[1]   Analysing the effect of sibling number on input and output in the first 18 months

²   Abstract

³ Prior research suggests that across a wide range of cognitive, educational, and health-based

⁴ measures, first-born children outperform their later-born peers. Expanding on this literature

⁵ using naturalistic home-recorded data and parental vocabulary report, we find that early

⁶ language outcomes vary by number of siblings in a sample of 43 English-learning U.S.

⁷ children from mid-to-high socioeconomic status homes. More specifically, we find that

⁸ children in our sample with two or more - but not one - older siblings had smaller productive

⁹ vocabularies at 18 months, and heard less input from caregivers across several measures than

¹⁰ their peers with less than two siblings. We discuss implications regarding what infants

¹¹ experience and learn across a range of family sizes in infancy.

¹² *Keywords:* Siblings, Lexical Development, Input Effects, Language Acquisition

13   Analysing the effect of sibling number on input and output in the first 18 months

14   A common simplifying assumption in research on language development is that there is

15 a theoretical "optimum" environment for early language, whereby the input is tailored to a

16 single infant's needs, changing over time as language capacity grows (e.g. Soderstrom, 2007;

17 Stern, Spieker, Barnett, & MacKain, 1983). However, for many infants and for many reasons,

18 language acquisition occurs across diverse social contexts that can influence the learning

19 environment, including the presence of older siblings in the home (Fenson et al., 1994).

20 According to the United States Census Bureau (2010), around one third of children are born

21 into households with at least one other infant present, and one in every five infants is

22 acquiring language in a household shared with two or more other children. Similar statistics

23 are reported for British infants (Office for National Statistics, 2018), where the average

24 household has 1.75 children, and 15% of households have three children or more. In this

25 paper, we consider the role of siblings in the early language environment of English-learning

26 infants. We use naturalistic home-recorded data to measure input in earlier- and later-born

27 infants, alongside their productive vocabulary over the first 18 months of life.

28   Prior research suggests that infants born to households with older children may be

29 slower to learn language. Fenson and colleagues (1994) found that by 30 months of age,

30 children with older siblings performed worse than those with no siblings across

31 parent-reported measures of productive vocabulary, use of word combinations, and mean

32 length of utterance. This 'sibling effect' may be the result of differences in input between

33 first- and later-born children: some research finds that infants with older siblings hear less

34 speech aimed specifically at them, and what they do hear is understood to be linguistically

35 less supportive of early language development (Hoff-Ginsberg, 1998; Oshima-Takane &

36 Robbins, 2003). In contrast, some studies have noted linguistic *advantages* for later-borns,

37 who may have stronger social-communicative skills (Hoff, 2006), better understanding of

38 pronouns (Oshima-Takane, Goodz, & Derevensky, 1996), and better conversational abilities

39 (Dunn & Shatz, 1989). Overall, while the particulars differ across studies, prior work

40 suggests that the presence of siblings in the home leads to differences in infants' early

41 linguistic experiences and skills, though the direction of these differences varies depending on

42 what aspects of language are being measured.

43   Numerous studies have attempted to better understand how siblings affect the

44 language development trajectory, with comparisons of language acquisition across first- and

45 later-borns. Here again, findings are mixed, but overall two general conclusions can be

46 drawn. First, analyses consistently show that infants with older siblings generally have

47 slower *vocabulary development* (Berglund, Eriksson, & Westerlund, 2005; Fenson et al., 1994;

48 Pine, 1995; Zambrana, Ystrom, & Pons, 2012), and this effect increases with number of older

49 siblings (Gurgand et al., 2022; Karwath, Relikowski, & Schmitt, 2014; Peyre et al., 2016).

50 Furthermore, this finding is consistent across cultures (e.g. European French (Gurgand et al.,

51 2022; Havron et al., 2019); Singaporean (Havron et al., 2022); Kenyan (Jakiela, Ozier,

52 Fernald, & Knauer, 2020); and German (Karwath et al., 2014)). However, this finding is not

53 as clear-cut as has been previously assumed: Hoff-Ginsberg (1998) shows first-borns to have

54 better lexical and syntactic skills up until 2;5, but later-born infants had better

55 conversational abilities during the same time-period. Recent studies have also identified

56 effects for age gap between the target child and their siblings (whereby larger age gaps

57 correlate with lower vocabulary scores (Gurgand et al., 2022; Havron et al., 2022)) and

58 sibling sex (whereby older brothers have a negative effect on vocabulary outcomes, but not

59 older sisters (Havron et al., 2019; Jakiela et al., 2020)) though neither of these effects are

60 found consistently across datasets; Havron et al. (2022) and Gurgand et al. (2022) find no

61 effect for sibling sex, whereas Havron et al. (2019) find no effect for age gap. Some of these

62 differences across studies may relate to insufficient power to detect relatively small effects,

63 perhaps leading to under- or over-estimation of effect sizes, or simultaneous contributing

64 factors that are difficult to disentangle.

65   The second general finding pertains to sibling-related differences in the early linguistic

66 *environment*: infants with no siblings receive more input overall, and this more closely

67 reflects what is typically considered to be 'high quality' input in the extant literature

68 (i.e. more input in an infant-directed speech style (Ramírez-Esparza, García-Sierra, & Kuhl,

69 2014); longer utterance length (Barnes, Gutfreund, Satterly, & Wells, 1983); higher lexical

70 diversity (Rowe & Snow, 2020)). Indeed, the very presence of a sibling in the linguistic

71 environment changes the way language is used. When siblings are present (i.e. triadic

72 interactions), mothers' input has been found to be more focused on regulating behaviour, as

73 opposed to the language-focused speech that is common in dyadic contexts (Oshima-Takane

74 & Robbins, 2003). Reports show that the mean length of utterance is longer in the input of

75 first-born infants (Hoff-Ginsberg, 1998; but see also Oshima-Takane & Robbins, 2003 for a

76 comparison of dyadic and triadic contexts), who also hear more questions directed at them

77 than later-borns. Both Jones and Adamson (1987) and Oshima-Takane and Robbins (2003)

78 report no difference between the overall number of word types produced by mothers in

79 dyadic and triadic settings, but the proportion of speech directed at the target infant is

80 drastically reduced when input is shared with siblings.

81   As Hoff (2006) explains, infants with siblings have less experience of speech directed at

82 them, but they do have an advantage over their first-born peers in that they are subject to

83 more overheard speech. This may be an important source of input for infants with one or

84 more older siblings. Akhtar, Jipson and Callanan (2001) show that, by age 2;6, children can

85 learn both novel object labels and novel verbs through overhearing. Slightly younger children

86 (aged 1;11-2;2) were also able to learn the novel object labels, but not verb labels.

87 Two-year-old infants can even learn novel object labels while doing activities that distract

88 them from the language input, and when the novel words are produced non-saliently (Akhtar,

89 2005). This suggests that, while the learning environment for later-borns might differ from

90 that of first-born infants, there may be ample opportunity for them to learn from the speech

91 that surrounds them; namely overheard speech directed at their older sibling(s). Evidence is

92 mainly drawn from work testing infants aged 2 and above (e.g. Akhtar, 2005; Fitch,

Liberman, Luyster, & Arunachalam, 2020; Foushee, Srinivasan, & Xu, 2021), and generally relies on experimental work rather than observations of the home environment. However, Floor and Akhtar (2006) tested younger infants to find that the capacity to learn from overheard speech is available from as early as 16 months, at least in an experimental setting.

There thus may be a trade-off, even in early development, between highly supportive one-to-one input from a caregiver (cf. Ramírez-Esparza et al., 2014) and the potential benefits drawn from communicating with (or overhearing communication with) a sibling. In the present study, we test the extent to which having more versus fewer siblings in the home environment may affect the linguistic environment in ways that could lead to differences in vocabulary development over the course of the first 18 months of life. In analyzing infants' growing productive vocabulary and linguistic environment in relation to the presence of older siblings in their household, the present work expands on the extant literature in two key ways. First, much of the existing literature identifying links between sibling number and vocabulary outcomes draws on large-scale questionnaire data, rather than naturalistic day-to-day interactions in the home. In contrast, we analyze an existing corpus of home recordings in concert with vocabulary checklists, in order to capture the reality of the early linguistic environment and how this is affected by sibling number. Second, we consider the opportunities that overheard speech might present in the infant's linguistic environment. We examine the effect of sibling number on overall amount of input produced in our naturalistic recordings, as well as, crucially, the extent to which parents label objects being attended to by the infant (*object presence*[1]). The analysis of object presence will allow us to gain insight into the kinds of learning opportunities being presented to infants in the early input, based on the previous research showing that object labeling - even when not directed specifically at the target child - can be a valuable source for acquiring linguistic knowledge. Based on work

---

[1] We've retained the term object presence for continuity with prior work using this variable but note that what this variable captures isn't merely whether the object was present but rather whether it was present when the word for it was said aloud

117 summarized above, we expect that both the language environment and infants' early

118 productive vocabulary will vary as a function of how many older siblings they have.

119 **Hypotheses**

120   Synthesising the work above in broad strokes, given prior research showing that early

121 lexical development is more advanced among first-born infants (e.g. Hoff-Ginsberg, 1998), we

122 predict that children with more siblings will have lower productive vocabularies than their

123 peers with fewer siblings. However, we have no *a priori* predictions about how these

124 differences will manifest gradiently (e.g. linear decrease for each additional sibling, a

125 threshold effect where we see a drop after a certain sibship size, etc.). With regard to the

126 infants' linguistic environment, we hypothesise that infants with more siblings will

127 experience lower prevalence of two aspects of the language input previously shown to support

128 language development: **amount of input** and **amount of object presence.** Just as for

129 productive vocabulary size, we do not make *a priori* predictions regarding the shape of these

130 effects, beyond predicting a decrease with sibling number.

131   More specifically regarding input, following previous studies that show infants with

132 siblings to receive less speech directed at them (Jones & Adamson, 1987; Oshima-Takane &

133 Robbins, 2003), we expect to see the same effect in our sample. In terms of object presence,

134 as noted above, we mean word and object co-occurrence, e.g. mother saying "cat" when the

135 child is looking at a cat. We predict object presence will decrease as sibling number increases,

136 because as caregivers' attention is drawn away from one-to-one interactions with the infant,

137 there is likely less opportunity for contingent talk and joint attention. Prior research suggests

138 links between object presence and early word learning (Bergelson & Aslin, 2017; Cartmill et

139 al., 2013), though to our knowledge this has not been examined in relation to sibling number.

¹⁴⁰ **Methods**

¹⁴¹   We analyze data from the SEEDLingS corpus, a longitudinal set of data incorporating

¹⁴² home recordings, parental reports and experimental studies from the ages of 0;6 to 1;6. See

¹⁴³ Bergelson, Amatuni, Dailey, Koorathota, & Tor (2019) for further details on the full set of

¹⁴⁴ home-recorded data and its annotations. The present study draws on the parental report

¹⁴⁵ data to index child vocabulary size, and annotations of hour-long home video recordings,

¹⁴⁶ taken on a monthly basis during data collection, to index input.[2] We note at the outset that

¹⁴⁷ with such a multidimensional dataset there are always alternative ways of conducting

¹⁴⁸ analyses of input and output; due to limited power in our sample, we are unable to consider

¹⁴⁹ all potential contributing variables (e.g. the same dataset was analysed in a previous study

¹⁵⁰ and found that mothers' work schedules were associated with vocabulary knowledge at 17

¹⁵¹ months (Laing & Bergelson, 2019); we do not analyse that variable further here due to our

¹⁵² limited sample size). Our goal here is to make motivated decisions that we clearly describe,

¹⁵³ provide some alternative analytic choices in the supplementals, and to share the data with

¹⁵⁴ readers such that they are free to evaluate alternative approaches.

¹⁵⁵ **Participants**

¹⁵⁶   Forty-four families in New York State completed the year-long study. Infants (21

¹⁵⁷ females) were from largely middle-class households; 33 mothers had attained a B.A. degree

¹⁵⁸ or higher. Based on parental report, no infants had speech- or hearing-relevant diagnoses;

¹⁵⁹ none were low birth weight (all >2,500g); 42 were white, two were from multi-racial

¹⁶⁰ backgrounds. All infants heard >75% English on a regular basis and lived in two parent

¹⁶¹ homes. Two participants were dizygotic twins; we retain one twin in the current sample,

---

[2] We also ran our input analysis using data sub-sampled from day-long audio recordings taken on a different day from the video data reported below; results were consistent with those outlined below for most analyses (see Supplementary Materials, S1).

162 considering the other only as a sibling[3]. Thus our final sample size was 43 infants.

163 **Sibling Details.**   Sibling number was computed based on parental report in the

164 demographics questionnaires completed at 0;6 (Sibling number range: 0-4). Siblings were on

165 average 4.11 years older than the infants in this study (SD: 4.01 years, R: 0-17 years).[4] All

166 siblings lived in the household with the infant full time, apart from one infant who had two

167 older half siblings (and no other full siblings) who lived with their other parent part of the

168 time. Both older siblings were present for at least some of the monthly recordings. One

169 family had a foster child live in the home for 2 months of data collection, who is not

170 accounted for in our data; the target infant had one sibling. All siblings were older than or of

171 the same age as the infant in question.

Materials

172 **Materials**

173 **Parental report data.**   To index each child's language abilities, we draw on data

174 from vocabulary checklists (MacArthur-Bates Communicative Development Inventory,

175 hereafter CDI, Fenson et al. 1994), administered monthly from 0;6 to 1;6, along with a

176 demographics questionnaire; each month's CDI survey came pre-populated with the previous

177 month's answers to save on reduplicated effort. Because the majority of infants did not

178 produce their first word until around 0;11 according to CDI reports (M=10.67, SD=2.23), we

179 use CDI data from 0;10 onwards in our analysis. CDI production data for each month is

180 taken as a measure of the infants' lexical development. CDI data for production has been

181 well validated by prior work, including work in this sample (Frank, Braginsky, Yurovsky, &

182 Marchman, 2021; Moore, Dailey, Garrison, Amatuni, & Bergelson, 2019). Of the intended 13

183 CDIs collected for each of the 43 infants, 26 were missing across 11 infants (leaving 559 CDIs

---

[3] Results were consistent when both twins were removed from the dataset, see S2.

[4] For six infants, siblings' exact birthdates were not provided, and so age difference was estimated by subtracting the infant's age (6 months) from the sibling's age in years, as listed on the questionnaire (e.g. if a sibling was 5 years old, they were classed as being 4.5 years older than the infant).

184 in total). 4 infants had 4 CDI data-points missing, while the majority (n = 5) had only one

185 missing data-point.

186   **Home-recorded video data.**   Every month between 0;6 and 1;5, infants were

187 video-recorded for one hour in their home, capturing a naturalistic representation of each

188 infant's day-to-day input. We did not ask families to ensure certain family members were or

189 were not present; our video recordings capture whoever was home at the time families opted

190 to schedule. Here we draw on data from the two caregivers who produced the most words in

191 each recording; in 87% of cases this was the mother, and 8% of cases the father. Fathers

192 produced the second highest number of words in 50% of cases (see S3 for a full breakdown of

193 speakers classed as caregivers in the dataset). At the child level, the modal caretaker across

194 the 12 videos was the mother for 37 infants, father for 4 infants and grandmother for the

195 remaining 2 infants. Infants wore a hat with two small Looxcie video cameras attached, one

196 pointed slightly up, and one pointed slightly down; this captured the scene from the infants'

197 perspective. In the event that infants refused to wear the hats, caregivers wore the same

198 kind of camera on a headband. Additionally, a camcorder on a tripod was set up in the room

199 where infants and caretakers were interacting to capture a broader view; families were asked

200 to move this camcorder if they changed rooms. The dataset includes 12 videos for each child,

201 one for each month that we analyzed.

202   Object words (i.e. concrete nouns) deemed to be said to, by, or loudly and clearly near

203 the target child were annotated by trained coders for several properties of interest to the

204 broader project on noun learning. Here we examine annotations for speaker, i.e. who

205 produced each noun, and object presence, i.e. whether the noun's referent was present and

206 attended to by the infant (see "Derived Input Measures" below).

207   ***Derived Input measures.***   Two input measures were derived based on the

208 individual word level annotations of concrete nouns directed to or near the target child in

209 this corpus, each pertaining to an aspect of the input that is established as important in

210 early language learning: **overall household input** (how many concrete nouns does each

211 infant hear? Note that this measure only includes speech produced directly to or close by the

212 target child; see example below) and **object presence** (what proportion of this input is

213 referentially transparent?), detailed below. Neither of these measures are, in our view,

214 interpretable as "pure" quality or quantity input measures; we hold that quality and quantity

215 are inextricably linked in general, and specifically we include (by design) only object words

216 that the recordings suggest were possible learning instances for the infants who heard them,

217 wherein quantity and quality are conflated. This included only concrete, imageable words

218 that were addressed directly to the child (e.g. "Have you got your toy *bear*?"), or sufficiently

219 loud and proximal that they were clearly audible to the child (e.g. "Can you pass me the toy

220 *bear*?", directed at the sibling while mother, infant, and sibling play on the rug). As

221 mentioned above, only speech produced in the infant's immediate surroundings (i.e. speech

222 that would have been clearly heard by the target infant) was coded.

223 *Household Input* reflects how many nouns infants heard in the recordings from their

224 two main caregivers (operationalized as the two adults who produced the most nouns in each

225 recording; see above) and (where relevant) siblings. Input from speakers other than these

226 two caregivers was relatively rare during video recordings, accounting for 0.39% of input

227 overall (SD=2.37%), and is excluded from our analysis. This measure of the early language

228 environment is based on evidence showing strong links between the amount of speech heard

229 in the early input and later vocabulary size (Anderson, Graham, Prime, Jenkins, & Madigan,

230 2021). This analysis considers only nouns produced by speakers in the child's environment,

231 directed to or produced clearly near the child (which is what was annotated in the broader

232 SEEDLingS project); concrete nouns are acquired earlier in development in English and

233 cross-linguistically (Braginsky, Yurovsky, Marchman, & Frank, 2019). As in any sample of

234 naturalistic interaction, the number of nouns correlates highly with the number of words

235 overall (e.g. based on automated analyses of adult word counts vs. manual noun-only

236 annotations, Bulgarelli & Bergelson, 2020). Thus, noun count in the monthly hour of video

237 data serves as our household input proxy.

238 *Object Presence* was coded as "yes", "no", or "unsure" for each object word annotated

239 in the home recordings, as produced by the two main caregivers detailed above, based on

240 trained annotators' assessment of whether the referent of the word (i.e. the object) was

241 present and attended to or touched by the child or the caregiver. For example, if the

242 caregiver was pointing at a ball while the saying the word *ball*, this was coded as "yes". If

243 the infant was holding (but not looking at) a bottle while the caregiver said *bottle*, this

244 would also be coded as "yes". On the other hand, if the caregiver refers to a dog that is

245 barking in the other room, that would be coded as "no", as it was not present during object

246 labeling. In the video data, 182 instances (0.24% on average per infant) of object presence

247 were marked as unsure; these instances were not included in this analysis.

Data analysis

249 While we set out to test the hypotheses outlined above, aspects of our analysis were

250 exploratory in nature. In respect of this, and on the advice of a helpful anonymous reviewer,

251 we focus on descriptive and confirmatory measures of analysis through data visualization and

252 effect size reporting alongside significance testing. For each of key variable tested, we present

253 these three avenues for understanding the data, alongside any further follow-up exploratory

254 analyses, where appropriate.

255 All reported models were generated in R (R Core Team, 2019) using the *lmerTest*

256 package to run linear mixed-effects regression models when needed (Kuznetsova, Brockhoff,

257 & Christensen, 2017). *P*-values were generated by likelihood ratio tests resulting from nested

258 model comparison. All models include infant as a random effect. All post-hoc tests are

259 two-sample, two-tailed Wilcoxon Tests; given that all of our variables of interest (CDI score,

260 household input and object presence) differed significantly from normal by Shapiro tests, we

261 opted to run non-parametric tests for all post-hoc comparisons. Where multiple post-hoc

262 comparisons are run on the same dataset, Bonferroni corrections are applied (e.g. with an

263 adjusted *p*-value threshold of .025 for 2 between-group comparisons). While we have a

264 substantial amount of data for each participant, our limited n means we are under-powered

265 to consider multiple demographic variables simultaneously given the data distribution

266 (e.g. sibling number and sex, see Table 1; as luck would have it both infants with 3 siblings

267 were girls and both with 4 were boys). There were no correlations between sibling number or

268 child word production and maternal age/education. See S4 for further details.

<h1 style="text-align:center">Results</h1>

270 Our analyses consider infants' total productive vocabulary[5] alongside our two input

271 measures – nouns in household input and extent of object presence in the input – as a

272 function of sibling number. Since the raw data are highly skewed, log-transformed data

273 and/or proportions are used for statistical analysis (1 was added to the raw infant

274 production data of all infants before log-transformation to retain infants with vocabularies of

275 0.) Unless otherwise specified, all figures display non-transformed data for interpretive ease.

276 Vocabulary development was highly variable across the 43 infants, according to the

277 CDI data we had available. By 18 months, 2 infants produced no words (taken from 36

278 available CDIs at this time-point), while mean productive vocabulary size was 60.28 words

279 (SD=78.31, Mdn=30.50). Three infants had substantially larger-than-average (3SDs above

280 the monthly mean) vocabularies at certain time-points in the data; we counted one of these

281 infants as an outlier and remove this child's data from the CDI analysis given that their

282 vocabulary was higher for multiple consecutive months (1;1-1;6). The other two infants had

283 higher vocabularies at 10-11 months only (when variance was quite limited, see Figure 1),

284 and were retained to maximize data inclusion. This left 42 infants (19 females) in the

285 analysis of vocabulary size. Infants had one sibling on average (M=0.86, Mdn=1, SD=1.10).

286 See Table 1.

---

[5] While in principle we could have just used noun productive vocabulary, in practice noun and total

vocabulary is correlated >.95 in this age range; we opted to retain the overall total vocabulary, as lexical

class is not a straightforward notion in the early lexicon.

Table 1

Sibling number by female and male infants (n=42). One child was an outlier, and was removed from the CDI analysis and this table; see text for details.

n Siblings	Female	Male	Total
0	9	12	21
1	6	6	12
2	2	3	5
3	2	0	2
4	0	2	2
Total	19	23	42

Effect of siblings on infants' productive vocabulary

We first modeled the effect of siblings on reported productive vocabulary. We explored three possible variations on how to represent the sibling effect: a binary variable (0 vs. >0 siblings), aggregated groups (None vs. One vs. 2+ siblings), and discrete sibling number (0 to 4 siblings), comparing the following nested model structures, where (1) is the baseline model and (2) includes siblings as the variable of interest.

1. vocabulary size (log-transformed) ~ age (months) + (1|subject)
2. vocabulary size (log-transformed) ~ siblings [binary, group or discrete] + age (months) + (1|subject)

In our sample, simply having siblings (i.e. as a binary variable) did not predict CDI productive vocabulary size, while both discrete sibling number and sibling group did. See

298 Table 2.[6]

Table 2

Output from likelihood ratio tests comparing regression models that predict the effects of sibling number (binary, grouped and discrete variables) on vocabulary size. Month was included in each model as a fixed effect; subject was included as a random effect. R2 values are included to reflect model goodness-of-fit.

Model	Df	Chisq	p value	R2
0 vs. >0 siblings	1.00	2.13	0.14	0.81
Sibling group	2.00	8.00	0.02	0.81
Sibling number	1.00	6.08	0.01	0.81

Note. While we report R2 values for completeness, we note that utility and interpretability of this metric for this model type is debated, see https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html.

---

[6] While our sample size and distribution leaves it statistically questionable to consider both sex and sibling number, for completeness we did also run a model that included sex in addition to age and sibling number (our primary variable of interest). Sex did not improve model fit over and above the effect of siblings in any of the three comparisons (*p*s all >0.54).

Table 3

Full model output from linear mixed effects regression models comparing language development over time in relation to sibling group. Age in months was included as a fixed effect; subject was included as a random effect.

Effect	Estimate	Std. Error	df	t value	p
Intercept	-2.69	0.26	156.59	-10.27	<0.001
SibGroupOne	-0.01	0.30	42.08	-0.05	0.963
SibGroup2+	-0.94	0.33	42.84	-2.81	0.007
month	0.34	0.01	315.13	25.19	<0.001

Having more siblings was associated with a smaller vocabulary size over the course of early development. This is consistent with previous findings (Hoff-Ginsberg, 1998; Pine, 1995). We find that for each additional sibling, infants were reported to have produced 26% fewer words. The 'sibling effect' is thus present in our data.

In terms of our grouped sibling variable (i.e. 0 vs. 1 vs. 2+ siblings), infants with one sibling produced 64 words on average at 18 months, which is, on average, 5 words *more* than their firstborn peers. On the other hand, infants with two or more siblings produced 13 words on average, which is 353% fewer words than the firstborns in our data. See Table 3 and and Figure 1. Post-hoc Wilcoxon Rank Sum tests comparing reported productive vocabulary at 18 months (where there's the widest vocabulary range) revealed significantly larger vocabularies for infants with one sibling compared to those with two or more siblings (W=5, $p = $ .004, CI=[-72.00,-12.00]), but no difference between infants with one sibling and those with no siblings (W=79.50, $p = $ .631, CI=[-34.00,34.00]). See Table 4.
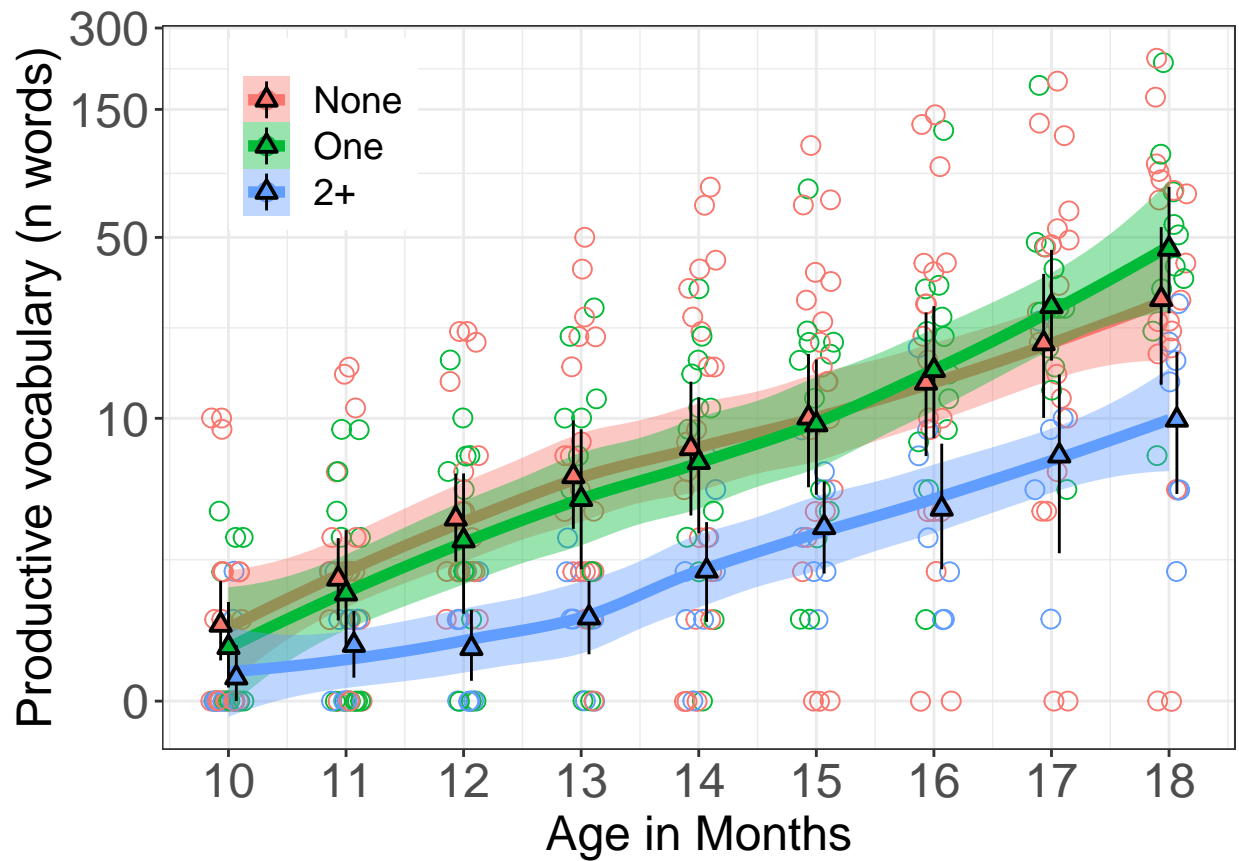
Figure 1. Reported productive vocabulary acquisition (CDI) over time (n=42; one child was an outlier, and was removed from the CDI analysis and this figure; see text for details). Colors denote sibling group; line with colored confidence band reflects local estimator (loess) fit over individual infants' vocabulary at each month. Triangles indicate mean with bootstrapped CIs computed over each month's data. Points (jittered horizontally) show individual infants' vocabulary size at each month. Y-axis utilizes log-transformed vertical spacing for visual clarity.

Effect of siblings on infants' input

312

313   Having established that infants' productive vocabulary varied as a function of sibling

314   number in all but the binary version of the measure (0 vs. >0 siblings), we turn to our input

315   measures to test whether *input* varied by a child's sibling status. For these analyses we

316   report here the group sibling division (0 vs. 1 vs. 2+) as this lets us keep relatively similar

317   Ns across groups, thus making variance more comparable for post-hoc comparisons (the

318   discrete sibling number (0-4) version is reported for completeness in S5; results hold for both

319   input variables). We also now include the child who was a multi-month vocabulary outlier

320   above, given that input and vocabulary are not tested in the same model. One infant of the

321   full sample of 43 infants was an outlier in that they heard substantially more input words

322   and words with object presence than all the other infants in the sample in four of their

323   recording sessions. Given that these sessions were not contiguous, we opted to keep this

324   infant in the analyses reported below, though all results hold when they are removed from

325   our sample (see S6).

326   While we didn't have strong a priori expectations about how overall input or object

327   presence would vary by age or sex, these were included in initial model comparisons to see if

328   they improved fit alongside a random effect of infant. Both variables improved fit for the

329   input model, and only age did for the object presence model. Thus our baseline models

330   include these sets of control variables, respectively. See Table 5 for final model estimates.

331   **Caregiver Input.**   We tested overall quantity of input (aggregated across the two

332   main caregivers, as outlined above, and siblings) in our model alongside age, sex and subject,

333   as noted above, and a significant effect was found for the effect of sibling group ($\chi^2(2) =$

334   8.88, p=.012, R^2=0.59). Infants with one sibling heard on average 9% fewer words than

335   those with no siblings in any given hour-long recording, while infants with two or more

336   siblings heard 81% fewer words.

337   We then ran post-hoc tests to compare mean amount of input across sibling groups;

338   these showed a significant difference in average input received between infants with one

339   sibling versus those with two or more siblings (W=11, p=.002, CI=[-120.87,-38.75];

340   Bonferroni-corrected p-threshold = .025 for all reported Wilcoxon tests) while amount of

341   input did not differ between infants with no siblings and those with one sibling (W=146, p

342   = .736, CI=[-39.50,57.50]). See Table 4 for overall group differences (M and SD) in amount

343   of input.

344   While we operationalized caregiver input in our models as input speech from the two

345   adults who produced the most words in any given session, in 86.73% of cases this was the

346   mother or father. Considering mothers and father specifically, maternal input accounted for

347   75% of object words in the data overall (M=195.49 words, Mdn=162.31, SD=108.93)[7].

348   Fathers accounted for an average of 18% (M=58.72, Mdn=33, SD=65.05), while infants with

349   siblings received around 12% of their input from their brothers and sisters (M=22.97,

350   Mdn=18, SD=18.49). See Table 4 and Figure 2. As well as the overall input being greater

351   for firstborns, compared with infants with one or 2+ siblings, note also that the variance is

352   greater for this group, and decreases as sibling number increases. This is shown in the SDs

353   reported in Table 4, and in the data points visualized in Figure 2.

354   Overall, for infants who had siblings, at least one other child was present in 68% of

355   recordings (n = 188 recordings, SD = 27%). Wilcoxon Rank Sum tests comparing mean

356   monthly input showed no difference between the amount of sibling input received by infants

357   with one sibling compared with those with two or more siblings (W=31, p=.071,

358   CI=[-14.50,2]). Looking at mothers and fathers individually, infants with two or more

359   siblings heard significantly less input from their mothers than those with one sibling (W=5,

360   p< .001, CI=[-124.88,-41.92]), while there was no difference between those with one vs. no

361   siblings (W=125, p = .985, CI=[-48.23,51.50]). Finally, amount of paternal input did not

---

[7] One family in our sample had two mothers; rather than artificially assigning one parent to another
category, we averaged both mothers' input for this child; we acknowledge that this is an imperfect solution
but found it better than the alternatives.

Table 4

Data summary of our two input measures and reported vocabulary size at 18 months.

	No siblings		1 sibling		2+ siblings	
Variable	Mean	SD	Mean	SD	Mean	SD
% object presence in input, 10-17 months	68.61	14.13	55.90	15.72	46.46	16.85
N Input utterances, 10-17 months	213.54	122.98	196.28	81.04	117.67	46.96
Productive Vocabulary 18m (CDI)	58.89	60.76	64.10	61.97	13.00	9.49

362   differ between groups (one vs. none: W=102, p=.393, CI=[-10.21,61.30]; one vs. 2+: W=21,

363   p=.945, CI=[-74.33,56.54]).

364   **Object presence.**   On average, 60% of annotated utterances included a referent that

365   was present and attended to by the infant (Mdn=0.61, SD=0.12). See Table 4. Consistent

366   with our hypothesis that infants with more siblings would hear fewer words in referentially

367   transparent conditions (i.e. they would experience lower object presence) than those with

368   fewer siblings, our models reveal a significant effect for sibling group on object presence

369   ($\chi^2(2) = 27.33$, $p<$ .001, R^2=0.55). See Table 5 and Figure 3.

370   Descriptively, infants with no siblings experienced on average 48% more object

371   presence in their input than those with two or more siblings, and 23% more than those with

372   one sibling. Post-hoc comparisons revealed significant between-group differences: infants

373   with no siblings experienced significantly more object presence than those with one sibling

374   (W=240, $p<$ .001, CI=[0.07,0.20]; Bonferroni-corrected p-threshold = .025). Likewise,

375   infants with one sibling experienced significantly more object presence those with two or

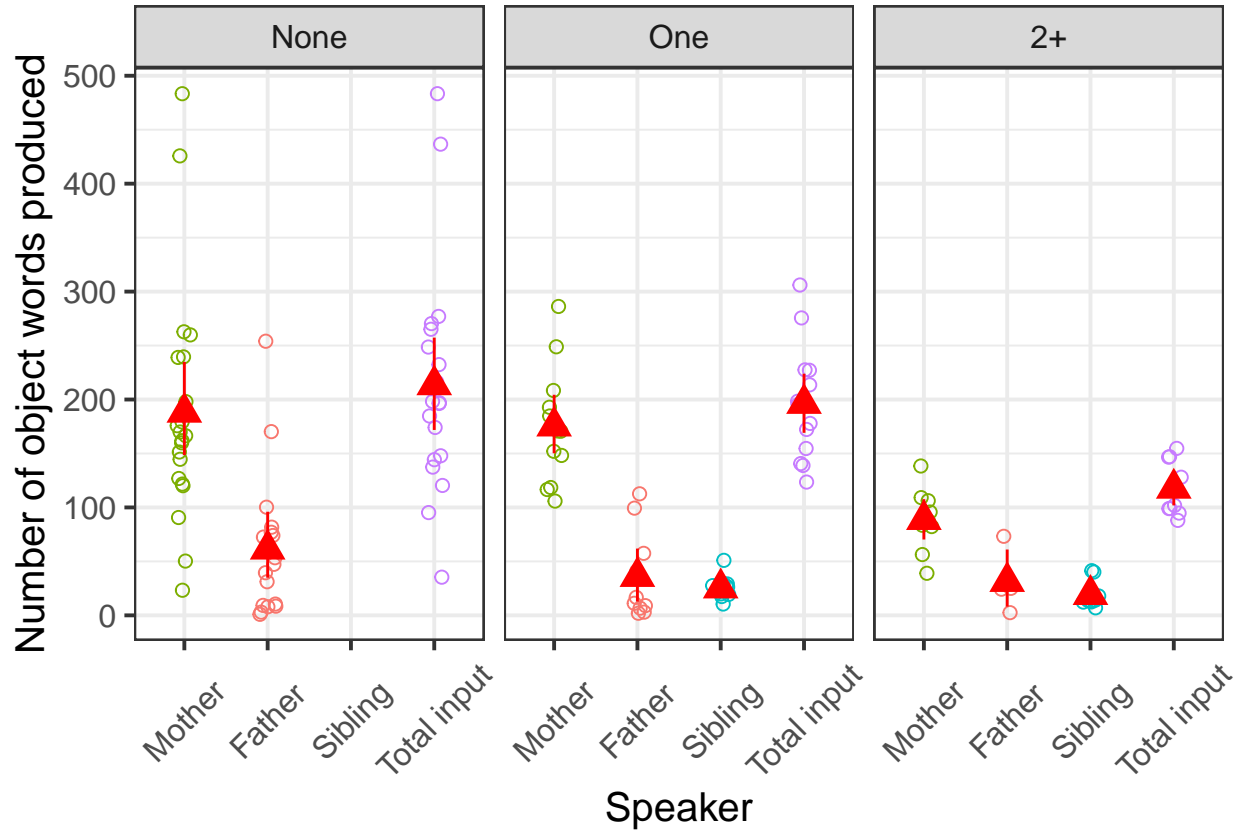376   more siblings (W=25, p=.025, CI=[-0.18,-0.01]).

Figure 2. Mean number of words produced by Mothers, Fathers and Siblings, as well as total family input (mother + father + sibling(s)), across sessions recorded between 10-17 months. Circles represent values for individual infants; red triangles show group means. In the case where the infant had two mothers, mean maternal input is shown. See S7 for a month-by-month visualisation of caregiver input.

Table 5

Full model output from linear mixed effects regression models comparing our two input measures (object words produced in caregiver input and object presence) over time in relation to sibling group. Age in months was included as a fixed effect in both models, sex was included in the caregiver input model only; subject was included as a random effect.

Variable	Effect	Estimate	Std. Error	df	t value	p value
Caregiver input	Intercept	4.87	0.18	182.83	27.57	<0.001

	SibGroupOne	0.01	0.15	43.00	0.04	0.965
	SibGroup2+	-0.49	0.17	43.00	-2.94	0.005
	month	0.03	0.01	301.00	3.02	0.003
	sexM	-0.18	0.13	43.00	-1.38	0.173
Object presence	Intercept	0.57	0.04	320.78	12.72	<0.001
	SibGroupOne	-0.13	0.03	43.00	-3.80	<0.001
	SibGroup2+	-0.22	0.04	43.00	-5.87	<0.001
	month	0.01	0.00	301.00	2.98	0.003

[377] **Sibling presence.** So far, our analysis takes into account the differences in input

[378] based on whether or not the target child has a sibling, but does not directly consider

[379] whether sibling presence in the recordings affected these variables. That is, if it is the active

[380] presence of the sibling that affects how the caretaker interacts with the target child, then we

[381] would expect to see a difference in our input measures when the sibling is present vs. absent.

[382] On the other hand, if the very fact of having a sibling changes the way that a caregiver

[383] interacts with the infant regardless of whether any sibling is actually present on the scene,

[384] then no difference should be observed. While sibling presence in each recordings was not

[385] coded directly in the dataset, for this exploratory analysis we can get at this with an

[386] admittedly imperfect proxy: did the sibling produce nouns in the recording. If yes, we can

[387] safely assume they are present; if not we (less safely, but reasonably for initial exploratory

[388] purposes) assume they are not. As reported above, by this measure, at least one sibling was

[389] present in 68% of recordings for the infants who had a sibling.

[390] Since the presence of a sibling in any given infant's data changed month-on-month

[391] (i.e. sometimes the sibling was present and sometimes they were not), and since our measure

[392] of sibling presence is imperfect, we opt here to describe the pattern of data without drawing

[393] any strong conclusions from statistical models. Descriptively, the presence of a sibling
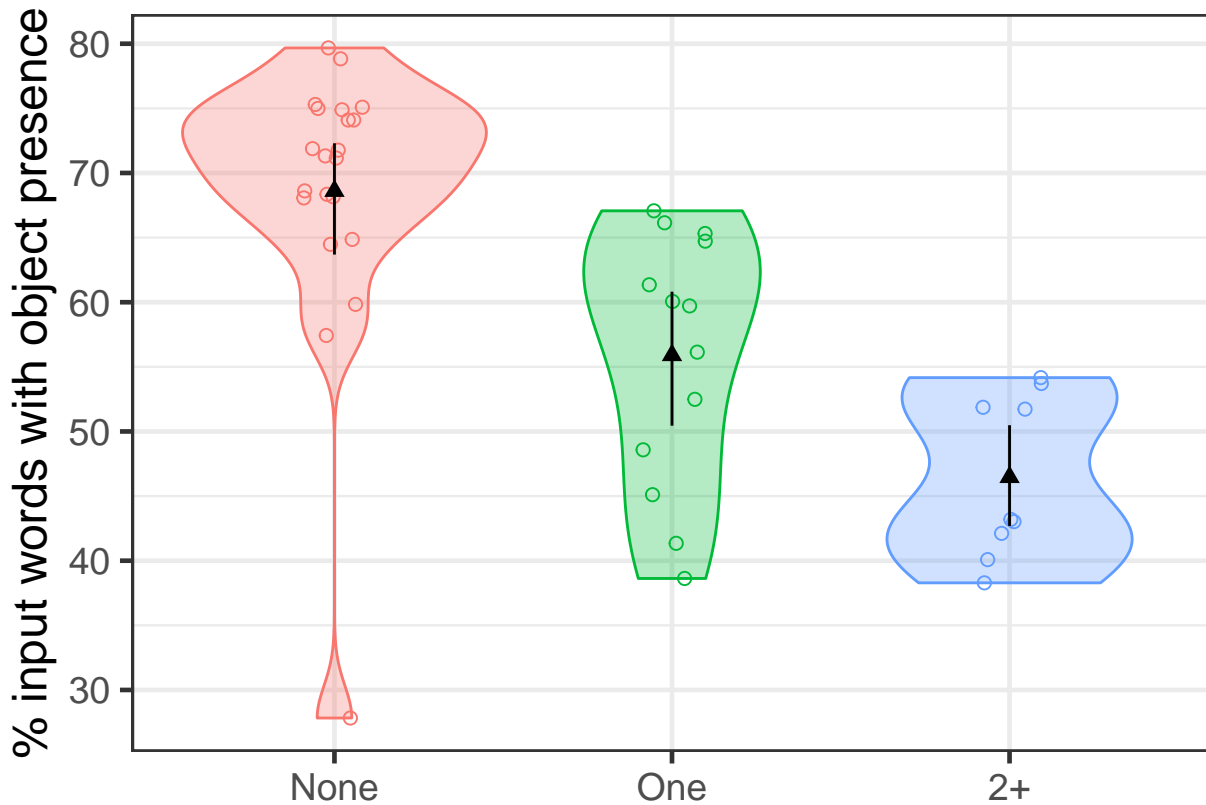
Figure 3. Percentage of input words produced with object presence across sibling groups. Error bars and black triangles show 95% CIs and mean proportion of object presence across sibling groups. Dots indicate mean proportion of object presence per infant, collapsing across age and jittered horizontally for visual clarity. See S7 for a month-by-month visualisation of object presence.

394 affected the amount of object presence in the data, but not the amount of input. See Table 6

395 and Figures 4 and 5. Overall, the presence of a sibling negatively affected object presence,

396 and this was consistent over time; when a sibling was present, infants in both groups heard

397 less object presence. This effect was stronger for infants with two or more siblings (though

398 note that it is unclear from our measure how many siblings were present, and it is possible

399 that only one sibling was present in the recording), and overall this was true regardless of

400 whether the sibling was present, or whether the infant was alone with the caregiver. The

401 picture is less clear for caregiver input, where the presence of a sibling has a more variable

Table 6

Data summary of our two input measures according to presence or absence of siblings during the recording.

Variable	Sibling presence	1 sibling		2+ siblings	
		Mean	SD	Mean	SD
N Input utterances, 10-17 months	Sibling not present	136.60	107.68	84.58	58.43
	Sibling present	124.41	93.55	73.51	48.34
% object presence in input, 10-17 months	Sibling not present	68.81	12.36	59.22	14.25
	Sibling present	52.63	14.82	40.85	14.81

402 effect on the number of object words produced by caregivers (see Figure 4), particularly for

403 infants with only one sibling. This supports our findings above, suggesting that the presence

404 of one additional child does not have any negative effects on the amount of input that

405 caregiver provide. However, input was consistently lower for the group with two or more

406 siblings, and this was true regardless of whether or not a sibling was present; again, this is

407 consistent with the findings reported above for this input measure.

Discussion

409   We investigated the nature of infant language development in relation to number of

410 children in the household. Previous research found a delay in lexical acquisition for

411 later-born infants (Fenson et al., 1994; Hoff, 2006), with differences in input across birth

412 order reported as a root cause. Our results add several new dimensions to this, by testing for

413 differences across more vs. fewer older siblings, and by looking at input during child-centered

414 home recordings. Infants with more siblings were reported to say fewer words by 18 months,

415 heard fewer nouns from their parents, and were less likely to be attending to an object when

416 hearing its label. Importantly, and in contrast with some previous research (Hoff-Ginsberg,
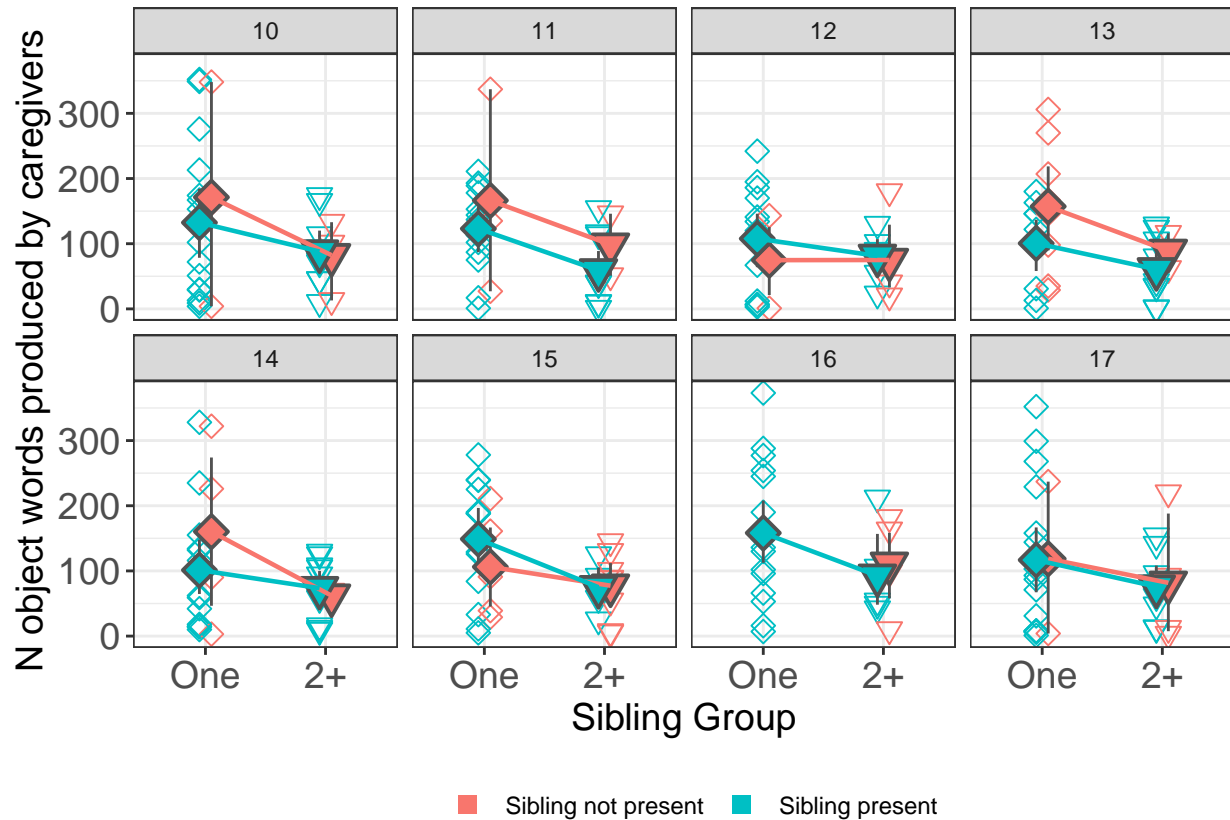
Figure 4. Difference in number of input words in infants' input according to whether or not a sibling or siblings were present during the time of recording, for each month of data. Infants with no siblings were not included in the plots for visual ease. White shapes represent individual infants in the data; filled shapes represent means and 95%CIs, colours represent presence or absence of siblings during the recording session.

1998; Oshima-Takane & Robbins, 2003), infants with one sibling showed no decrement in lexical production and minimal reduction in input in comparison to first-born infants. That is, our results suggest that simply having a sibling does not contribute to input or vocabulary differences across children (as measured here), while having more than one siblings seems to do so. Indeed, infants with zero and one sibling had similar results for productive vocabulary, and parental noun input overall (though not object presence). Moreover, parental input was not affected by the presence or absence of the sibling in the room. In contrast, infants with two or more siblings said fewer words, and also heard fewer input words overall.
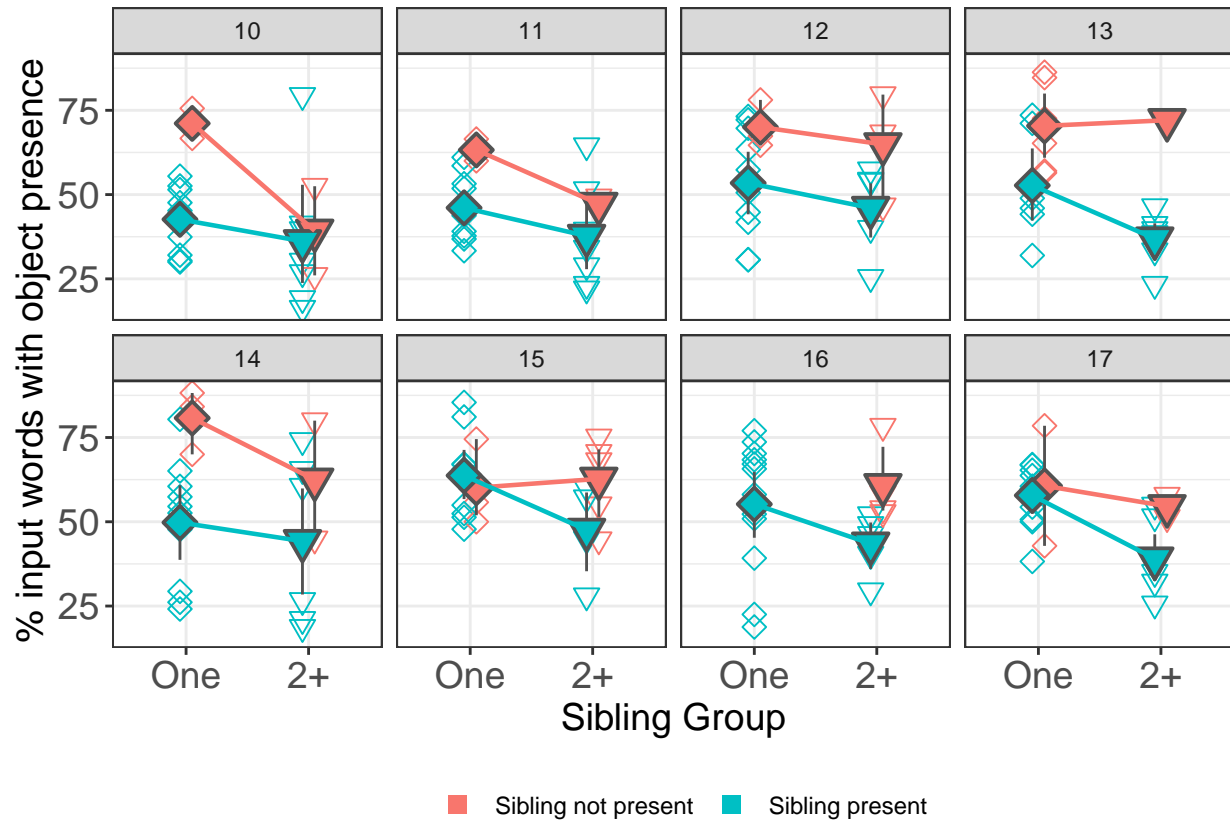
Figure 5. Difference in % of object presence in infants' input according to whether or not a sibling or siblings were present during the time of recording, for each month of data. Infants with no siblings were not included in the plots for visual ease. White shapes represent individual infants in the data; filled shapes represent means and 95%CIs, colours represent presence or absence of siblings during the recording session.

With regards to object presence, having siblings made it increasingly less likely to hear an object label when attending to it, and this effect increased with sibship size (i.e. children with more siblings heard input containing a lower proportion of object presence). Unlike for total parental noun input (which was reduced for 2+ siblings but not modulated based on whether siblings were present in the recordings), reduced object presence for children with more siblings was particularly notable in recordings with siblings present.

The sibling effect. When we considered the effect of sibling status – that is, whether or not infants had any siblings, disregarding specific sibling number – our findings

433 showed that having siblings made no difference to infants' lexical production capacities. This

434 contrasts with Hoff-Ginsberg (1998), who found that, by 18 months, laterborns exhibit lower

435 language skills. However, Oshima-Takane and colleagues (1996) found no overall differences

436 between first- and second-born children across a range of language measures taken at 21

437 months. Our results suggest that considering *sibling quantity* may be a more sensitive way to

438 reveal demographic effects than their (coarser-grained) first- vs. later-born status. We find

439 that the more older siblings a child had, the lower their reported productive vocabulary at 18

440 months. This adds to findings from Fenson and colleagues (1994), who found a weak but

441 significant negative correlation between birth order and production of both words and

442 gestures. Controlling for age, our model showed that for each additional older sibling, infants

443 produced more than 30% fewer words by 18 months.

444   While infants with more siblings heard less input speech overall, having one sibling did

445 not significantly reduce the number of nouns in an infant's input. This is in direct contrast

446 with reports from the literature; Hoff (2006) states that "when a sibling is present, each child

447 receives less speech directed solely at... her *because mothers produce the same amount of*

448 *speech whether interacting with one or two children*" (p.67, italics added). While this does

449 not appear to be the case in the present dataset, it may be due to the circumstances of the

450 home-recorded data: while siblings were present in many of the recordings (68% of

451 recordings in which the target child had a sibling), given the focus of the data collection,

452 parents may have had a tendency to direct their attention - and consequently their linguistic

453 input - more towards the target child; our samples also differed in other ways

454 (e.g. sociocultural context) that may have influenced the results as well. Alternatively, our

455 results may diverge from those of Hoff (2006) due to the nature of our input measure, which

456 only took nouns into account. That said, we find this alternative explanation unlikely given

457 work by Bulgarelli and Bergelson (2020) showing that nouns are a reliable proxy for overall

458 input in this dataset, suggesting that this measure provides an appropriate representation of

459 overall input directed at the target child.

460   In contrast to the other results, our analysis of object presence showed a more linear

461 'sibling effect'. In this case, even having one sibling led to fewer word-object pairs presented

462 in the input. This was true regardless of whether or not other siblings were present, but

463 object presence was further negatively affected by the presence of a sibling in the room.

464 Presence of a labeled object with congruent input speech has been found to support early

465 word learning across several studies. For instance, Bergelson and Aslin (2017) combined

466 analysis of this home-recorded data at six months with an experimental study to show that

467 word-object presence in naturalistic caregiver input correlated with comprehension of nouns

468 (tested using eye-tracking). Relatedly, Gogate and colleagues (2000) propose that contingent

469 word production supports the learning of novel word-object combinations, with "multimodal

470 motherese" - whereby a target object word is produced in movement or touch-based

471 synchrony with its referent - supporting word learning. More broadly, lower rates of

472 referential transparency for common non-nouns like *hi* and *uh-oh* have been proposed to

473 potentially explain why these words are learned later than common concrete nouns

474 (Bergelson & Swingley, 2013). While the present results on object presence don't speak

475 directly to word learning, they do suggest that this potentially helpful learning support is

476 less available for children with more siblings.

477   **Siblingese as a learning opportunity?**   We also found that infants with siblings

478 did not hear much speech from their older brothers and sisters. Similar findings are reported

479 in a lab-based interaction study by Oshima-Takane and Robbins (2003), who found that

480 older siblings rarely talked directly to the target child; instead, most input from siblings was

481 overheard speech from sibling-mother interactions. One possibility raised by these results is

482 that perhaps parents are able to compensate or provide relatively similar input and learning

483 support for one or two children, but once children outnumber parents, this balancing act of

484 attention, care, and time becomes unwieldy. While the current sample is relatively limited

485 and homogeneous in the family structures and demographics it includes, future work could

486 fruitfully investigate this possibility by considering whether (controlling for other potential

contributors like SES, Hoff-Ginsberg, 1998) the presence of more caregivers (whether parents, relatives, or other adults) helps foster language development.

Alternatively, second-borns might 'even out' with children with no siblings due to a trade-off between direct attention from the caregiver and the possibility of more sophisticated social-communicative interactions. For these infants there is still ample opportunity to engage with the mother in one-to-one interactions, allowing a higher share of her attention than is available to third- or later-borns. Furthermore, triadic interactions can benefit the development of a number of linguistic and communication skills (Barton & Tomasello, 1991; Dunn & Shatz, 1989). Second-borns may also benefit from overheard speech in their input, supporting the acquisition of nouns and even more complex lexical categories (Floor & Akhtar, 2006; Oshima-Takane et al., 1996). For infants with one sibling, the benefits of observing/overhearing interactions between sibling and caregiver, as well as the possibility for partaking in such interactions, may outweigh the decrease in some aspects of the input (i.e., in our data, only observed in object presence). Having more than one sibling may throw this off-balance, such that the possibilities for both supportive one-to-one input *and* more sophisticated interactions are simultaneously diminished.

Importantly, the present results make no claims about eventual outcomes for these children: generally speaking, regardless of sibling number, all typically-developing infants reach full and fluent language use. Indeed, some research suggests that sibling effects, while they may be clear in early development, are not always sustained into childhood; e.g. twins are known to experience a delay in language development into the third year, but are quick to catch up thereafter (Dales, 1969; Tomasello, Mannle, & Kruger, 1986). This demonstrates the cognitive adaptability of early development, which brings about the acquisition of language across varying and allegedly 'imperfect' learning environments. Infants' capacity to develop linguistic skills from the resources that are available to them – whether that is infant-directed object labels or overheard abstract concepts – highlights the dynamic and adaptable nature of early cognitive development, and a system that is sufficiently robust to

514 bring about the same outcome across populations.

515   **Limitations.**   Of course, the 'success' of early language development is defined by

516 how success is measured. Here we chose word production as our measure of linguistic

517 capability; we did not consider other equally valid measures such as language comprehension

518 or early social-interaction skills. Similarly, our input measures focused on nouns; other lexical

519 classes may reveal different effects, though they are generally far sparser in production until

520 toddlerhood. Our analysis of vocabulary relied on parental report data; this method could

521 have biased our first-born sample towards more accurate or larger vocabulary reports owing

522 to their parents having more time and attention to spend observing their vocabulary

523 development (see Kartushina et al., 2022 for a discussion of this possibility in light of the

524 COVID-19 pandemic, though note the present data were collected in 2014-16). However, we

525 were able to validate this measure by running correlation tests between reported (CDI)

526 vocabulary and the number of words produced by each infant in both the audio and video

527 data. Across both type and token counts, reported productive vocabulary size correlated

528 strongly with words produced in the audio/video data (rho values all between 0.54-0.60).

529   There is also some imbalance in group sizes across our data; our sample was not

530 pre-selected for sibling number, and so group sizes are unmatched across the analysis.

531 Including a larger number of infants with 2+ siblings may have revealed a different pattern

532 of results. Finally, more work across wider and larger populations is necessary to unpack the

533 generalizability of the present results. Our sample is reflective of average household sizes in

534 middle-class families across North America and Western Europe (Office for National

535 Statistics, 2018; United States Census Bureau, 2010), but it is not unusual in some

536 communities and parts of the world for households to include between three and six children

537 on average (Institute for Family Studies & Wheatley Institution, 2019). Adding to this, it is

538 also necessary to consider cross-cultural differences in the way children are addressed by

539 their parents, other caretakers, and other children (Bunce et al., 2020; Casillas, Brown, &

540 Levinson, 2019; Shneidman & Goldin-Meadow, 2012). For instance, Bunce and colleagues

541 (2020) find relatively similar rates of target child directed speech across US, Canadian,

542 Argentinian, UK, Papuan and Mayan samples, some differences in who the input comes from,

543 and large effects of number of talkers present. These results suggest that caution is advisable

544 before generalizing the current results to any other socio-cultural contexts, but also pose

545 exciting open questions regarding what variability in experiences do – or don't – change

546 about early language interaction and development.

Conclusion

548 Our results with English-learning infants in the US support prior findings from the

549 literature showing that later-born infants have slower lexical acquisition than their first-born

550 peers. However, we highlight an important difference from previous findings, namely that in

551 the present sample, second-born infants show no such effect, while infants with more than

552 two siblings have significantly smaller productive vocabularies at age 18 months.

553 Correspondingly, we identified parallel group differences in overall noun input and object

554 presence. While we did not test these corresponding measures directly, our results suggest

555 that having more siblings affects a child's early language environment, which in turn may

556 lead to slower vocabulary growth in the first 18 months of life. We look forward to future

557 studies considering the granularity of more versus fewer siblings, and how this relates to

558 language abilities over the course of development.

References

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*(2), 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning Words Through Overhearing. *Child Development*, *72*(2), 416–430. Retrieved from http://www.jstor.org/stable/1132404

Anderson, N. J., Graham, S. A., Prime, H., Jenkins, J. M., & Madigan, S. (2021). Linking Quality and Quantity of Parental Linguistic Input to Child Language Skills: A Meta-Analysis. *Child Development*, *92*(2), 484–501. https://doi.org/10.1111/cdev.13508

Barnes, S., Gutfreund, M., Satterly, D., & Wells, G. (1983). Characteristics of adult speech which predict children's language development. *Journal of Child Language*, *10*(1), 65–84. https://doi.org/10.1017/S0305000900005146

Barton, M. E., & Tomasello, M. (1991). *Joint Attention and Conversation in Mother-Infant-Sibling Triads. 62*(3), 517–529.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. https://doi.org/10.1111/desc.12715

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921. https://doi.org/10.1073/pnas.1712966114

Bergelson, E., & Swingley, D. (2013). The Acquisition of Abstract Words by Young Infants. *Cognition*, *127*(3), 391–397. https://doi.org/10.1038/jid.2014.371

Berglund, E., Eriksson, M., & Westerlund, M. (2005). Communicative skills in relation to gender, birth order, childcare and socioeconomic status in 18-month-old children. *Scandinavian Journal of Psychology*, *46*(6), 485–491. https://doi.org/10.1111/j.1467-9450.2005.00480.x

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and Variability in Children's Word Learning Across Languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Bulgarelli, F., & Bergelson, E. (2020). Look who's talking: A comparison of automated and human-generated speaker tags in naturalistic day-long recordings. *Behavior Research Methods*, *52*(2), 641–653. https://doi.org/10.3758/s13428-019-01265-7

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Migdalek, M., & Casillas, M. (2020). *A cross-cultural examination of young children's everyday language experiences* [Preprint].

Cartmill, E. a., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the United States of America.* https://doi.org/10.1073/pnas.1309518110

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early Language Experience in a Tseltal Mayan Village. *Child Development, EarlyView article.* https://doi.org/10.1111/cdev.13349

Dales, R. J. (1969). Motor and language development of twins during the first three years. *The Journal of Genetic Psychology; Provincetown, Mass., Etc.*, *114*(2), 263–271. Retrieved from https://search.proquest.com/docview/1297124434/citation/D928716F9A7E4AEFPQ/1

Dunn, J., & Shatz, M. (1989). Becoming a Conversationalist despite (Or Because of) Having an Older Sibling. *Child Development*, *60*(2), 399–410.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, M., Stephen J. Tomasello, ... Stiles, J. (1994). Variability in Early Communicative Development. *Monographs of the Society for Research in Child Development*, *59*. https://doi.org/10.2307/1166093

Fitch, A., Liberman, A. M., Luyster, R., & Arunachalam, S. (2020). Toddlers' word learning through overhearing: Others' attention matters. *Journal of Experimental Child Psychology*, *193*, 104793. https://doi.org/10.1016/j.jecp.2019.104793

Floor, P., & Akhtar, N. (2006). Can 18-month-old infants learn words by listening in on conversations? *Infancy*, *9*(3), 327–339. https://doi.org/10.1207/s15327078in0903_4

Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, *206*, 104415. https://doi.org/10.1016/j.cognition.2020.104415

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and Consistency in Early Language Learning: The Wordbank Project*. MIT Press.

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. https://doi.org/10.1111/1467-8624.00197

Gurgand, L., Lamarque, L., Havron, N., Bernard, J. Y., Ramus, F., & Peyre, H. (2022). The influence of sibship composition on language development at 2 years of age in the ELFE birth cohort study. *Developmental Science*, *n/a*(n/a), e13356. https://doi.org/10.1111/desc.13356

Havron, N., Lovcevic, I., Kee, M., Chen, H., Chong, Y., Daniel, M., . . . Tsuji, S. (2022). The Effect of Older Sibling, Postnatal Maternal Stress, and Household Factors on Language Development in Two- to Four-Year-Old Children. *Developmental Psychology*, *58*(11), 2096–2113. https://doi.org/10.31234/osf.io/m9w48

Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., . . . Thiebaugeorges, O. (2019). The Effect of Older Siblings on Language Development as a Function of Age Difference and Sex. *Psychological Science*, *30*(9), 1333–1343. https://doi.org/10.1177/0956797619861436

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics*, *19*(4), 603–629. https://doi.org/10.1017/S0142716400010389

Institute for Family Studies, & Wheatley Institution. (2019). *World family map 2019: Mapping family change and child well-being outcomes*. Charlottesville, VA: Institute for Family Studies. Retrieved from https://ifstudies.org/reports/world-family-map/2019/executive-summary

Jakiela, P., Ozier, O., Fernald, L., & Knauer, H. (2020). Big sisters. *Policy Research Working Papers*, *9454*. Retrieved from https://openknowledge.worldbank.org/server/api/core/bitstreams/54b5d934-3263-5e51-93af-7d1e9e619804/content

Jones, C. P., & Adamson, L. B. (1987). *Language Use in Mother-Child and Mother-Child-Sibling Interactions*. *58*(2), 356–366.

Kartushina, N., Mani, N., Aktan-Erciyes, A., Alaslani, K., Aldrich, N. J., Almohammadi, A., . . . Mayor, J. (2022). COVID-19 first lockdown as a window into language acquisition: Associations between caregiver-child activities and vocabulary gains. *Language Development Research*, *2*(1). https://doi.org/10.34842/abym-xv34

Karwath, C., Relikowski, I., & Schmitt, M. (2014). Sibling structure and educational achievement: How do the number of siblings, birth order, and birth spacing affect children's vocabulary competences? *Journal of Family Research*, *26*(3), 372–396. https://doi.org/10.3224/zff.v26i3.18993

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). {lmerTest} Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laing, C. E., & Bergelson, E. (2019). Mothers' Work Status and 17-Month-Olds'

Productive Vocabulary. *Infancy*, *24*(1), 101–109.

https://doi.org/10.1111/infa.12265

Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point,

Walk, Talk: Links Between Three Early Milestones, From Observation and

Parental Report. *Developmental Psychology.* https://doi.org/10.1037/dev0000738

Office for National Statistics. (2018). *Families with dependent children by number of*

children, UK, 1996 to 2017 (No. 008855). Office for National Statistics. Retrieved

from Office for National Statistics website: https://www.ons.gov.uk/

peoplepopulationandcommunity/birthsdeathsandmarriages/families/adhocs/

008855familieswithdependentchildrenbynumberofchildrenuk1996to2017

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). *Birth Order Effects on*

Early Language Development : Do Secondborn Children Learn from Overheard

Speech ? Author (s): Yuriko Oshima-Takane , Elizabeth Goodz and Jeffrey L .

Derevensky Published by : Wiley on behalf of the Society for Research in Child De.

67(2), 621–634.

Oshima-Takane, Y., & Robbins, M. (2003). Linguistic environment of secondborn

children. *First Language*, *23*(1), 21–40.

https://doi.org/http://dx.doi.org/10.1177/0142723703023001002

Peyre, H., Bernard, J. Y., Hoertel, N., Forhan, A., Charles, M.-A., De Agostini, M.,

. . . Ramus, F. (2016). Differential effects of factors influencing cognitive

development at the age of 5-to-6 years. *Cognitive Development*, *40*, 152–162.

https://doi.org/10.1016/j.cogdev.2016.10.001

Pine, J. M. (1995). Variation in Vocabulary Development as a Function of Birth

Order. *Child Development*, *66*(1), 272–281.

R Core Team. (2019). *R: A Language Environment for Statistical Computing.* R

Foundation for Statistical Computing. Retrieved from

https://www.R-project.org/

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1016/j.surg.2006.10.010.Use

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language*, *47*(1), 5–21. https://doi.org/10.1017/S0305000919000655

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. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

Stern, D. N., Spieker, S., Barnett, R. K., & MacKain, K. (1983). The Prosody of Maternal Speech: Infant Age and Context Related Changes. *Journal of Child Language*, *10*(1), 1–15. https://doi.org/10.1017/S0305000900005092

Tomasello, M., Mannle, S., & Kruger, A. C. (1986). Linguistic environment of 1- to 2-year-old twins. *Developmental Psychology*, *22*(2), 169–176. https://doi.org/10.1037/0012-1649.22.2.169

United States Census Bureau. (2010). *Household Type by Number of People Under 18 Years* (No. PCT16). Retrieved from https: //data.census.gov/cedsci/table?q=number%20of%20children&hidePreview= false&tid=DECENNIALSF12010.PCT16&t=Children&vintage=2018

Zambrana, I. M., Ystrom, E., & Pons, F. (2012). Impact of Gender, Maternal Education, and Birth Order on the Development of Language Comprehension: A Longitudinal Study from 18 to 36 Months of Age. *Journal of Developmental &*

721   *Behavioral Pediatrics, 33*(2), 146–155.

722   https://doi.org/10.1097/DBP.0b013e31823d4f83