

¹ Peekbank: An open, large-scale repository for developmental eye-tracking data of children's
² word recognition

³ Martin Zettersten¹, Daniel Yurovsky², Tian Linger Xu³, Sarp Uner⁴, Angeline Sin Mei Tsui⁵,
⁴ Rose M. Schneider⁶, Annissa N. Saleh⁷, Stephan Meylan^{8,9}, Jessica Mankowitz⁵, Kyle
⁵ MacDonald¹⁰, Bria Long⁵, Molly Lewis², George Kachergis⁵, Kunal Handa¹¹, Benjamin
⁶ deMayo¹, Alexandra Carstensen⁶, Mika Braginsky⁹, Veronica Boyce⁵, Naiti S. Bhatt¹²,
⁷ Claire Bergey¹³, & Michael C. Frank⁵

⁸ ¹ Department of Psychology, Princeton University

⁹ ² Department of Psychology, Carnegie Mellon University

¹⁰ ³ Department of Psychological and Brain Sciences, Indiana University

¹¹ ⁴ Data Science Institute, Vanderbilt University

¹² ⁵ Department of Psychology, Stanford University

¹³ ⁶ Department of Psychology, University of California, San Diego

¹⁴ ⁷ Department of Psychology, The University of Texas at Austin

¹⁵ ⁸ Department of Psychology and Neuroscience, Duke University

¹⁶ ⁹ Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

¹⁷ ¹⁰ Core Technology, McD Tech Labs

¹⁸ ¹¹ Brown University

¹⁹ ¹² Department of Psychology, New York University

²⁰ ¹³ Department of Psychology, University of Chicago

²¹ Author Note

²² All code is available at <https://github.com/langcog/peekbank-paper>. Raw and
²³ standardized datasets are available on the Peekbank OSF repository (<https://osf.io/pr6wu/>)
²⁴ and can be accessed using the peekbankr R package
²⁵ (<https://github.com/langcog/peekbankr>).

26

Abstract

27 The ability to rapidly recognize words and link them to referents is central to children's
28 early language development. This ability, often called word recognition in the developmental
29 literature, is typically studied in the looking-while-listening paradigm, which measures
30 infants' fixation on a target object (vs. a distractor) after hearing a target label. We present
31 a large-scale, open database of infant and toddler eye-tracking data from
32 looking-while-listening tasks. The goal of this effort is to address theoretical and
33 methodological challenges in measuring vocabulary development. We first present how we
34 created the database, its features and structure, and associated tools for processing and
35 accessing infant eye-tracking datasets. Using these tools, we then work through two
36 illustrative examples to show how researchers can use Peekbank to interrogate theoretical
37 and methodological questions about children's developing word recognition ability.

38 *Keywords:* word recognition; eye-tracking; vocabulary development;

39 looking-while-listening; visual world paradigm; lexical processing

40 Word count: 6623

41 Peekbank: An open, large-scale repository for developmental eye-tracking data of children's
42 word recognition

43 Across their first years of life, children learn words at an accelerating pace (Frank,
44 Braginsky, Yurovsky, & Marchman, 2021). While many children will only produce their first
45 word at around one year of age, most children show signs of understanding many common
46 nouns (e.g., *mommy*) and phrases (e.g., *Let's go bye-bye!*) much earlier in development
47 (Bergelson & Swingley, 2012, 2013). Although early word understanding is a critical element
48 of first language learning, the processes involved are less directly apparent in children's
49 behaviors and are less accessible to observation than developments in speech production
50 (Fernald, Zangl, Portillo, & Marchman, 2008). To understand a spoken word, children must
51 process the incoming auditory signal and link that signal to relevant meanings – a process
52 often referred to as word recognition. One of the primary means of measuring word
53 recognition in young infants is using eye-tracking techniques that gauge where children look
54 in response to linguistic stimuli (Fernald, Zangl, Portillo, & Marchman, 2008). The logic of
55 these methods is that if, upon hearing a word, a child preferentially looks at a target
56 stimulus rather than a distractor, the child is able to recognize the word and activate its
57 meaning during real-time language processing. Measuring early word recognition offers
58 insight into children's early word representations: children's speed of response (i.e., moving
59 their eyes; turning their heads) to the unfolding speech signal can reveal children's level of
60 comprehension (Bergelson, 2020; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).
61 Word recognition skills are also thought to build a foundation for children's subsequent
62 language development. Past research has found that early word recognition efficiency is
63 predictive of later linguistic and general cognitive outcomes (Bleses, Makransky, Dale, Højlen,
64 & Ari, 2016; Marchman et al., 2018).

65 While word recognition is a central part of children's language development, mapping
66 the trajectory of word recognition skills has remained elusive. Studies investigating children's

67 word recognition are typically limited in scope to experiments in individual labs involving
68 small samples tested on a handful of items. The limitations of single datasets makes it
69 difficult to understand developmental changes in children’s word knowledge at a broad scale.

70 One way to overcome this challenge is to compile existing datasets into a large-scale
71 database in order to expand the scope of research questions that can be asked about the
72 development of word recognition abilities. This strategy capitalizes on the fact that the
73 looking-while-listening paradigm is widely used, and vast amounts of data have been
74 collected across labs on infants’ word recognition over the past 35 years (Golinkoff, Ma, Song,
75 & Hirsh-Pasek, 2013). Such datasets have largely remained isolated from one another, but
76 once combined, they have the potential to offer insights into lexical development at a broad
77 scale. Similar efforts to collect other measures of language development have borne fruit in
78 recent years. For example, WordBank aggregated data from the MacArthur-Bates
79 Communicative Development Inventory, a parent-report measure of child vocabulary, to
80 deliver new insights into cross-linguistic patterns and variability in vocabulary development
81 (Frank, Braginsky, Yurovsky, & Marchman, 2017, 2021). In this paper, we introduce
82 *Peekbank*, an open database of infant and toddler eye-tracking data aimed at facilitating the
83 study of developmental changes in children’s word recognition.

84 Measuring Word Recognition: The “Looking-While-Listening” Paradigm

85 Word recognition is traditionally studied in the “looking-while-listening” paradigm
86 (Fernald, Zangl, Portillo, & Marchman, 2008; alternatively referred to as the intermodal
87 preferential looking procedure, Hirsh-Pasek, Cauley, Golinkoff, & Gordon, 1987). In these
88 studies, infants listen to a sentence prompting a specific referent (e.g., *Look at the dog!*)
89 while viewing two images on the screen (e.g., an image of a dog – the target image – and an
90 image of a bird – the distractor image). Infants’ word recognition is evaluated by how
91 quickly and accurately they fixate on the target image after hearing its label. Past research

92 has used this basic method to study a wide range of questions in language development. For
93 example, the looking-while-listening paradigm has been used to investigate early noun
94 knowledge, phonological representations of words, prediction during language processing, and
95 individual differences in language development (Bergelson & Swingley, 2012; Golinkoff, Ma,
96 Song, & Hirsh-Pasek, 2013; Lew-Williams & Fernald, 2007; Marchman et al., 2018; Swingley
97 & Aslin, 2002).

98 While this research has been fruitful in advancing understanding of early word
99 knowledge, fundamental questions remain. One central question is how to accurately capture
100 developmental change in the speed and accuracy of word recognition. There is ample
101 evidence demonstrating that infants get faster and more accurate in word recognition over
102 the first few years of life (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).
103 However, precisely measuring developmental increases in the speed and accuracy of word
104 recognition remains challenging due to the difficulty of distinguishing developmental changes
105 in word recognition skill from changes in knowledge of specific words. This problem is
106 particularly thorny in studies with young children, since the number of items that can be
107 tested within a single session is limited and items must be selected in an age-appropriate
108 manner (Peter et al., 2019). More broadly, other key differences in the design choices (e.g.,
109 how distractor items are selected) and analytic decisions (e.g., how the analysis window is
110 defined) between studies could obscure developmental change if not appropriately taken into
111 account.

112 One approach to addressing these challenges is to conduct meta-analyses
113 aggregating effects across studies while testing for heterogeneity due to researcher choices
114 (Bergmann et al., 2018; Lewis et al., 2016). However, meta-analyses typically lack the
115 granularity to estimate participant-level and item-level variation or to model behavior
116 beyond coarse-grained effect size estimates. An alternative way to approach this challenge is
117 to aggregate trial-level data from smaller studies measuring word recognition with a wide

118 range of items and design choices into a large-scale dataset that can be analyzed using a
119 unified modeling approach. A sufficiently large dataset would allow researchers to estimate
120 developmental change in word recognition speed and accuracy while generalizing across
121 changes related to specific words or the design features of particular studies.

122 A related open theoretical question is understanding changes in children's word
123 recognition at the level of individual items. Looking-while-listening studies have been limited
124 in their ability to assess the development of specific words. One limitation is that studies
125 typically test only a small number of trials for each item, limiting the power the accurately
126 measure the development of word-specific accuracy (DeBolt, Rhemtulla, & Oakes, 2020). A
127 second limitation is that target stimuli are often yoked with a limited set of distractor
128 stimuli (i.e., a child sees a target with only one or two distractor stimuli over the course of
129 an experiment), leaving ambiguous whether accurate looking to a particular target word can
130 be attributed to children's recognition of the target word or their knowledge about the
131 distractor. Aggregating across many looking-while-listening studies has the potential to meet
132 these challenges by increasing the number of observations for specific items at different ages
133 and by increasing the size of the inventory of distractor stimuli that co-occur with each
134 target.

135 Replicability and Reproducibility

136 A core challenge facing psychology in general, and the study of infant development in
137 particular, are threats to the replicability and reproducibility of core empirical results (Frank
138 et al., 2017; Nosek et al., 2021). In infant research, many studies are not adequately powered
139 to detect the main effects of interest (Bergmann et al., 2018). This issues is compounded by
140 low reliability in infant measures, often due to limits on the number of trials that can be
141 collected from an individual infant in an experimental session (Byers-Heinlein, Bergmann, &
142 Savalei, 2021). One hurdle to improving power in infant research is that it can be difficult to

143 develop a priori estimates of effect sizes and how specific design decisions (e.g., the number
144 of test trials) will impact power and reliability. Large-scale databases of infant behavior can
145 aid researchers in their decision-making by allowing them to directly test how different
146 design decisions affect power and reliability. For example, if a researcher is interested in
147 understanding how the number of test trials could impact the power and reliability of their
148 looking-while-listening design, a large-scale infant eye-tracking database would allow them to
149 simulate possible outcomes across a range of test trials, providing the basis for data-driven
150 design decisions.

151 In addition to threats to replicability, the field of infant development also faces
152 concerns about analytic *reproducibility* – the ability for researchers to arrive at the same
153 analytic conclusion reported in the original research article, given the same dataset. A recent
154 estimate based on studies published in a prominent cognitive science journal suggests that
155 analyses can remain difficult to reproduce, even when data are made available to other
156 research teams (Hardwicke et al., 2018). Aggregating data in centralized databases can aid
157 in improving reproducibility in several ways. First, building a large-scale database requires
158 defining a standardized data specification. Recent examples include the **brain imaging**
159 **data structure** (BIDS), an effort to specify a unified data format for neuroimaging
160 experiments (Gorgolewski et al., 2016), and the data formats associated with **ChildProject**,
161 for managing long-form at-home language recordings (Gautheron, Rochat, & Cristia, under
162 review). Defining a data standard – in this case, for infant eye-tracking experiments –
163 supports reproducibility by guaranteeing that critical information will be available in openly
164 shared data and by making it easier for different research teams to understand the data
165 structure. Second, open databases make it easy for researchers to generate open and
166 reproducible analytic pipelines, both for individual studies and for analyses aggregating
167 across datasets. Creating open analytic pipelines across many datasets also serves a
168 pedagogical purpose, providing teaching examples illustrating how to implement analytic
169 techniques used in influential studies and how to conduct reproducible analyses with infant

170 eye-tracking data.

171 **Peekbank: An open database of developmental eye-tracking studies.**

172 What all of these open challenges share is that they are difficult to address at the scale
173 of a single research lab or in a single study. To address this challenge, we developed
174 *Peekbank*, a flexible and reproducible interface to an open database of developmental
175 eye-tracking studies. The Peekbank project (a) collects a large set of eye-tracking datasets
176 on children’s word recognition, (b) introduces a data format and processing tools for
177 standardizing eye-tracking data across heterogeneous data sources, and (c) provides an
178 interface for accessing and analyzing the database. In the current paper, we introduce the
179 key components of the project and give an overview of the existing database. We then
180 provide two worked examples of how researchers can use Peekbank. In the first, we examine
181 a classic result in the word recognition literature, and in the second we aggregate data across
182 studies to investigate developmental trends for the recognition of individual words.

183 **Design and Technical Approach**

184 **Database Framework**

185 One of the main challenges in compiling a large-scale eye-tracking database is the lack
186 of a shared data format: both labs and individual experiments can record their results in a
187 wide range of formats. For example, different experiments encode trial-level and
188 participant-level information in many different ways. Therefore, we have developed a
189 common tabular format to support analyses of all studies simultaneously.

190 As illustrated in Figure 1, the Peekbank framework consists of four main components:
191 (1) a set of tools to *convert* eye-tracking datasets into a unified format, (2) a relational

192 database populated with data in this unified format, (3) a set of tools to *retrieve* data from
193 this database, and (4) a web app (using the Shiny framework) for visualizing the data. These
194 components are supported by three packages. The `peekds` package (for the R language, R
195 Core Team, 2021) helps researchers convert existing datasets to use the standardized format
196 of the database. The `peekbank` module (Python) creates a database with the relational
197 schema and populates it with the standardized datasets produced by `peekds`. The database
198 is served through MySQL, an industry standard relational database server, which may be
199 accessed by a variety of programming languages, and can be hosted on one machine and
200 accessed by many others over the Internet. As is common in relational databases, records of
201 similar types (e.g., participants, trials, experiments, coded looks at each timepoint) are
202 grouped into tables, and records of various types are linked through numeric identifiers. The
203 `peekbankr` package (R) provides an application programming interface, or API, that offers
204 high-level abstractions for accessing the tabular data stored in Peekbank. Most users will
205 access data through this final package, in which case the details of data formatting,
206 processing, and the specifics of connecting to the database are abstracted away from the user.

207 Database Schema

208 The Peekbank database contains two major types of data: (1) metadata regarding
209 experiments, participants, and trials, and (2) time course looking data, detailing where a
210 child is looking on the screen at a given point in time (Fig. 2).

211 **Metadata.** Metadata can be separated into four parts: (1) participant-level
212 information (e.g., demographics), (2) experiment-level information (e.g., the type of eye
213 tracker used to collect the data), (3) session information (e.g. a participant's age for a
214 specific experimental session), and (4) trial information (e.g., which images or videos were
215 presented onscreen, and paired with which audio).

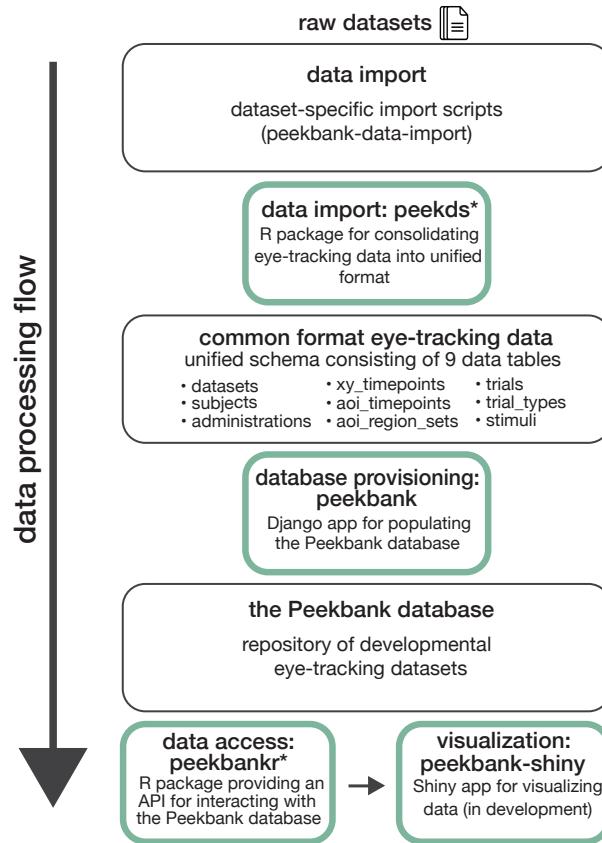


Figure 1. Overview of the Peekbank data ecosystem. Peekbank tools are highlighted in green. * indicates R packages introduced in this work.

216

Participant Information.

217

Invariant information about individuals who participate in one or more studies (e.g., a participant's first language) is recorded in the `subjects` table, while the `administrations` table contains information about each individual session in a given study (see Session Information, below). This division allows Peekbank to gracefully handle longitudinal designs: a single participant can complete multiple sessions and thus be associated with multiple administrations.

223

Participant-level data includes all participants who have experiment data. In general, we include as many participants as possible in the database and leave it to end-users to apply the appropriate exclusion criteria for their analysis.

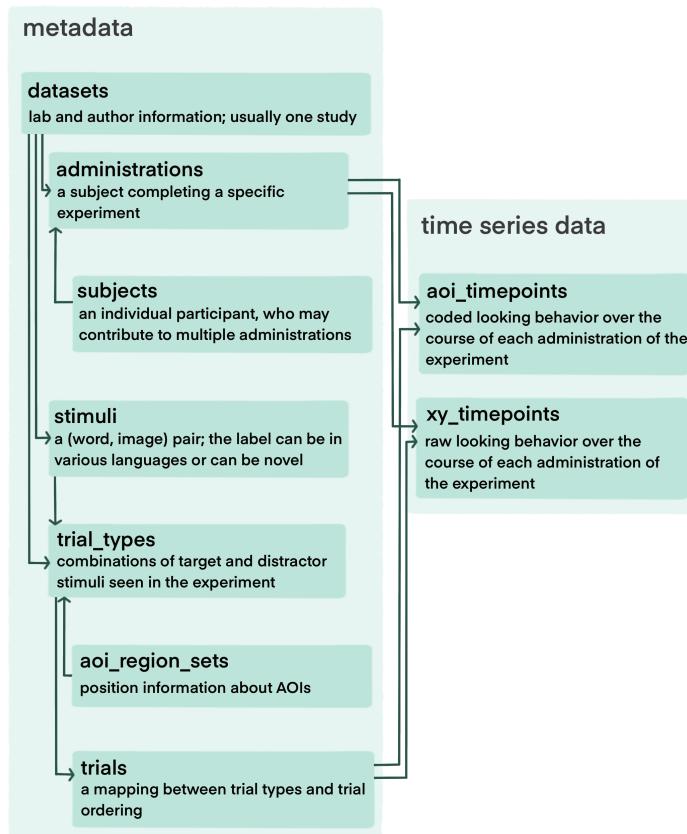


Figure 2. The Peekbank schema. Each darker rectangle represents a table in the relational database.

226

Experiment Information.

227

The **datasets** table includes information about the lab conducting the study and the relevant publications to cite regarding the data. In most cases, a dataset corresponds to a single study.

230

Information about the experimental design is split across the **trial_types** and **stimuli** tables. The **trial_types** table encodes information about each trial *in the design of the experiment*,² We note that the term *trial* is ambiguous and could be used to refer to both a particular combination of stimuli seen by many participants and a participant seeing that particular combination at a particular point in the experiment. We track the former in the **trial_types** table and the latter in the **trials** table.} including the target

236 stimulus and location (left vs. right), the distractor stimulus and location, and the point of
237 disambiguation for that trial. If a dataset used automatic eye-tracking rather than manual
238 coding, each trial type is additionally linked to a set of area of interest (x, y) coordinates,
239 encoded in the `aoi_region_sets` table. The `trial_types` table links trial types to the
240 `aoi_region_sets` table and the `trials` table. Each `trial_type` record links to two records
241 in the `stimuli` table, identified by the `distractor_id` and the `target_id` fields.

242 Each record in the `stimuli` table is a (word, image) pair. In most experiments, there is
243 a one-to-one mapping between images and labels (e.g., each time an image of a dog appears
244 it is referred to as *dog*). For studies in which there are multiple potential labels per image
245 (e.g., *dog* and *chien* are both used to refer to an image of a dog), images can have multiple
246 rows in the `stimuli` table with unique labels as well as a row with no label to be used when
247 the image appears solely as a distractor (and thus its label is ambiguous). This structure is
248 useful for studies on synonymy or using multiple languages. For studies in which the same
249 label refers to multiple images (e.g., the word *dog* refers to an image of a dalmatian and a
250 poodle), the same label can have multiple rows in the `stimuli` table with unique images.

251 ***Session Information.***

252 The `administrations` table includes information about the participant or experiment
253 that may change between sessions of the same study, even for the same participant. This
254 includes the age of the participant, the coding method (eye-tracking vs. hand-coding), and
255 the properties of the monitor that was used.

256 ***Trial Information.***

257 The `trials` table includes information about a specific participant completing a
258 specific instance of a trial type. This table links each record in the time course looking data
259 (described below) to the trial type and specifies the order of the trials seen by a specific
260 participant.

261 **Time course data.** Raw looking data is a series of looks to areas of interest (AOIs),

262 such as looks to the left or right of the screen, or to (x, y) coordinates on the experiment

263 screen, linked to points in time. For data generated by eye-trackers, we typically have (x, y)

264 coordinates at each time point, which we encode in the `xy_timepoints` table. These looks

265 are also recoded into AOIs according to the AOI coordinates in the `aoi_region_sets` table

266 using the `add_aois()` function in `peekds`, and encoded in the `aoi_timepoints` table. For

267 hand-coded data, we typically have a series of AOIs (i.e., looks to the left vs. right of the

268 screen), but lack information about exact gaze positions on-screen; in these cases the AOIs

269 are recoded into the categories in the Peekbank schema (target, distractor, other, and

270 missing) and encoded in the `aoi_timepoints` table; however, these datasets do not have any

271 corresponding data in the `xy_timepoints` table.

272 Typically, timepoints in the `xy_timepoints` table and `aoi_timepoints` table need to

273 be regularized to center each trial's time around the point of disambiguation – such that 0 is

274 the time of target word onset in the trial (i.e., the beginning of *dog* in *Can you find the*

275 *dog?*). We re-centered timing information to the onset of the target label to facilitate

276 comparison of target label processing across all datasets.¹ If time values run throughout the

277 experiment rather than resetting to zero at the beginning of each trial, `rezero_times()` is

278 used to reset the time at each trial. After this, each trial's times are centered around the

279 point of disambiguation using `normalize_times()`. When these steps are complete, the

280 time course is ready for resampling.

281 To facilitate time course analysis and visualization across datasets, time course data

282 must be resampled to a uniform sampling rate (i.e., such that every trial in every dataset has

283 observations at the same time points). All data in the database is resampled to 40 Hz

¹ While information preceding the onset of the target label in some datasets such as co-articulation cues (Mahr, McMillan, Saffran, Ellis Weismer, & Edwards, 2015) or adjectives (Fernald, Marchman, & Weisleder, 2013) can in principle disambiguate the target referent, we use a standardized point of disambiguation based on the onset of the label for the target referent. Onset times for other potentially disambiguating information (such as adjectives) can typically be recovered from the raw data provided on OSF.

(observations every 25 ms), which represents a compromise between retaining fine-grained timing information from datasets with dense sampling rates (maximum sampling rate among current datasets: 500 Hz) while minimizing the possibility of introducing artifacts via resampling for datasets with lower sampling rates (minimum sampling rate for current datasets: 30 Hz). Further, 25 ms is a mathematically convenient interval for ensuring consistent resampling; we found that using 33.333 ms (30 Hz) as our interval simply introduced a large number of technical complexities. The resampling operation is accomplished using the `resample_times()` function. During the resampling process, we interpolate using constant interpolation, selecting for each interpolated timepoint the looking location for the earlier-observed time point in the original data for both `aoi_timepoints` and `xy_timepoints` data. Compared to linear interpolation (see e.g., Wass, Smith, & Johnson, 2013) – which fills segments of missing or unobserved time points by interpolating between the observed locations of timepoints at the beginning and end of the interpolated segment –, constant interpolation has the advantage that it is more conservative, in the sense that it does not introduce new look locations beyond those measured in the original data.

One possible application of our new dataset is investigating the consequences of other interpolation functions for data analysis.

301 Processing, Validation, and Ingestion

The `peekds` package offers functions to extract the above data. Once these data have been extracted in a tabular form, the package also offers a function to check whether all tables have the required fields and data types expected by the database. In an effort to double check the data quality and to make sure that no errors are made in the importing script, as part of the import procedure we create a time course plot based on our processed tables to replicate the results in the paper that first presented each dataset. Once this plot has been created and checked for consistency and all tables pass our validation functions, the

³⁰⁹ processed dataset is ready for reprocessing into the database using the `peekbank` library.
³¹⁰ This library applies additional data checks, and adds the data to the MySQL database using
³¹¹ the Django web framework.

³¹² Currently, the import process is carried out by the Peekbank team using data offered
³¹³ by other research teams. In the future, we hope to allow research teams to carry out their
³¹⁴ own import processes with checks from the Peekbank team before reprocessing. To this end,
³¹⁵ import script templates are available for both hand-coded datasets and automatic
³¹⁶ eye-tracking datasets for research teams to adapt to their data.

³¹⁷ Current Data Sources

Table 1
Overview of the datasets in the current database.

| Citation | Dataset name | N | Mean age (mos.) | Age range (mos.) | Method | Language |
|------------------------------------|-----------------|-----|-----------------|------------------|---------------|------------------|
| Adams et al., 2018 | ft_pt | 69 | 17.1 | 13–20 | manual coding | English |
| Byers-Heinlein et al., 2017 | mix | 48 | 20.1 | 19–21 | eye-tracking | English, French |
| Casillas et al., 2017 | tseltal | 23 | 31.3 | 9–48 | manual coding | Tseltal |
| Fernald et al., 2013 | fmw | 80 | 20.0 | 17–26 | manual coding | English |
| Frank et al., 2016 | tablet | 69 | 35.5 | 12–60 | eye-tracking | English |
| Garrison et al., 2020 | yoursmy | 35 | 14.5 | 12–18 | eye-tracking | English |
| Hurtado et al., 2007 | xsectional | 49 | 23.8 | 15–37 | manual coding | Spanish |
| Hurtado et al., 2008 | input_uptake | 76 | 21.0 | 17–27 | manual coding | Spanish |
| Mahr et al., 2015 | coartic | 29 | 20.8 | 18–24 | eye-tracking | English |
| Perry et al., 2017 | cowpig | 45 | 20.5 | 19–22 | manual coding | English |
| Pomper & Saffran, 2016 | switchingCues | 60 | 44.3 | 41–47 | manual coding | English |
| Pomper & Saffran, 2019 | salientme | 44 | 40.1 | 38–43 | manual coding | English |
| Potter & Lew-Williams, unpublished | canine | 36 | 23.8 | 21–27 | manual coding | English |
| Potter et al., 2019 | remix | 44 | 22.6 | 18–29 | manual coding | Spanish, English |
| Ronfard et al., 2021 | lsc | 40 | 20.0 | 18–24 | manual coding | English |
| Swingley & Aslin, 2002 | mispron | 50 | 15.1 | 14–16 | manual coding | English |
| Weisleder & Fernald, 2013 | stl | 29 | 21.6 | 18–27 | manual coding | Spanish |
| Yurovsky & Frank, 2017 | attword | 288 | 25.5 | 13–59 | eye-tracking | English |
| Yurovsky et al., 2013 | reflook_socword | 435 | 33.6 | 12–70 | eye-tracking | English |
| Yurovsky et al., unpublished | reflook_v4 | 45 | 34.2 | 11–60 | eye-tracking | English |

³¹⁸ The database currently includes 20 looking-while-listening datasets comprising $N=1594$
³¹⁹ total participants (Table 1). The current data represents a convenience sample of datasets
³²⁰ that were (a) datasets collected by or available to Peekbank team members, (b) made
³²¹ available to Peekbank after informal inquiry or (c) datasets that were openly available. Most
³²² datasets (14 out of 20 total) consist of data from monolingual native English speakers. They

323 span a wide age spectrum with participants ranging from 9 to 70 months of age, and are
324 balanced in terms of gender (47% female). The datasets vary across a number of
325 design-related dimensions, and include studies using manually coded video recordings and
326 automated eye-tracking methods (e.g., Tobii, EyeLink) to measure gaze behavior. All studies
327 tested familiar items, but the database also includes 5 datasets that tested novel
328 pseudo-words in addition to familiar words. Users interested in a subset of the data (e.g.,
329 only trials testing familiar words) can filter out unwanted trials using columns available in
330 the schema (e.g., using the column `stimulus_novelty`).

331 Versioning and Reproducibility

332 The content of Peekbank will change as we add additional datasets and revise previous
333 ones. To facilitate reproducibility of analyses, we use a versioning system by which
334 successive releases are assigned a name reflecting the year and version, e.g., 2022.1. By
335 default, users will interact with the most recent version of the database available, though the
336 `peekbankr` API allows researchers to run analyses against any previous version of the
337 database. For users with intensive use-cases, each version of the database may be
338 downloaded as a compressed .sql file and installed on a local MySQL server.

339 Peekbank allows for fully reproducible analyses using our source data, but the goal is
340 not to reproduce precisely the analyses – or even the datasets – in the publications whose
341 data we archive. Because of our emphasis on a standardized data importing and formatting
342 pipeline, there may be minor discrepancies in the timecourse data that we archive compared
343 with those reported in original publications. Further, we archive all of the data that are
344 provided to us – including participants that might have been excluded in the original studies,
345 if these data are available – rather than attempting to reproduce specific exclusion criteria.
346 We hope that Peekbank can be used as a basis for comparing different exclusion and filtering
347 criteria – as such, an inclusive policy regarding importing all available data helps us provide

348 a broad base of data for investigating these decisions.

349 **Interfacing with Peekbank**

350 **Peekbankr**

351 The `peekbankr` API offers a way for users to access data from the database and
352 flexibly analyze it in R. The majority of API calls simply allow users to download tables (or
353 subsets of tables) from the database. In particular, the package offers the following functions:

- 354 • `connect_to_peekbank()` opens a connection with the Peekbank database to allow
355 tables to be downloaded with the following functions
- 356 • `get_datasets()` gives each dataset name and its citation information
- 357 • `get_subjects()` gives information about persistent participant identifiers (e.g., native
358 languages, sex)
- 359 • `get_administrations()` gives information about specific experimental
360 administrations (e.g., participant age, monitor size, gaze coding method)
- 361 • `get_stimuli()` gives information about word–image pairings that appeared in
362 experiments
- 363 • `get_trial_types()` gives information about pairings of stimuli that appeared in the
364 experiment (e.g., point of disambiguation, target and distractor stimuli, condition,
365 language)
- 366 • `get_trials()` gives the trial orderings for each administration, linking trial types to
367 the trial IDs used in time course data
- 368 • `get_aoi_region_sets()` gives coordinate regions for each area of interest (AOI)
369 linked to trial type IDs
- 370 • `get_xy_timepoints()` gives time course data for each participant’s looking behavior
371 in each trial, as (x, y) coordinates on the experiment monitor

- 372 • `get_aoi_timepoints()` gives time course data for each participant's looking behavior
373 in each trial, coded into areas of interest

374 Once users have downloaded tables, they can be merged using `join` command via their
375 linked IDs. A set of standard merges are shown below in the "Peekbank in Action" section;
376 these allow the common use-case of examining time course data and metadata jointly.

377 Because of the size of the XY and AOI data tables, downloading data across multiple
378 studies can be time-consuming. Many of the most common analyses of the Peekbank data
379 require download of the `aoi_timepoints` table, thus we have put substantial work into
380 optimizing transfer times. In particular, `connect_to_peekbank` offers a data compression
381 option, and `get_aoi_timepoints` by default downloads time-courses via a compressed
382 (run-length encoded) representation, which is then uncompressed on the client side. More
383 information about these options (including how to modify them) can be found in the
384 package documentation.

385 **Shiny App**

386 One goal of the Peekbank project is to allow a wide range of users to easily explore and
387 learn from the database. We therefore have created an interactive web application –
388 `peekbank-shiny` – that allows users to quickly and easily create informative visualizations
389 of individual datasets and aggregated data. `peekbank-shiny` is built using Shiny, a software
390 package for creating web apps for data exploration with R, as well as the `peekbankr` package.
391 The Shiny app allows users to create commonly used visualizations of looking-while-listening
392 data, based on data from the Peekbank database. Specifically, users can visualize:

- 393 1. the time course of looking data in a profile plot depicting infant target looking across
394 trial time
- 395 2. overall accuracy (proportion target looking) within a specified analysis window

- 396 3. reaction times (speed of fixating the target image) in response to a target label
397 4. an onset-contingent plot, which shows the time course of participant looking as a
398 function of their look location at the onset of the target label

399 Users are given various customization options for each of these visualizations, e.g.,
400 choosing which datasets to include in the plots, controlling the age range of participants,
401 splitting the visualizations by age bins, and controlling the analysis window for time course
402 analyses. Plots are then updated in real time to reflect users' customization choices, and
403 users are given options to share the visualizations they created. A screenshot of the app is
404 shown in Figure 3. The Shiny app thus allows users to quickly inspect basic properties of
405 Peekbanks datasets and create reproducible visualizations without incurring any of the
406 technical overhead required to access the database through R.

407 **OSF site**

408 In addition to the Peekbank database proper, all data is openly available on the
409 Peekbank OSF webpage (<https://osf.io/pr6wu/>). The OSF site also includes the original raw
410 data (both time series data and metadata, such as trial lists and participant logs) that was
411 obtained for each study and subsequently processed into the standardized Peekbank format.
412 Users who are interested in inspecting or reproducing the processing pipeline for a given
413 dataset can use the respective import script (openly available on GitHub,
414 <https://github.com/langcog/peekbank-data-import>) to download and process the raw data
415 from OSF into its final standardized format. Where available, the OSF page also includes
416 additional information about the stimuli used in each dataset, including in some instances
417 the original stimulus sets (e.g., image and audio files).

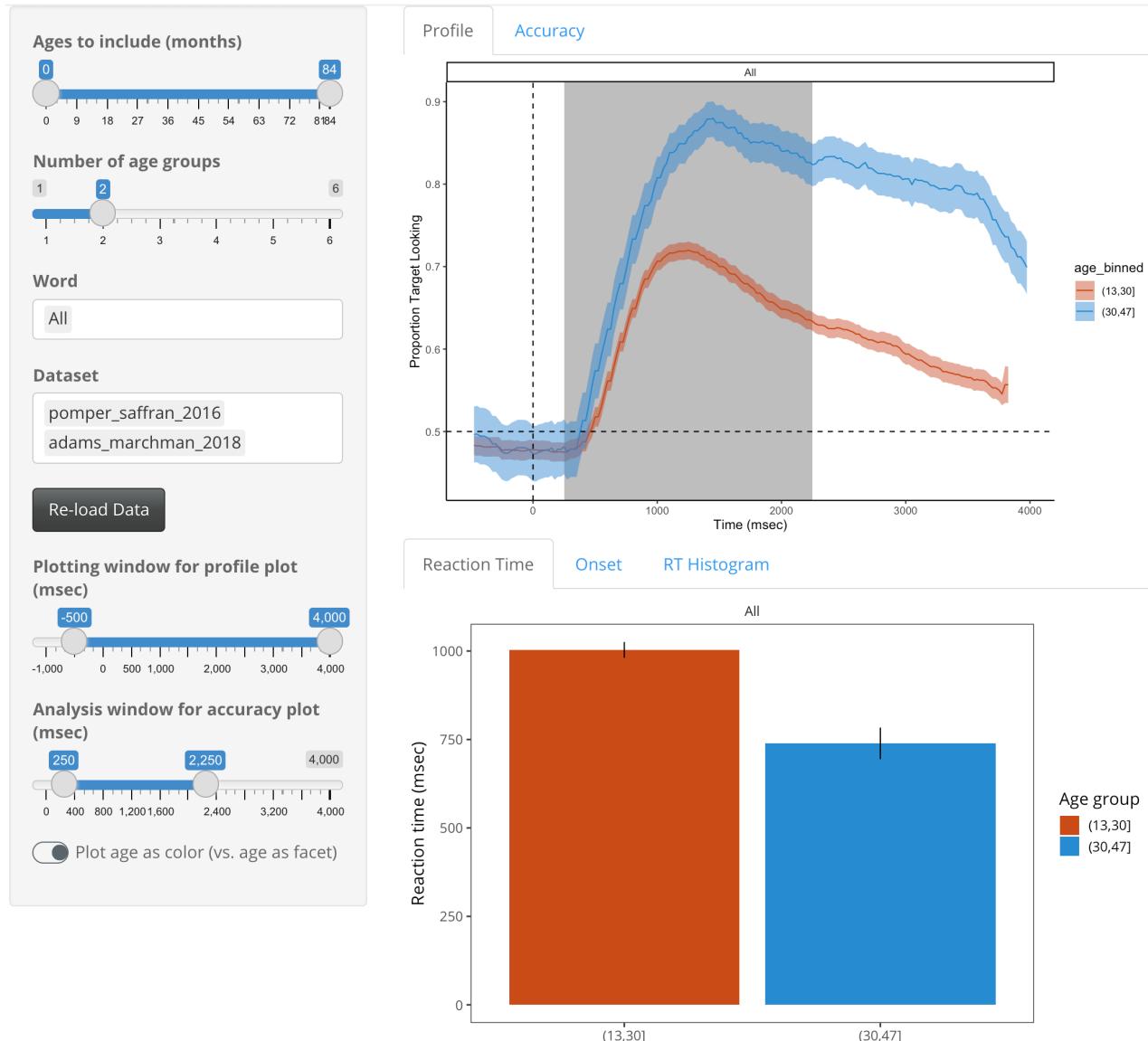


Figure 3. Screenshot of the Peekbank Shiny app, which shows a variety of standard analysis plots as a function of user-selected datasets, words, age ranges, and analysis windows. Shown here are mean reaction time and proportion target looking over time by age group for two selected datasets.

| Dataset Name | Unique Items | Prop. Target | 95% CI |
|-----------------|--------------|--------------|--------------|
| attword | 6 | 0.63 | [0.62, 0.65] |
| canine | 16 | 0.65 | [0.61, 0.68] |
| coartic | 10 | 0.71 | [0.68, 0.74] |
| cowpig | 12 | 0.61 | [0.58, 0.63] |
| fmw | 12 | 0.65 | [0.63, 0.67] |
| ft_pt | 8 | 0.65 | [0.63, 0.67] |
| input_uptake | 12 | 0.61 | [0.59, 0.63] |
| lsc | 8 | 0.69 | [0.65, 0.73] |
| mispron | 22 | 0.57 | [0.55, 0.59] |
| mix | 6 | 0.55 | [0.52, 0.58] |
| reflook_socword | 6 | 0.61 | [0.6, 0.63] |
| reflook_v4 | 10 | 0.61 | [0.57, 0.65] |
| remix | 8 | 0.63 | [0.58, 0.67] |
| salientme | 16 | 0.74 | [0.72, 0.75] |
| stl | 12 | 0.63 | [0.6, 0.66] |
| switchingCues | 40 | 0.77 | [0.75, 0.8] |
| tablet | 24 | 0.64 | [0.6, 0.68] |
| tseltal | 30 | 0.59 | [0.54, 0.63] |
| xsectional | 8 | 0.59 | [0.55, 0.63] |
| yoursmy | 87 | 0.60 | [0.56, 0.64] |

Table 2

Average proportion target looking in each dataset.

418

Peekbank in Action

419 In the following section, we provide examples of users can access and analyze the data
 420 in Peekbank. First, we provide an overview of some general properties of the datasets in the
 421 database. We then demonstrate two potential use-cases for Peekbank data. In each case, we
 422 provide sample code to demonstrate the ease of doing simple analyses using the database.
 423 Our first example shows how we can investigate the findings of a classic study. This type of
 424 investigation can be a very useful exercise for teaching students about best practices for data
 425 analysis (e.g., Hardwicke et al., 2018) and also provides an easy way to explore
 426 looking-while-listening time course data in a standardized format. Our second example
 427 shows an in-depth exploration of developmental changes in the recognition of particular
 428 words. Besides its theoretical interest (which we will explore more fully in subsequent work),
 429 this type of analysis could in principle be used for optimizing the stimuli for new
 430 experiments, especially as the Peekbank dataset grows and gains coverage over a greater

431 number of items. All analyses are conducted using R [Version 4.1.1; R Core Team (2021)]²

432 **General Descriptives**

433 One of the values of the uniform data format we use in Peekbank is the ease of
 434 providing cross-dataset descriptions that can give an overview of some of the general
 435 patterns found in our data. A first broad question is about the degree of accuracy in word
 436 recognition found across studies. In general, participants demonstrated robust, above-chance
 437 word recognition in each dataset (chance=0.5). Table 2 shows the average proportion of
 438 target looking within a standard critical window of 367-2000ms after the onset of the label
 439 for each dataset (Swingley & Aslin, 2002). Proportion target looking was generally higher for
 440 familiar words ($M = 0.66$, 95% CI = [0.65, 0.67], $n = 1543$) than for novel words learned
 441 during the experiment ($M = 0.59$, 95% CI = [0.58, 0.61], $n = 822$).

442 A second question of interest is about the variability across items (i.e., target labels)
 443 within specific studies. Some studies use a smaller set of items [e.g., 8 nouns; Adams et al.
 444 (2018)] while others use dozens of different items (e.g., Garrison, Baudet, Breitfeld, Aberman,
 445 & Bergelson, 2020). Figure 4 gives an overview of the variability in proportion looking to the
 446 target item for individual words in each dataset. Although all datasets show a gradual rise in
 447 average proportion target looking over chance performance, the number of unique target
 448 labels and their associated accuracy vary widely across datasets.

² We, furthermore, used the R-packages *dplyr* [Version 1.0.7; Wickham, François, Henry, and Müller (2021)], *forcats* [Version 0.5.1; Wickham (2021a)], *ggplot2* [Version 3.3.5; Wickham (2016)], *ggthemes* [Version 4.2.4; Arnold (2021)], *here* [Version 1.0.1; Müller (2020)], *papaja* [Version 0.1.0.9997; Aust and Barth (2020)], *peekbankr* [Version 0.1.1.9002; Braginsky, MacDonald, and Frank (2021)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *readr* [Version 2.0.1; Wickham and Hester (2021)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.1.4; Müller and Wickham (2021)], *tidyR* [Version 1.1.3; Wickham (2021b)], *tidyverse* [Version 1.3.1; Wickham et al. (2019)], *viridis* [Version 0.6.1; Garnier et al. (2021a); Garnier et al. (2021b)], *viridisLite* [Version 0.4.0; Garnier et al. (2021b)], and *xtable* [Version 1.8.4; Dahl, Scott, Roosen, Magnusson, and Swinton (2019)].

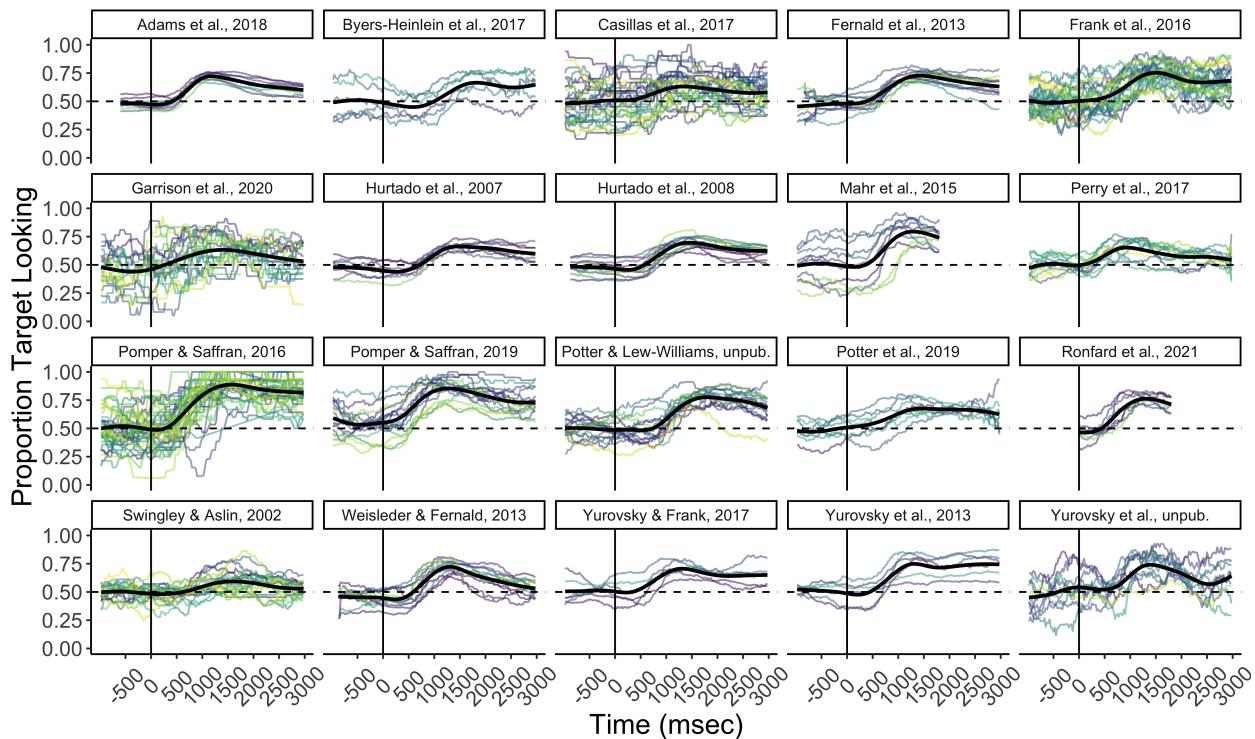


Figure 4. Item-level variability in proportion target looking within each dataset (chance=0.5). Time is centered on the onset of the target label (vertical line). Colored lines represent specific target labels. Black lines represent smoothed average fits based on a general additive model using cubic splines.

449 Investigating prior findings: Swingley and Aslin (2002)

450 Swingley and Aslin (2002) investigated the specificity of 14-16 month-olds' word
 451 representations using the looking-while-listening paradigm, asking whether recognition would
 452 be slower and less accurate for mispronunciations, e.g. *oppel* (close mispronunciation) or *opel*
 453 (distant mispronunciation) instead of *apple* (correct pronunciation). In this short vignette,
 454 we show how easily the data in Peekbank can be used to visualize this result. Our goal here
 455 is not to provide a precise analytical reproduction of the analyses reported in the original
 456 paper, but rather to demonstrate the use of the Peekbank framework to analyze datasets of
 457 this type. In particular, because Peekbank uses a uniform data import standard, it is likely
 458 that there will be minor numerical discrepancies between analyses on Peekbank data and
 459 analyses that use another processing pipeline.

```
library(peekbankr)
aoi_timepoints <- get_aoi_timepoints(dataset_name = "swingley_aslin_2002")
administrations <- get_administrations(dataset_name = "swingley_aslin_2002")
trial_types <- get_trial_types(dataset_name = "swingley_aslin_2002")
trials <- get_trials(dataset_name = "swingley_aslin_2002")
```

460 We begin by retrieving the relevant tables from the database, `aoi_timepoints`,
 461 `administrations`, `trial_types`, and `trials`. As discussed above, each of these can be
 462 downloaded using a simple API call through `peekbankr`, which returns dataframes that
 463 include ID fields. These ID fields allow for easy joining of the data into a single dataframe
 464 containing all the information necessary for the analysis.

```
swingley_data <- aoi_timepoints |>
  left_join(administrations) |>
  left_join(trials) |>
  left_join(trial_types) |>
  filter(condition != "filler") |>
  mutate(condition = if_else(condition == "cp", "Correct", "Mispronounced"))
```

465 As the code above shows, once the data are joined, condition information for each
 466 timepoint is present and so we can easily filter out filler trials and set up the conditions for
 467 further analysis. For simplicity, here we combine both mispronunciation conditions since the
 468 close vs. distant mispronunciation manipulation showed no effect in the original paper.

```
accuracies <- swingley_data |>
  group_by(condition, t_norm, administration_id) |>
  summarize(correct = sum(aoi == "target") /
    sum(aoi %in% c("target", "distractor"))) |>
  group_by(condition, t_norm) |>
  summarize(mean_correct = mean(correct),
            ci = 1.96 * sd(correct) / sqrt(n()))
```

469 The final step in our analysis is to create a summary dataframe using `dplyr`
 470 commands. We first group the data by timestep, participant, and condition and compute the
 471 proportion looking at the correct image. We then summarize again, averaging across

⁴⁷² participants, computing both means and 95% confidence intervals (via the approximation of
⁴⁷³ 1.96 times the standard error of the mean). The resulting dataframe can be used for
⁴⁷⁴ visualization of the time course of looking.

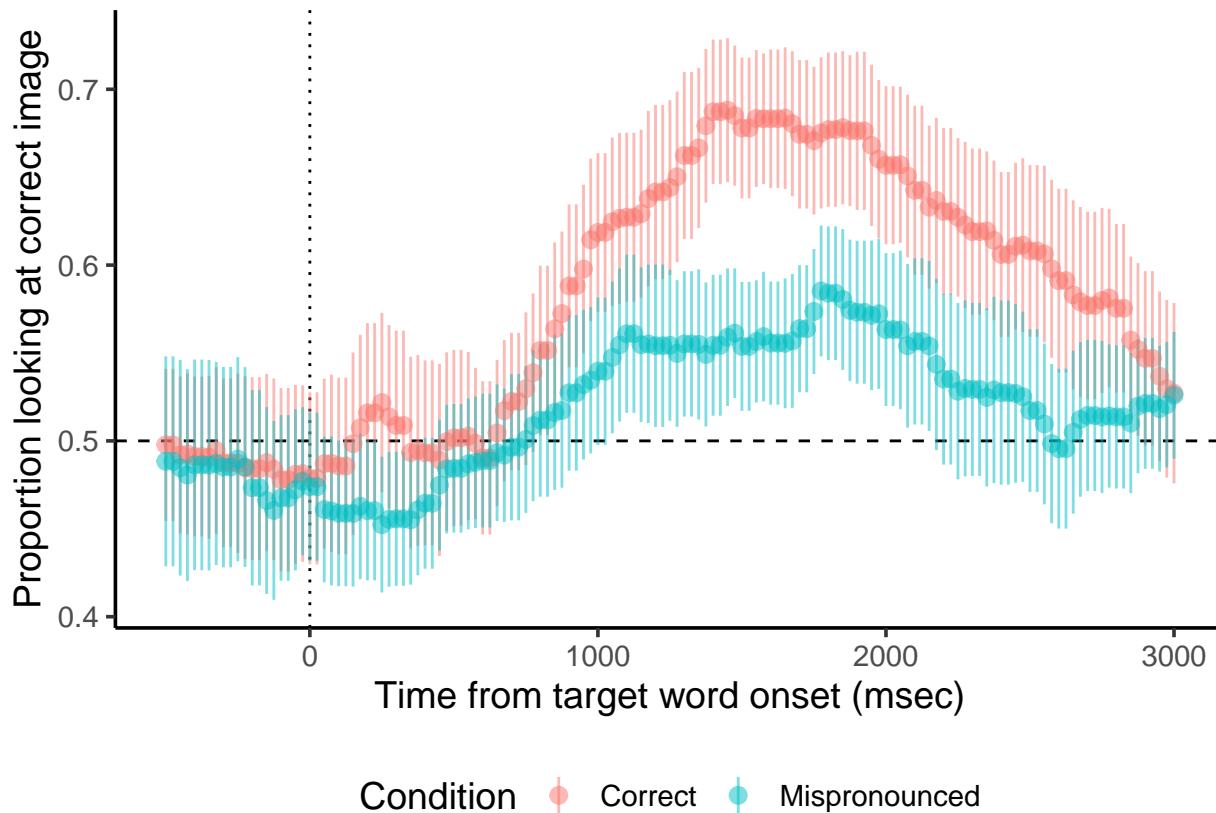


Figure 5. Proportion looking at the correct referent by time from the point of disambiguation (the onset of the target noun) in Swingley & Aslin (2002). Colors show the two pronunciation conditions; points give means and ranges show 95% confidence intervals. The dotted line shows the point of disambiguation and the dashed line shows chance performance.

⁴⁷⁵ Figure 5 shows the average time course of looking for the two conditions, as produced
⁴⁷⁶ by the code above. Looks after the correctly pronounced noun appeared both faster
⁴⁷⁷ (deviating from chance earlier) and more accurate (showing a higher asymptote). Overall,
⁴⁷⁸ this example demonstrates the ability to produce this visualization in just a few lines of code.

479 **Item analyses**

480 A second use case for Peekbank is to examine item-level variation in word recognition.
 481 Individual datasets rarely have enough statistical power to show reliable developmental
 482 differences within items. To illustrate the power of aggregating data across multiple datasets,
 483 we select the four words with the most data available across studies and ages (apple, book,
 484 dog, and frog) and show average recognition trajectories.

485 Our first step is to collect and join the data from the relevant tables including
 486 timepoint data, trial and stimulus data, and administration data (for participant ages). We
 487 join these into a single dataframe for easy manipulation; this dataframe is a common
 488 starting point for analyses of item-level data.

```
all_aoi_timepoints <- get_aoi_timepoints()
all_stimuli <- get_stimuli()
all_administrations <- get_administrations()
all_trial_types <- get_trial_types()
all_trials <- get_trials()

aoi_data_joined <- all_aoi_timepoints |>
  right_join(all_administrations) |>
  right_join(all_trials) |>
  right_join(all_trial_types) |>
  mutate(stimulus_id = target_id) |>
  right_join(all_stimuli) |>
  select(administration_id, english_stimulus_label, age, t_norm, aoi)
```

489 Next we select a set of four target words (chosen based on having more than 100
 490 children contributing data for each word across several one-year age groups). We create age

491 groups, aggregate, and compute timepoint-by-timepoint confidence intervals using the z
 492 approximation.

```
target_words <- c("book", "dog", "frog", "apple")

target_word_data <- aoi_data_joined |>
  filter(english_stimulus_label %in% target_words) |>
  mutate(age_group = cut(age, breaks = seq(12, 48, 12))) |>
  filter(!is.na(age_group)) |>
  group_by(t_norm, administration_id, age_group, english_stimulus_label) |>
  summarise(correct = mean(aoi == "target") /
    mean(aoi %in% c("target", "distractor"), na.rm=TRUE)) |>
  group_by(t_norm, age_group, english_stimulus_label) |>
  summarise(ci = 1.96 * sd(correct, na.rm=TRUE) / sqrt(length(correct)),
            correct = mean(correct, na.rm=TRUE),
            n = n())
```

493 Finally, we plot the data as time courses split by age. Our plotting code is shown
 494 below (with styling commands again removed for clarity). Figure 6 shows the resulting plot,
 495 with time courses for each of three (rather coarse) age bins. Although some baseline effects
 496 are visible across items, we still see clear and consistent increases in looking to the target,
 497 with the increase appearing earlier and in many cases asymptoting at a higher level for older
 498 children. On the other hand, this simple averaging approach ignores study-to-study variation
 499 (perhaps responsible for the baseline effects we see in the *apple* and *frog* items especially). In
 500 future work, we hope to introduce model-based analytic methods that use mixed effects
 501 regression to factor out study-level and individual-level variance in order to recover
 502 developmental effects more appropriately (see e.g., Zettersten et al., 2021 for a prototype of
 503 such an analysis).

```
ggplot(target_word_data,
       aes(x = t_norm, y = correct, col = age_group)) +
  geom_line() +
  geom_linerange(aes(ymin = correct - ci, ymax = correct + ci),
                 alpha = .2) +
  facet_wrap(~english_stimulus_label)
```

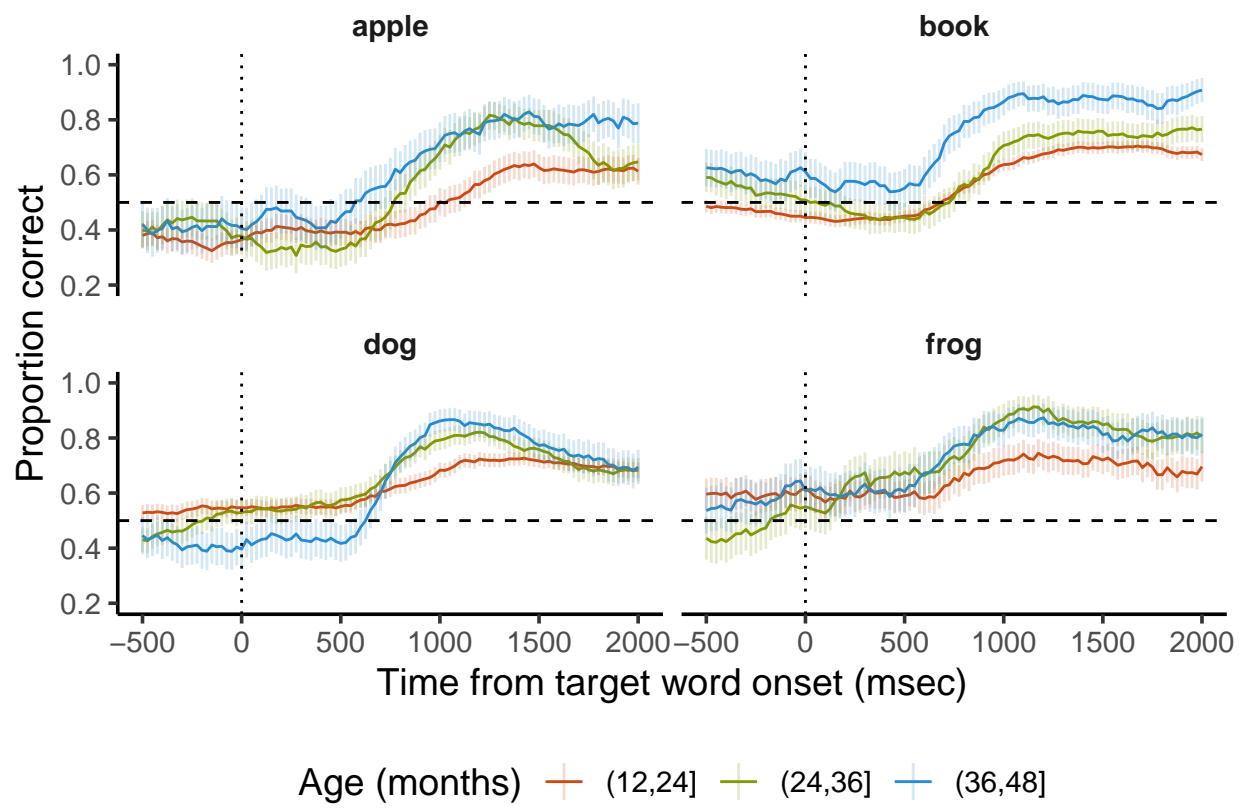


Figure 6. Time course plot for four well-represented target items in the Peekbank dataset, split by three age groups. Each line represents children's average looking to the target image after the onset of the target label (dashed vertical line). Error bars represent 95% CIs.

504

Discussion

505 Theoretical progress in understanding child development requires rich datasets, but
 506 collecting child data is expensive, difficult, and time-intensive. Recent years have seen a
 507 growing effort to build open source tools and pool research efforts to meet the challenge of

508 building a cumulative developmental science (Bergmann et al., 2018; Frank, Braginsky,
509 Yurovsky, & Marchman, 2017; Sanchez et al., 2019; The ManyBabies Consortium, 2020).
510 The Peekbank project expands on these efforts by building an infrastructure for aggregating
511 eye-tracking data across studies, with a specific focus on the looking-while-listening
512 paradigm. This paper presents an overview of the structure of the database, shows how users
513 can access the database, and demonstrates how it can be used both to investigate prior
514 experiments and to synthesize data across studies.

515 The current database has a number of limitations, particularly in its number and
516 diversity of datasets. With 20 datasets currently available in the database, idiosyncrasies of
517 particular designs and condition manipulations still have substantial influence on modeling
518 results. Expanding the set of distinct datasets will allow us to increase the number of
519 observations per item across datasets, leading to more robust generalizations across item-level
520 variability. The current database is also limited by the relatively homogeneous background of
521 its participants, both with respect to language (almost entirely monolingual native English
522 speakers) and cultural background (Henrich, Heine, & Norenzayan, 2010; Muthukrishna et
523 al., 2020). Increasing the diversity of participant backgrounds and languages will expand the
524 scope of the generalizations we can form about child word recognition.

525 Finally, while the current database is focused on studies of word recognition, the tools
526 and infrastructure developed in the project can in principle be used to accommodate any
527 eye-tracking paradigm, opening up new avenues for insights into cognitive development. Gaze
528 behavior has been at the core of many of the key advances in our understanding of infant
529 cognition (Bergelson & Swingley, 2012; Fernald, Pinto, Swingley, Weinberg, & McRoberts,
530 1998; Lew-Williams & Fernald, 2007; Weisleder & Fernald, 2013; Yurovsky & Frank, 2017).
531 Aggregating large datasets of infant looking behavior in a single, openly-accessible format
532 promises to bring a fuller picture of infant cognitive development into view.

CRediT author statement

533

534 Outside of the position of the first and the last author, authorship position was
535 determined by sorting authors' last names in reverse alphabetical order. An overview over
536 authorship contributions following the CRediT taxonomy can be viewed here:
537 https://docs.google.com/spreadsheets/d/e/2PACX-1vRD-LJD_dTAQaAynyBlwXvGpfAVzP-3Pi6JTDoG15m3PYZe0c44Y12U2a_hwdmhIstpjyigG2o3na4y/pubhtml.
538

Acknowledgements

539

540 We would like to thank the labs and researchers that have made their data publicly
541 available in the database. For further information about contributions, see
542 <https://langcog.github.io/peekbank-website/docs/contributors/>.

References

- 543 Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., & Feldman,
544 H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes
545 in full term and preterm toddlers. *Child Development*, 89(5), 1674–1690.
- 546
- 547 Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*.
548 Retrieved from <https://CRAN.R-project.org/package=ggthemes>
- 549 Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*.
550 Retrieved from <https://github.com/crsh/papaja>
- 551 Bergelson, E. (2020). The comprehension boost in early word learning: Older infants
552 are better learners. *Child Development Perspectives*, 14(3), 142–149.
- 553 Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the
554 meanings of many common nouns. *PNAS*, 109(9), 3253–3258.
- 555 Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young
556 infants. *Cognition*, 127(3), 391–397.
- 557 Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C.,
558 & Cristia, A. (2018). Promoting replicability in developmental research through
559 meta-analyses: Insights from language acquisition research. *Child Development*,
560 89(6), 1996–2009.
- 561 Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early
562 productive vocabulary predicts academic achievement 10 years later. *Applied
563 Psycholinguistics*, 37(6), 1461–1476.

- 564 Braginsky, M., MacDonald, K., & Frank, M. (2021). *Peekbankr: Accessing the*
565 *peekbank database*. Retrieved from <http://github.com/langcog/peekbankr>
- 566 Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable
567 infant research. *PsyArXiv*. <https://doi.org/https://doi.org/10.31234/osf.io/ksfvq>
- 568 Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable:*
569 *Export tables to LaTeX or HTML*. Retrieved from
570 <https://CRAN.R-project.org/package=xtable>
- 571 DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in
572 infant research: A case study of the effect of number of infants and number of
573 trials in visual preference procedures. *Infancy*, 25(4), 393–419.
574 <https://doi.org/10.1111/infa.12337>
- 575 Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language
576 processing skill and vocabulary are evident at 18 months. *Developmental Science*,
577 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- 578 Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998).
579 Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological*
580 *Science*, 9(3), 228–231.
- 581 Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while
582 listening: Using eye movements to monitor spoken language comprehension by
583 infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen
584 (Eds.), *Developmental psycholinguistics: On-line methods in children's language*
585 *processing* (pp. 97–135). Amsterdam: John Benjamins.
- 586 Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ...
587 Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting

- 588 Reproducibility, Best Practices, and Theory-Building. *Infancy*, 22(4), 421–435.
- 589 <https://doi.org/10.1111/infa.12182>
- 590 Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank:
- 591 An open repository for developmental vocabulary data. *Journal of Child*
- 592 *Language*, 44(3), 677–694.
- 593 Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability*
- 594 and *Consistency in Early Language Learning: The Wordbank Project*. Cambridge,
- 595 MA: MIT Press.
- 596 Garnier, Simon, Ross, Noam, Rudis, Robert, . . . Cédric. (2021a). *viridis -*
- 597 *colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>
- 598 Garnier, Simon, Ross, Noam, Rudis, Robert, . . . Cédric. (2021b). *viridis -*
- 599 *colorblind-friendly color maps for r*. <https://doi.org/10.5281/zenodo.4679424>
- 600 Garrison, H., Baudet, G., Breitfeld, E., Aberman, A., & Bergelson, E. (2020).
- 601 Familiarity plays a small role in noun comprehension at 12–18 months. *Infancy*,
- 602 25(4), 458–477.
- 603 Gautheron, L., Rochat, N., & Cristia, A. (under review). Managing, storing, and
- 604 sharing long-form recordings and their annotations. Retrieved from
- 605 <https://doi.org/10.31234/osf.io/w8trm>
- 606 Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years
- 607 using the intermodal preferential looking paradigm to study language acquisition:
- 608 What have we learned? *Perspectives on Psychol. Science*, 8(3), 316–339.
- 609 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P.,
- 610 . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for

- organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., ... Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, 5(8). <https://doi.org/10.1098/rsos.180448>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Retrieved from <https://CRAN.R-project.org/package=purrr>
- Hirsh-Pasek, K., Cauley, K. M., Golinkoff, R. M., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language*, 14(1), 23–45.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by Latino children learning Spanish as their first language. *Journal of Child Language*, 34(2), 227–249. <https://doi.org/10.1017/S0305000906007896>
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11(6), 31–39. <https://doi.org/10.1111/j.1467-7687.2008.00768.x>
- Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., & Frank, M. C. (2016). *A Quantitative Synthesis of Early Language Acquisition*

- 635 *Using Meta-Analysis.* <https://doi.org/10.31234/osf.io/htsjm>
- 636 Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid
637 use of grammatical gender in spoken word recognition. *Psychological Science*,
638 18(3), 193–198.
- 639 Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015).
640 Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*,
641 142, 345–350. <https://doi.org/10.1016/j.cognition.2015.05.009>
- 642 Marchman, V. A., Loi, E. C., Adams, K. A., Ashland, M., Fernald, A., & Feldman, H.
643 M. (2018). Speed of language comprehension at 18 months old predicts
644 school-relevant outcomes at 54 months old in children born preterm. *Journal of
645 Dev. & Behav. Pediatrics*, 39(3), 246–253.
- 646 Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A.,
647 McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich,
648 and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of
649 Cultural and Psychological Distance. *Psychological Science*, 31(6), 678–701.
- 650 Müller, K. (2020). *Here: A simpler way to find your files*. Retrieved from
651 <https://CRAN.R-project.org/package=here>
- 652 Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from
653 <https://CRAN.R-project.org/package=tibble>
- 654 Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A.,
655 . . . Vazire, S. (2021). Replicability, Robustness, and Reproducibility in
656 Psychological Science. *PsyArXiv*.
657 <https://doi.org/https://doi.org/10.31234/osf.io/ksfvq>

658 Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F.
659 (2019). Does speed of processing or vocabulary size predict later language growth
660 in toddlers? *Cognitive Psychology*, 115, 101238.

661 Potter, C., & Lew-Williams, C. (unpublished). Behold the canine!: How does
662 toddlers' knowledge of typical frames and familiar words interact to influence their
663 sentence processing?

664 R Core Team. (2021). *R: A language and environment for statistical computing*.
665 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
666 <https://www.R-project.org/>

667 Ronfard, S., Wei, R., & Rowe, M. L. (2021). Exploring the linguistic, cognitive, and
668 social skills underlying lexical processing efficiency as measured by the
669 looking-while-listening paradigm. *Journal of Child Language*, 1–24.
670 <https://doi.org/10.1017/S0305000921000106>

671 Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank,
672 M. C. (2019). childe-db: A flexible and reproducible interface to the child
673 language data exchange system. *Behavior Research Methods*, 51(4), 1928–1941.
674 <https://doi.org/10.3758/s13428-018-1176-7>

675 Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form
676 representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
677 <https://doi.org/10.1111/1467-9280.00485>

678 The ManyBabies Consortium. (2020). Quantifying sources of variability in infancy
679 research using the infant-directed speech preference. *Advances in Methods and*
680 *Practices in Psychological Science*, 3(1), 24–52.

- 681 Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of
682 variable quality to provide accurate fixation duration estimates in infants and
683 adults. *Behavior Research Methods*, 45(1), 229–250.
684 <https://doi.org/10.3758/s13428-012-0245-6>
- 685 Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language
686 Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*,
687 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- 688 Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag
689 New York. Retrieved from <https://ggplot2.tidyverse.org>
- 690 Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string*
691 operations. Retrieved from <https://CRAN.R-project.org/package=stringr>
- 692 Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors)*.
693 Retrieved from <https://CRAN.R-project.org/package=forcats>
- 694 Wickham, H. (2021b). *Tidyr: Tidy messy data*. Retrieved from
695 <https://CRAN.R-project.org/package=tidyr>
- 696 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ...
697 Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*,
698 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 699 Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of*
700 *data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 701 Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data*. Retrieved from
702 <https://CRAN.R-project.org/package=readr>

703 Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: salience and
704 social cues in early word learning. *Dev Sci*, 20(2).

705 Yurovsky, D., Wade, A., Kraus, A. M., Gengoux, G. W., Hardan, A. Y., & Frank, M.
706 C. (under review). Developmental changes in the speed of social attention in early
707 word learning.

708 Zettersten, M., Bergey, C., Bhatt, N., Boyce, V., Braginsky, M., Carstensen, A., ...
709 others. (2021). Peekbank: Exploring children's word recognition through an open,
710 large-scale repository for developmental eye-tracking data.