	z-value	df	p-value
intercept	1.81	0.07	25.37	43.17	$<$ .001
second appearance	-0.08	0.04	-2.18	16.74	.044
known animal	0.25	0.08	3.21	43.04	.003
prop. known	-0.17	0.07	-2.32	15.30	.034
scaled(vocab size)	-0.02	0.04	-0.54	40.22	.595
second $\cdot$ known	-0.14	0.02	-6.17	5,294.43	$<$ .001

One possible explanation for the smaller effect of parents' a priori beliefs about the child's knowledge on the second appearance of each animal is that they gathered information from its first appearance. To test this hypothesis, we analyzed the second appearance of each animal, predicting (log) length of referring expressions from accuracy on the first appearance, the parents' belief about the child's knowledge of that animal, and the interaction. We found that parents who thought their child knew an animal, but discovered that they did not, used a longer referring expressions in its second appearance (Table S7).

Table S7: Predicting length on an animal's second appearance. `log(length)` $\sim$ `first accuracy` $\cdot$ `known + (known | subj) + (1 | animal)`

term	estimate	std.error	z-value	df	p-value
intercept	1.84	0.08	21.86	59.68	$<$ .001
first accuracy	-0.15	0.09	-1.69	600.46	.091
known animal	-0.35	0.10	-3.57	57.82	.001
first accuracy $\cdot$ known animal	0.48	0.12	3.98	615.91	$<$ .001

S4 Content of referring expressions

In our primary analyses, we focused on the length of parents' referring expressions as a theory-agnostic proxy for the amount of information in them. To assess how the content of these utterances changed in accord with parents' estimates of their children's animal knowledge, we manually coded utterances for the following features: (1) Use of the animal's canonical label (e.g., "leopard"), (2) Use of a descriptor (e.g., "spotted"), (3) Use of a comparison (e.g., "like a cat"), (4) Use of a subordinate category label (e.g., "Limelight Larry" for peacock), and (5) Use of a superordinate level category label (e.g., "bird" for peacock). Because the rates of usage of each of these kinds of reference varied widely, we fit a logistic mixed effects model separately for each reference kind, estimating whether it would be used on each trial from whether the parent thought their child knew the animal (Table S8). We also coded two other features: use of anaphora (e.g. "the spotted one") and use of animal sounds (e.g. "moo"). However, these were so rare that they could not be analyzed quantitatively.

Table S8: Qualitative analysis of referential expressions. Models were specified as `usage ~ unknown animal + (1 | subj) + (1 | animal)`

measure	term	estimate	std.error	z-value	p-value
descriptor	intercept	-6.18	0.67	-9.19	$< .001$
descriptor	unknown animal	3.09	0.59	5.23	$< .001$
canonical name	intercept	4.44	0.53	8.31	$< .001$
canonical name	unknown animal	0.43	0.34	1.24	.216
comparison	intercept	-7.69	1.11	-6.91	$< .001$
comparison	unknown animal	2.29	0.72	3.19	.001
subordinate	intercept	-6.70	1.74	-3.86	$< .001$
subordinate	unknown animal	-2.19	0.98	-2.24	.025
suoperordinate	intercept	-12.45	2.93	-4.26	$< .001$
suoperordinate	unknown animal	3.01	1.35	2.23	.026

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & others. (2007). *MacArthur-bates communicative development inventories: User's guide and technical manual*. Baltimore, MD: Brookes.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods*, *44*(4), 978–990.

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.