

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

25 **Data and Code Availability.** All code for reproducing the paper is available at
26 <https://github.com/langcog/peekbank-paper>. Raw and standardized datasets are available
27 on the Peekbank OSF repository (<https://osf.io/pr6wu/>) and can be accessed using the
28 peekbankr R package (<https://github.com/langcog/peekbankr>).

29 **CRediT author statement.** Outside of the position of the first and the last author,
30 authorship position was determined by sorting authors' last names in reverse alphabetical
31 order. An overview over authorship contributions following the CRediT taxonomy can be
32 viewed here: https://docs.google.com/spreadsheets/d/e/2PACX-1vRD-LJD_dTAQaAynyBlwXvGpfAVzP-3Pi6JTDG15m3PYZe0c44Y12U2a_hwdmhIstpjyigG2o3na4y/pubhtml.

34 Correspondence concerning this article should be addressed to Martin Zettersten,
35 Department of Psychology, Princeton University, 218 Peretsman Scully Hall, Princeton, NJ
36 08540. E-mail: martincz@princeton.edu

37

Abstract

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

49 *Keywords:* word recognition; eye-tracking; vocabulary development;
50 looking-while-listening; visual world paradigm; lexical processing

51 Word count: 6600

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

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

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

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

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

95 Measuring Word Recognition: The Looking-While-Listening Paradigm

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

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

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

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

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

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

146 Replicability and Reproducibility

147 A core challenge facing psychology in general, and the study of infant development in
148 particular, are threats to the replicability and reproducibility of core empirical results (Frank
149 et al., 2017; Nosek et al., 2022). In infant research, many studies are not adequately powered
150 to detect the main effects of interest (Bergmann et al., 2018). This issues is compounded by
151 low reliability in infant measures, often due to limits on the number of trials that can be
152 collected from an individual infant in an experimental session (Byers-Heinlein, Bergmann, &
153 Savalei, 2021). One hurdle to improving power in infant research is that it can be difficult to

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

162 In addition to threats to replicability, the field of infant development also faces
163 concerns about analytic *reproducibility* – the ability for researchers to arrive at the same
164 analytic conclusion reported in the original research article, given the same dataset. A recent
165 estimate based on studies published in a prominent cognitive science journal suggests that
166 analyses can remain difficult to reproduce, even when data are made available to other
167 research teams (Hardwicke et al., 2018). Aggregating data in centralized databases can aid
168 in improving reproducibility in several ways. First, building a large-scale database requires
169 defining a standardized data specification. Recent examples include the **brain imaging**
170 **data structure** (BIDS), an effort to specify a unified data format for neuroimaging
171 experiments (Gorgolewski et al., 2016), and the data formats associated with **ChildProject**,
172 for managing long-form at-home language recordings (Gautheron, Rochat, & Cristia, 2021).
173 Defining a data standard – in this case, for infant eye-tracking experiments – supports
174 reproducibility by guaranteeing that critical information will be available in openly shared
175 data and by making it easier for different research teams to understand the data structure.
176 Second, open databases make it easy for researchers to generate open and reproducible
177 analytic pipelines, both for individual studies and for analyses aggregating across datasets.
178 Creating open analytic pipelines across many datasets also serves a pedagogical purpose,
179 providing teaching examples illustrating how to implement analytic techniques used in
180 influential studies and how to conduct reproducible analyses with infant eye-tracking data.

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

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

193 Design and Technical Approach**194 Database Framework**

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

200 As illustrated in Figure 1, the Peekbank framework consists of four main components:
201 (1) a set of tools to *convert* eye-tracking datasets into a unified format, (2) a relational
202 database populated with data in this unified format, (3) a set of tools to *retrieve* data from
203 this database, and (4) a web app (using the Shiny framework) for visualizing the data. These

204 components are supported by three packages. The `peekds` package (for the R language, R
205 Core Team, 2021) helps researchers convert existing datasets to use the standardized format
206 of the database. The `peekbank` module (Python) creates a database with the relational
207 schema and populates it with the standardized datasets produced by `peekds`. The database
208 is served through MySQL, an industry standard relational database server, which may be
209 accessed by a variety of programming languages, and can be hosted on one machine and
210 accessed by many others over the Internet. As is common in relational databases, records of
211 similar types (e.g., participants, trials, experiments, coded looks at each timepoint) are
212 grouped into tables, and records of various types are linked through numeric identifiers. The
213 `peekbankr` package (R) provides an application programming interface, or API, that offers
214 high-level abstractions for accessing the tabular data stored in Peekbank. Most users will
215 access data through this final package, in which case the details of data formatting,
216 processing, and the specifics of connecting to the database are abstracted away from the user.

217 Database Schema

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

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

226 *Participant Information.*

227 Invariant information about individuals who participate in one or more studies (e.g, a

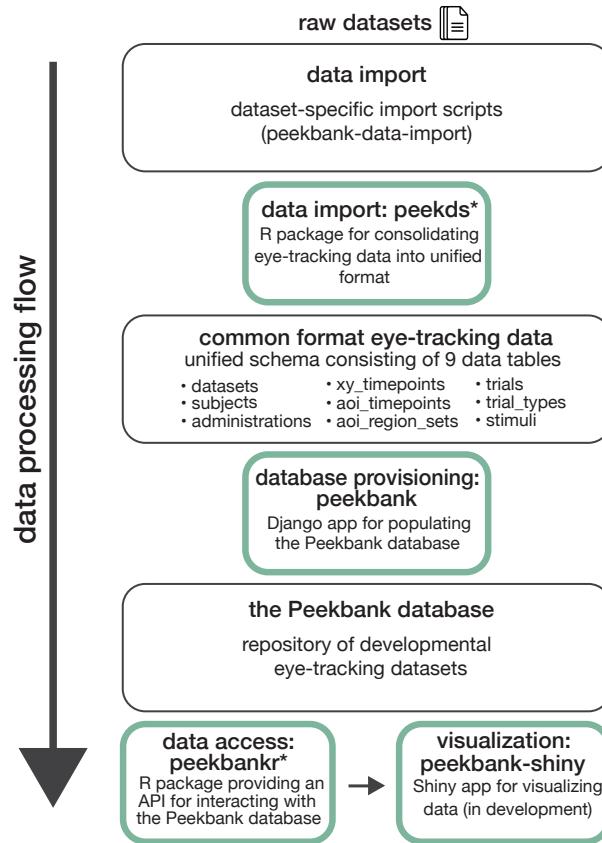


Figure 1. Overview of the Peekbank data ecosystem. Peekbank tools are highlighted in green.
* indicates R packages introduced in this work.

228 participant's first language) is recorded in the **subjects** table, while the **administrations**
 229 table contains information about each individual session in a given study (see Session
 230 Information, below). This division allows Peekbank to gracefully handle longitudinal designs:
 231 a single participant can complete multiple sessions and thus be associated with multiple
 232 administrations.

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

236 ***Experiment Information.***

237 The **datasets** table includes information about the lab conducting the study and the

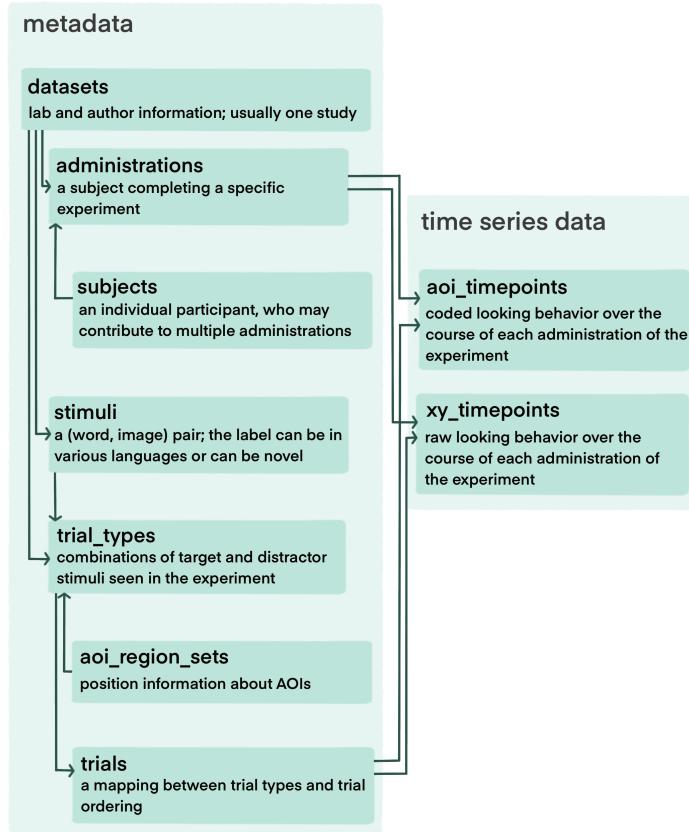


Figure 2. The Peekbank schema. Each darker rectangle represents a table in the relational database.

238 relevant publications to cite regarding the data. In most cases, a dataset corresponds to a
 239 single study.

240 Information about the experimental design is split across the `trial_types` and
 241 `stimuli` tables. The `trial_types` table encodes information about each trial *in the design*
 242 *of the experiment*,^{\footnote{We note that the term *trial* is ambiguous and could be used to}
 243 refer to both a particular combination of stimuli seen by many participants and a participant
 244 seeing that particular combination at a particular point in the experiment. We track the
 245 former in the `trial_types` table and the latter in the `trials` table.[}] including the target
 246 stimulus and location (left vs. right), the distractor stimulus and location, and the point of
 247 disambiguation for that trial. If a dataset used automatic eye-tracking rather than manual
 248 coding, each trial type is additionally linked to a set of area of interest (x, y) coordinates,

249 encoded in the `aoi_region_sets` table. The `trial_types` table links trial types to the
250 `aoi_region_sets` table and the `trials` table. Each trial_type record links to two records
251 in the `stimuli` table, identified by the `distractor_id` and the `target_id` fields.

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

261 ***Session Information.***

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

266 ***Trial Information.***

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

271 **Time course data.** Raw looking data is a series of looks to areas of interest (AOIs),
272 such as looks to the left or right of the screen, or to (x, y) coordinates on the experiment

273 screen, linked to points in time. For data generated by eye-trackers, we typically have (x, y)
274 coordinates at each time point, which we encode in the `xy_timepoints` table. These looks
275 are also recoded into AOIs according to the AOI coordinates in the `aoi_region_sets` table
276 using the `add_aois()` function in `peekds`, and encoded in the `aoi_timepoints` table. For
277 hand-coded data, we typically have a series of AOIs (i.e., looks to the left vs. right of the
278 screen), but lack information about exact gaze positions on-screen; in these cases the AOIs
279 are recoded into the categories in the Peekbank schema (target, distractor, other, and
280 missing) and encoded in the `aoi_timepoints` table; however, these datasets do not have any
281 corresponding data in the `xy_timepoints` table.

282 Typically, timepoints in the `xy_timepoints` table and `aoi_timepoints` table need to
283 be regularized to center each trial's time around the point of disambiguation – such that 0 is
284 the time of target word onset in the trial (i.e., the beginning of *dog* in *Can you find the*
285 *dog?*). We re-centered timing information to the onset of the target label to facilitate
286 comparison of target label processing across all datasets.¹ If time values run throughout the
287 experiment rather than resetting to zero at the beginning of each trial, `rezero_times()` is
288 used to reset the time at each trial. After this, each trial's times are centered around the
289 point of disambiguation using `normalize_times()`. When these steps are complete, the
290 time course is ready for resampling.

291 To facilitate time course analysis and visualization across datasets, time course data
292 must be resampled to a uniform sampling rate (i.e., such that every trial in every dataset has
293 observations at the same time points). All data in the database is resampled to 40 Hz
294 (observations every 25 ms), which represents a compromise between retaining fine-grained
295 timing information from datasets with dense sampling rates (maximum sampling rate among

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.

296 current datasets: 500 Hz) while minimizing the possibility of introducing artifacts via
297 resampling for datasets with lower sampling rates (minimum sampling rate for current
298 datasets: 30 Hz). Further, 25 ms is a mathematically convenient interval for ensuring
299 consistent resampling; we found that using 33.333 ms (30 Hz) as our interval simply
300 introduced a large number of technical complexities. The resampling operation is
301 accomplished using the `resample_times()` function. During the resampling process, we
302 interpolate using constant interpolation, selecting for each interpolated timepoint the looking
303 location for the earlier-observed time point in the original data for both `aoi_timepoints`
304 and `xy_timepoints` data. Compared to linear interpolation (see e.g., Wass, Smith, &
305 Johnson, 2013) – which fills segments of missing or unobserved time points by interpolating
306 between the observed locations of timepoints at the beginning and end of the interpolated
307 segment –, constant interpolation has the advantage that it is more conservative, in the sense
308 that it does not introduce new look locations beyond those measured in the original data.
309 One possible application of our new dataset is investigating the consequences of other
310 interpolation functions for data analysis.

311 Processing, Validation, and Ingestion

312 The `peekds` package offers functions to extract the above data. Once these data have
313 been extracted in a tabular form, the package also offers a function to check whether all
314 tables have the required fields and data types expected by the database. In an effort to
315 double check the data quality and to make sure that no errors are made in the importing
316 script, as part of the import procedure we create a time course plot based on our processed
317 tables to replicate the results in the paper that first presented each dataset. Once this plot
318 has been created and checked for consistency and all tables pass our validation functions, the
319 processed dataset is ready for reprocessing into the database using the `peekbank` library.
320 This library applies additional data checks, and adds the data to the MySQL database using

³²¹ the Django web framework.

³²² Currently, the import process is carried out by the Peekbank team using data offered
³²³ by other research teams. In the future, we hope to allow research teams to carry out their
³²⁴ own import processes with checks from the Peekbank team before reprocessing. To this end,
³²⁵ import script templates are available for both hand-coded datasets and automatic
³²⁶ eye-tracking datasets for research teams to adapt to their data.

³²⁷ 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

³²⁸ The database currently includes 20 looking-while-listening datasets comprising $N=1594$
³²⁹ total participants (Table 1). The current data represents a convenience sample of datasets
³³⁰ that were (a) datasets collected by or available to Peekbank team members, (b) made
³³¹ available to Peekbank after informal inquiry or (c) datasets that were openly available. Most
³³² datasets (14 out of 20 total) consist of data from monolingual native English speakers. They
³³³ span a wide age spectrum with participants ranging from 9 to 70 months of age, and are
³³⁴ balanced in terms of gender (47% female). The datasets vary across a number of

335 design-related dimensions, and include studies using manually coded video recordings and
336 automated eye-tracking methods (e.g., Tobii, EyeLink) to measure gaze behavior. All studies
337 tested familiar items, but the database also includes 5 datasets that tested novel
338 pseudo-words in addition to familiar words. Users interested in a subset of the data (e.g.,
339 only trials testing familiar words) can filter out unwanted trials using columns available in
340 the schema (e.g., using the column `stimulus_novelty`).

341 Versioning and Reproducibility

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

349 Peekbank allows for fully reproducible analyses using our source data, but the goal is
350 not to reproduce precisely the analyses – or even the datasets – in the publications whose
351 data we archive. Because of our emphasis on a standardized data importing and formatting
352 pipeline, there may be minor discrepancies in the timecourse data that we archive compared
353 with those reported in original publications. Further, we archive all of the data that are
354 provided to us – including participants that might have been excluded in the original studies,
355 if these data are available – rather than attempting to reproduce specific exclusion criteria.
356 We hope that Peekbank can be used as a basis for comparing different exclusion and filtering
357 criteria – as such, an inclusive policy regarding importing all available data helps us provide
358 a broad base of data for investigating these decisions.

359

Interfacing with Peekbank

360 **Peekbankr**

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

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

384 Once users have downloaded tables, they can be merged using `join` command via their

385 linked IDs. A set of standard merges are shown below in the “Peekbank in Action” section;

386 these allow the common use-case of examining time course data and metadata jointly.

387 Because of the size of the XY and AOI data tables, downloading data across multiple

388 studies can be time-consuming. Many of the most common analyses of the Peekbank data

389 require download of the `aoi_timepoints` table, thus we have put substantial work into

390 optimizing transfer times. In particular, `connect_to_peekbank` offers a data compression

391 option, and `get_aoi_timepoints` by default downloads time-courses via a compressed

392 (run-length encoded) representation, which is then uncompressed on the client side. More

393 information about these options (including how to modify them) can be found in the

394 package documentation.

395 Shiny App

396 One goal of the Peekbank project is to allow a wide range of users to easily explore and

397 learn from the database. We therefore have created an interactive web application –

398 `peekbank-shiny` – that allows users to quickly and easily create informative visualizations

399 of individual datasets and aggregated data. `peekbank-shiny` is built using Shiny, a software

400 package for creating web apps for data exploration with R, as well as the `peekbankr`

401 package. All code for the Shiny app is publicly available

402 (<https://github.com/langcog/peekbank-shiny>). The Shiny app allows users to create

403 commonly used visualizations of looking-while-listening data, based on data from the

404 Peekbank database. Specifically, users can visualize:

405 1. the *time course of looking data* in a profile plot depicting infant target looking across

406 trial time

407 2. *overall accuracy*, defined as the proportion target looking within a specified analysis

408 window

- 409 3. *reaction times* in response to a target label, defined as how quickly participants shift
410 fixation to the target image on trials in which they were fixating the distractor image
411 at onset of the target label
- 412 4. an *onset-contingent plot*, which shows the time course of participant looking as a
413 function of their look location at the onset of the target label

414 Users are given various customization options for each of these visualizations, e.g.,

415 choosing which datasets to include in the plots, controlling the age range of participants,
416 splitting the visualizations by age bins, and controlling the analysis window for time course
417 analyses. Plots are then updated in real time to reflect users' customization choices, and
418 users are given options to share the visualizations they created. A screenshot of the app is
419 shown in Figure 3. The Shiny app thus allows users to quickly inspect basic properties of
420 Peekbanks datasets and create reproducible visualizations without incurring any of the
421 technical overhead required to access the database through R.

422 **OSF site**

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

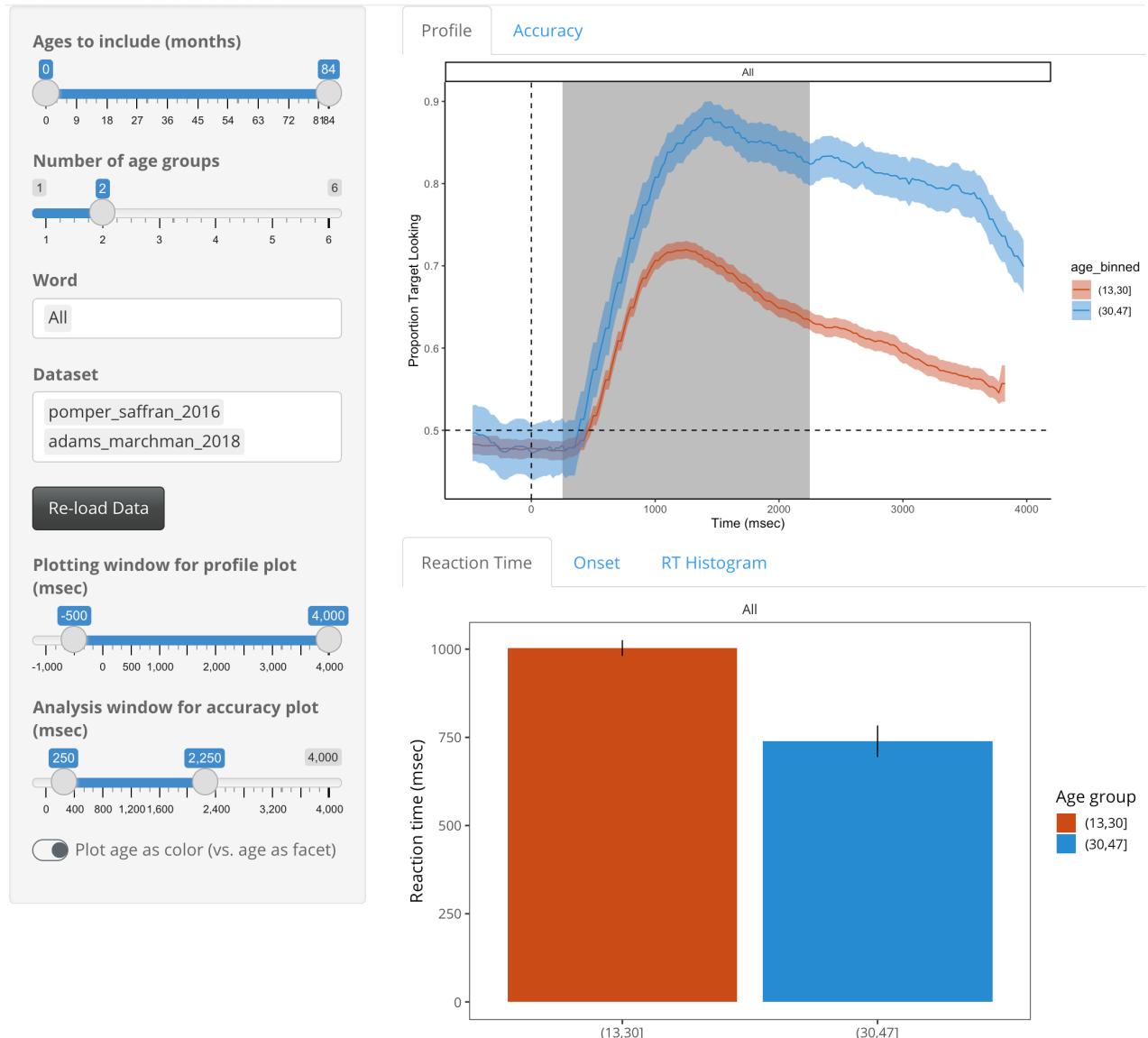


Figure 3. Screenshot of the Peekbank Shiny app, which shows a variety of standard analysis plots as a function of user-selected datasets, words, age ranges, and analysis windows. Shown here are mean reaction time and proportion target looking over time by age group for two selected datasets.

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.

433

Peekbank in Action

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

⁴⁴⁶ number of items. All analyses are conducted using R [Version 4.1.1; R Core Team (2021)]²

⁴⁴⁷ General Descriptives

⁴⁴⁸ One of the values of the uniform data format we use in Peekbank is the ease of
⁴⁴⁹ providing cross-dataset descriptions that can give an overview of some of the general
⁴⁵⁰ patterns found in our data. A first broad question is about the degree of accuracy in word
⁴⁵¹ recognition found across studies. In general, participants demonstrated robust, above-chance
⁴⁵² word recognition in each dataset (chance=0.5). Table 2 shows the average proportion of
⁴⁵³ target looking within a standard critical window of 367-2000ms after the onset of the label
⁴⁵⁴ for each dataset (Swingley & Aslin, 2002). Proportion target looking was generally higher for
⁴⁵⁵ familiar words ($M = 0.66$, 95% CI = [0.65, 0.67], $n = 1543$) than for novel words learned
⁴⁵⁶ during the experiment ($M = 0.59$, 95% CI = [0.58, 0.61], $n = 822$).

⁴⁵⁷ A second question of interest is about the variability across items (i.e., target labels)
⁴⁵⁸ within specific studies. Some studies use a smaller set of items [e.g., 8 nouns; Adams et al.
⁴⁵⁹ (2018)] while others use dozens of different items (e.g., Garrison, Baudet, Breitfeld, Aberman,
⁴⁶⁰ & Bergelson, 2020). Figure 4 gives an overview of the variability in proportion looking to the
⁴⁶¹ target item for individual words in each dataset. Although all datasets show a gradual rise in
⁴⁶² average proportion target looking over chance performance, the number of unique target
⁴⁶³ 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)].

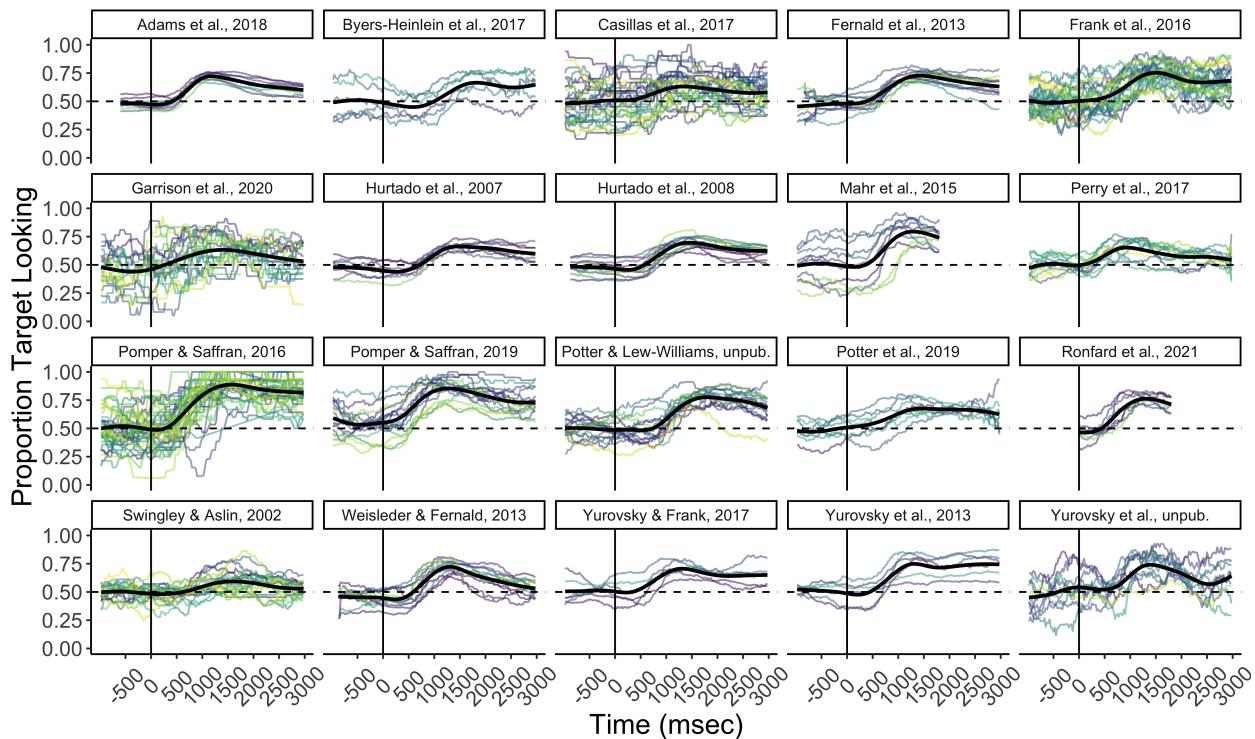


Figure 4. Item-level variability in proportion target looking within each dataset (chance=0.5). Time is centered on the onset of the target label (vertical line). Colored lines represent specific target labels. Black lines represent smoothed average fits based on a general additive model using cubic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

464 Investigating prior findings: Swingley and Aslin (2002)

465 Swingley and Aslin (2002) investigated the specificity of 14-16 month-olds' word
 466 representations using the looking-while-listening paradigm, asking whether recognition would
 467 be slower and less accurate for mispronunciations, e.g. *opal* (mispronunciation) instead of
 468 *apple* (correct pronunciation).³ In this short vignette, we show how easily the data in
 469 Peekbank can be used to visualize this result. Our goal here is not to provide a precise
 470 analytical reproduction of the analyses reported in the original paper, but rather to
 471 demonstrate the use of the Peekbank framework to analyze datasets of this type. In
 472 particular, because Peekbank uses a uniform data import standard, it is likely that there will

³ The original paper investigated both close (e.g., *opple*, /apl/) and distant (e.g., *opal*, /opl/) mispronunciations. For simplicity, here we combine both mispronunciation conditions since the close vs. distant mispronunciation manipulation showed no effect in the original paper.

- 473 be minor numerical discrepancies between analyses on Peekbank data and analyses that use
 474 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")
```

- 475 We begin by retrieving the relevant tables from the database, `aoi_timepoints`,
 476 `administrations`, `trial_types`, and `trials`. As discussed above, each of these can be
 477 downloaded using a simple API call through `peekbankr`, which returns dataframes that
 478 include ID fields. These ID fields allow for easy joining of the data into a single dataframe
 479 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"))
```

- 480 As the code above shows, once the data are joined, condition information for each
 481 timepoint is present and so we can easily filter out filler trials and set up the conditions for
 482 further analysis.

```
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()))
```

- 483 The final step in our analysis is to create a summary dataframe using `dplyr`
 484 commands. We first group the data by timestep, participant, and condition and compute the

485 proportion looking at the correct image. We then summarize again, averaging across
 486 participants, computing both means and 95% confidence intervals (via the approximation of
 487 1.96 times the standard error of the mean). The resulting dataframe can be used for
 488 visualization of the time course of looking.

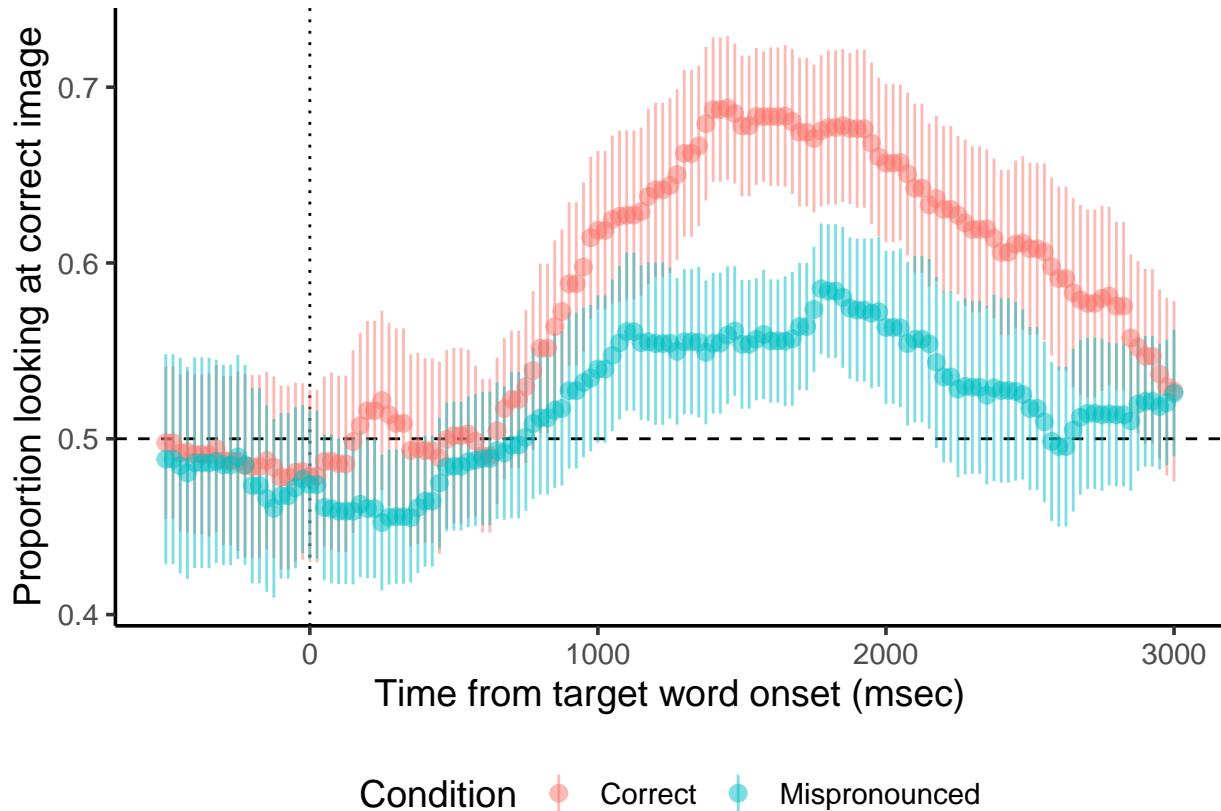


Figure 5. Proportion looking at the correct referent by time from the point of disambiguation (the onset of the target noun) in Swingley & Aslin (2002). Colors show the two pronunciation conditions; points give means and ranges show 95% confidence intervals. The dotted line shows the point of disambiguation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

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

493 **Item analyses**

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

499 Our first step is to collect and join the data from the relevant tables including
 500 timepoint data, trial and stimulus data, and administration data (for participant ages). We
 501 join these into a single dataframe for easy manipulation; this dataframe is a common
 502 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)
```

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

505 groups, aggregate, and compute timepoint-by-timepoint confidence intervals using the z
 506 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())
```

507 Finally, we plot the data as time courses split by age. Our plotting code is shown
 508 below (with styling commands again removed for clarity). Figure 6 shows the resulting plot,
 509 with time courses for each of three (rather coarse) age bins. Although some baseline effects
 510 are visible across items, we still see clear and consistent increases in looking to the target,
 511 with the increase appearing earlier and in many cases asymptoting at a higher level for older
 512 children. On the other hand, this simple averaging approach ignores study-to-study variation
 513 (perhaps responsible for the baseline effects we see in the *apple* and *frog* items especially). In
 514 future work, we hope to introduce model-based analytic methods that use mixed effects
 515 regression to factor out study-level and individual-level variance in order to recover
 516 developmental effects more appropriately (see e.g., Zettersten et al., 2021 for a prototype of
 517 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)
```

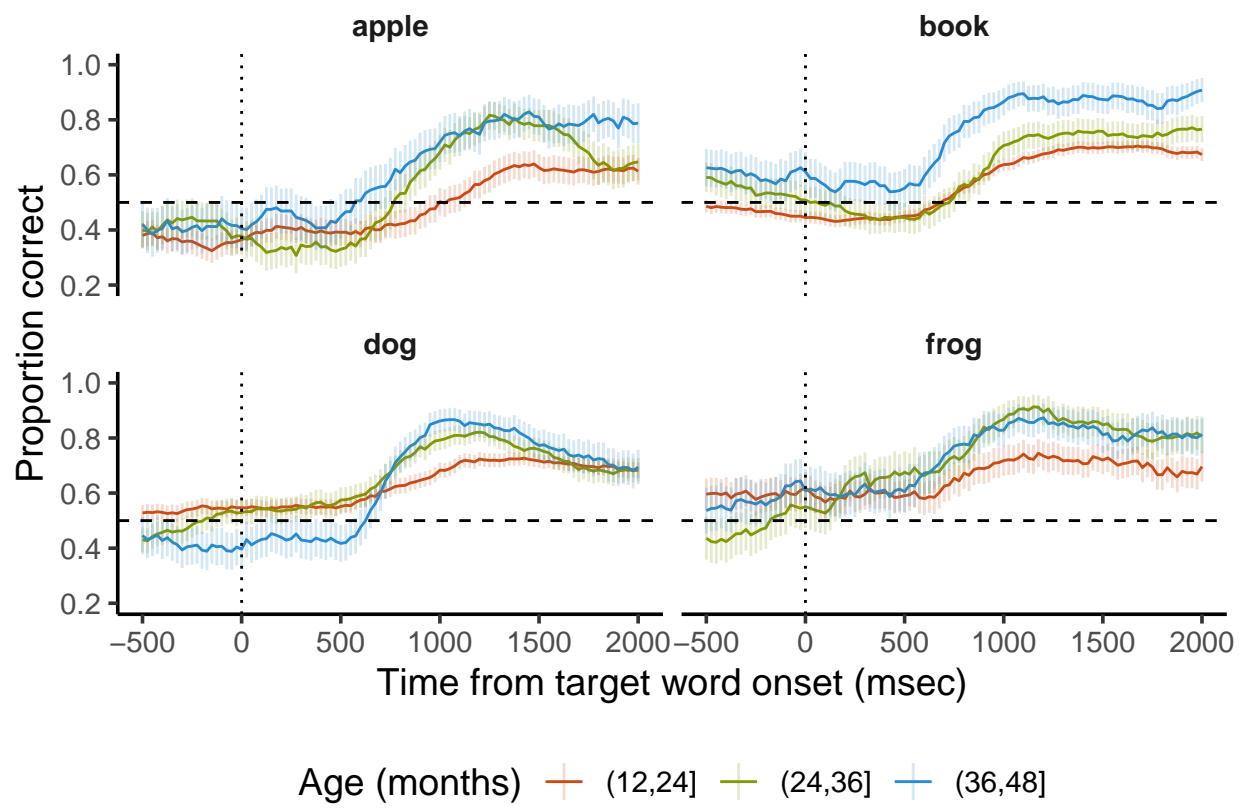


Figure 6. Time course plot for four well-represented target items in the Peekbank dataset, split by three age groups. Each line represents children's average looking to the target image after the onset of the target label (dashed vertical line). Error bars represent 95% CIs.

518

Discussion

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

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

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

539 Finally, while the current database is focused on studies of word recognition, the tools
540 and infrastructure developed in the project can in principle be used to accommodate any
541 eye-tracking paradigm, opening up new avenues for insights into cognitive development. Gaze
542 behavior has been at the core of many of the key advances in our understanding of infant
543 cognition (Bergelson & Swingley, 2012; Fernald, Pinto, Swingley, Weinberg, & McRoberts,
544 1998; Lew-Williams & Fernald, 2007; Weisleder & Fernald, 2013; Yurovsky & Frank, 2017).
545 Aggregating large datasets of infant looking behavior in a single, openly-accessible format
546 promises to bring a fuller picture of infant cognitive development into view.

References

- 547 Adams, K. A., Marchman, V. A., Loi, E. C., Ashland, M. D., Fernald, A., & Feldman,
548 H. M. (2018). Caregiver talk and medical risk as predictors of language outcomes
549 in full term and preterm toddlers. *Child Development*, 89(5), 1674–1690.
- 551 Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*.
552 Retrieved from <https://CRAN.R-project.org/package=ggthemes>
- 553 Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*.
554 Retrieved from <https://github.com/crsh/papaja>
- 555 Bergelson, E. (2020). The comprehension boost in early word learning: Older infants
556 are better learners. *Child Development Perspectives*, 14(3), 142–149.
- 557 Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the
558 meanings of many common nouns. *Proceedings of the National Academy of
559 Sciences*, 109(9), 3253–3258.
- 560 Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young
561 infants. *Cognition*, 127(3), 391–397.
- 562 Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C.,
563 & Cristia, A. (2018). Promoting replicability in developmental research through
564 meta-analyses: Insights from language acquisition research. *Child Development*,
565 89(6), 1996–2009.
- 566 Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early
567 productive vocabulary predicts academic achievement 10 years later. *Applied
568 Psycholinguistics*, 37(6), 1461–1476.

- 569 Braginsky, M., MacDonald, K., & Frank, M. (2021). *Peekbankr: Accessing the*
570 *peekbank database*. Retrieved from <http://github.com/langcog/peekbankr>
- 571 Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more
572 reliable infant research. *Infant and Child Development*.
- 573 <https://doi.org/https://doi.org/10.1002/icd.2296>
- 574 Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable:*
575 *Export tables to LaTeX or HTML*. Retrieved from
576 <https://CRAN.R-project.org/package=xtable>
- 577 DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in
578 infant research: A case study of the effect of number of infants and number of
579 trials in visual preference procedures. *Infancy*, 25(4), 393–419.
- 580 <https://doi.org/10.1111/infa.12337>
- 581 Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language
582 processing skill and vocabulary are evident at 18 months. *Developmental Science*,
583 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- 584 Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998).
585 Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological*
586 *Science*, 9(3), 228–231.
- 587 Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while
588 listening: Using eye movements to monitor spoken language comprehension by
589 infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen
590 (Eds.), *Developmental psycholinguistics: On-line methods in children's language*
591 *processing* (pp. 97–135). Amsterdam: John Benjamins.
- 592 Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ...

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

- 616 . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for
617 organizing and describing outputs of neuroimaging experiments. *Scientific Data*,
618 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- 619 Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C.,
620 Kidwell, M. C., . . . Frank, M. C. (2018). Data availability, reusability, and
621 analytic reproducibility: Evaluating the impact of a mandatory open data policy
622 at the journal *Cognition*. *Royal Society Open Science*, 5(8).
623 <https://doi.org/10.1098/rsos.180448>
- 624 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?
625 *Behavioral and Brain Sciences*, 33(2-3), 61–83.
626 <https://doi.org/10.1017/S0140525X0999152X>
- 627 Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools*. Retrieved
628 from <https://CRAN.R-project.org/package=purrr>
- 629 Hirsh-Pasek, K., Cauley, K. M., Golinkoff, R. M., & Gordon, L. (1987). The eyes
630 have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child
631 Language*, 14(1), 23–45.
- 632 Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by
633 Latino children learning Spanish as their first language. *Journal of Child
634 Language*, 34(2), 227–249. <https://doi.org/10.1017/S0305000906007896>
- 635 Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake?
636 Links between maternal talk, processing speed and vocabulary size in
637 Spanish-learning children. *Developmental Science*, 11(6), 31–39.
638 <https://doi.org/10.1111/j.1467-7687.2008.00768.x>
- 639 Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., &

- Frank, M. C. (2016). *A Quantitative Synthesis of Early Language Acquisition Using Meta-Analysis*. *PsyArXiv*. <https://doi.org/10.31234/osf.io/htsjm>
- Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18(3), 193–198.
- Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350. <https://doi.org/10.1016/j.cognition.2015.05.009>
- Marchman, V. A., Loi, E. C., Adams, K. A., Ashland, M., Fernald, A., & Feldman, H. M. (2018). Speed of language comprehension at 18 months old predicts school-relevant outcomes at 54 months old in children born preterm. *Journal of Developmental & Behavioral Pediatrics*, 39(3), 246–253.
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678–701.
- Müller, K. (2020). *Here: A simpler way to find your files*. Retrieved from <https://CRAN.R-project.org/package=here>
- Müller, K., & Wickham, H. (2021). *Tibble: Simple data frames*. Retrieved from <https://CRAN.R-project.org/package=tibble>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748. <https://doi.org/https://doi.org/10.31234/osf.io/ksfvq>

- 664 Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F.
665 (2019). Does speed of processing or vocabulary size predict later language growth
666 in toddlers? *Cognitive Psychology*, 115, 101238.
- 667 Potter, C., & Lew-Williams, C. (unpublished). Behold the canine!: How does
668 toddlers' knowledge of typical frames and familiar words interact to influence their
669 sentence processing?
- 670 R Core Team. (2021). *R: A language and environment for statistical computing*.
671 Vienna, Austria: R Foundation for Statistical Computing. Retrieved from
672 <https://www.R-project.org/>
- 673 Ronfard, S., Wei, R., & Rowe, M. L. (2021). Exploring the linguistic, cognitive, and
674 social skills underlying lexical processing efficiency as measured by the
675 looking-while-listening paradigm. *Journal of Child Language*, 1–24.
676 <https://doi.org/10.1017/S0305000921000106>
- 677 Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank,
678 M. C. (2019). childe-db: A flexible and reproducible interface to the child
679 language data exchange system. *Behavior Research Methods*, 51(4), 1928–1941.
680 <https://doi.org/10.3758/s13428-018-1176-7>
- 681 Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form
682 representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.
683 <https://doi.org/10.1111/1467-9280.00485>
- 684 The ManyBabies Consortium. (2020). Quantifying sources of variability in infancy
685 research using the infant-directed speech preference. *Advances in Methods and*
686 *Practices in Psychological Science*, 3(1), 24–52.
- 687 Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of

- variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, 45(1), 229–250.
<https://doi.org/10.3758/s13428-012-0245-6>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>
- Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations*. Retrieved from <https://CRAN.R-project.org/package=stringr>
- Wickham, H. (2021a). *Forcats: Tools for working with categorical variables (factors)*. Retrieved from <https://CRAN.R-project.org/package=forcats>
- Wickham, H. (2021b). *Tidyr: Tidy messy data*. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Hester, J. (2021). *Readr: Read rectangular text data*. Retrieved from <https://CRAN.R-project.org/package=readr>
- Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, 20(2), e12349.

711 Yurovsky, D., Wade, A., Kraus, A. M., Gengoux, G. W., Hardan, A. Y., & Frank, M.
712 C. (unpublished). Developmental changes in the speed of social attention in early
713 word learning.

714 Zettersten, M., Bergey, C., Bhatt, N., Boyce, V., Braginsky, M., Carstensen, A., ...
715 Frank, M. C. (2021). Peekbank: Exploring children's word recognition through an
716 open, large-scale repository for developmental eye-tracking data. *Proceedings of
717 the 43rd Annual Conference of the Cognitive Science Society.*