When do children first exhibit adult-like grammatical productivity, supported by abstract syntactic categories? Previous work has come to contradictory conclusions, ranging from full productivity in earliest speech samples [1,2] to gradual development from isolated phrases [3,4]. The English indefinite and definite determiners—'a' and 'the'—have emerged as a particular focus of attention in that their use indexes early productive knowledge of the noun category. We develop here a method for assessing the degree of grammatical productivity implicit in a child's use of determiners. We treat grammatical productivity as a continuous variable rather than a binary, allowing us to trace fine-grained input- and age-related changes in productivity. Our method furthermore accounts for the sparsity of noun frequency distributions, and explicitly accounts for the caregiver determiner uses observed by a child.

To address this question regarding early productivity, we construct a hierarchical beta-binomial model to characterize children's determiner preferences. For a given sample of child speech, we treat the sequence of definite and/or indefinite determiners for each noun as draws from a binomial distribution. Under this model, each noun that the child uses can thus be thought of as a coin with a weight, or determiner preference, indicating how often it should be used with the definite or the indefinite determiner. This determiner preference for a noun reflects two information sources available to the child: i) observed caregiver uses for that noun and ii) information about other nouns. The second information source—generative knowledge—is reflected in the beta prior in our model. The concentration parameter, a concise indicator of the degree to which determiner preference is shared across nouns, is the critical target of inference in that it reflects the salience of a NP -> Det+N rule in early language use.

By fitting the model at different time periods (e.g. comparing the first vs. the second half of a child's tokens from a dataset, or sliding a fixed-size window of tokens across the corpus), we can assess this concentration parameter at different points in developmental time. We fit this model to child—as well as caregiver—determiner uses from 27 children across nine longitudinal developmental corpora. 26 of these are from public-domain CHILDES datasets for US and UK English-speaking children. The 27th child is from Speechome, a new ultra-dense corpus with many more tokens around the onset of combinatorial speech. The Speechome dataset also has an unprecedented density of transcribed parental speech—providing a far more detailed picture of the language environment of the child in this early phase.

Children's earliest uses of determiners show little evidence of syntactic information shared across nouns (esp. Speechome, Naima from the Providence Corpus, and Warr and Liz from the Manchester Corpus). However, productivity rapidly increases between 20 and 30 months, followed by a return to moderate levels. We interpret this trajectory as evidence for 'fast abstraction' in development: while children may lack an abstract noun category at the outset, it is nonetheless quickly larned. Furthermore, abstract categories and production rules may be even *more* active for the child language learner than for the adult language user (e.g. 'a sky')—consistent with evidence of overgeneralization in the morphological domain during this same time period.

Word count: 530

[1] Valian V, Solt S, and Stewart J. Abstract categories or limited-scope formulae? The case of children's determiners. *J Child Lang*, 36:743–778, 2009.
[2] Yang C. Ontogeny and phylogeny of language. *Proc Natl Acad Sci USA*, 2013.
[3] Bannard C, Lieven E and Tomasello M. Modeling Children's Early Grammatical Knowledge. *Proc Natl Acad Sci USA*, 2009.
[4] Pine JM, Freudenthal D, Krajewski G, and Gobet F. Do young children have adult-like syntactic categories? Zipf's Law and the case of the determiner. *Cognition*, 127:345–360, 2013.