

<sup>1</sup> Peekbank: Exploring children's word recognition through an open, large-scale repository for  
<sup>2</sup> developmental eye-tracking data

<sup>3</sup> Martin Zettersten<sup>1</sup>, Claire Bergey<sup>2</sup>, Naiti S. Bhatt<sup>3</sup>, Veronica Boyce<sup>4</sup>, Mika Braginsky<sup>5</sup>,  
<sup>4</sup> Alexandra Carstensen<sup>4</sup>, Benny deMayo<sup>1</sup>, Kunal Handa<sup>12</sup>, George Kachergis<sup>4</sup>, Molly Lewis<sup>6</sup>,  
<sup>5</sup> Bria Long<sup>4</sup>, Kyle MacDonald<sup>7</sup>, Jessica Mankewitz<sup>4</sup>, Stephan Meylan<sup>5,8</sup>, Annissa N. Saleh<sup>9</sup>,  
<sup>6</sup> Rose M. Schneider<sup>10</sup>, Angeline Sin Mei Tsui<sup>4</sup>, Sarp Uner<sup>8</sup>, Tian Linger Xu<sup>11</sup>, Daniel  
<sup>7</sup> Yurovsky<sup>6</sup>, & Michael C. Frank<sup>4</sup>

<sup>8</sup> <sup>1</sup> Dept. of Psychology, Princeton University

<sup>9</sup> <sup>2</sup> Dept. of Psychology, University of Chicago

<sup>10</sup> <sup>3</sup> Scripps College

<sup>11</sup> <sup>4</sup> Dept. of Psychology, Stanford University

<sup>12</sup> <sup>5</sup> Dept. of Brain and Cognitive Sciences, MIT

<sup>13</sup> <sup>6</sup> Dept. of Psychology, Carnegie Mellon University

<sup>14</sup> <sup>7</sup> Core Technology, McD Tech Labs

<sup>15</sup> <sup>8</sup> Dept. of Psychology and Neuroscience, Duke University

<sup>16</sup> <sup>9</sup> Dept. of Psychology, UT Austin

<sup>17</sup> <sup>10</sup> Dept. of Psychology, UC San Diego

<sup>18</sup> <sup>11</sup> Dept. of Psychological and Brain Sciences, Indiana University

<sup>19</sup> <sup>12</sup> Brown University



20

## Abstract

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

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

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

33 Word count: X

- 34 Peekbank: Exploring children's word recognition through an open, large-scale repository for  
35 developmental eye-tracking data

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

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

word recognition are typically limited in scope to experiments in individual labs involving small samples tested on a handful of items. The limitations of single datasets makes it difficult to understand developmental changes in children’s word knowledge at a broad scale. One way to overcome this challenge is to compile existing datasets into a large-scale database in order to expand the scope of research questions that can be asked about the development word recognition abilities. This strategy capitalizes on the fact that the looking-while-listening paradigm is widely used, and vast amounts of data have been collected across labs on infants’ word recognition over the past 35 years (Golinkoff, Ma, Song, & Hirsh-Pasek, 2013). Such datasets have largely remained isolated from one another, but once combined, they have the potential to offer insights into the lexical development at a broad scale. Similar efforts in language development have born fruit in recent years. For example, WordBank aggregated data from the MacArthur-Bates Communicative Development Inventory, a parent-report measure of child vocabulary, to deliver new insights into cross-linguistic patterns and variability in vocabulary development (Frank, Braginsky, Yurovsky, & Marchman, 2017, 2021). In this paper, we introduce *Peekbank*, an open database of infant and toddler eye-tracking data aimed at facilitating the study of developmental changes in children’s word knowledge, especially recognition speed.

## 77 The “Looking-While-Listening” Paradigm

Word recognition is traditionally studied in the “looking-while-listening” paradigm (Fernald et al., 2008; alternatively referred to as the intermodal preferential looking procedure, Hirsh-Pasek, Cauley, Golinkoff, & Gordon, 1987). In such studies, infants listen to a sentence prompting a specific referent (e.g., *Look at the dog!*) while viewing two images on the screen (e.g., an image of a dog – the target image – and an image of a bird – the distractor image). Infants’ word recognition is measured in terms of how quickly and accurately they fixate on the correct target image after hearing its label. Past research has used this same basic method to study a wide range of questions in language development.

<sup>86</sup> For example, the looking-while-listening paradigm has been used to investigate early noun  
<sup>87</sup> knowledge, phonological representations of words, prediction during language processing, and  
<sup>88</sup> individual differences in language development (Bergelson & Swingley, 2012; Golinkoff et al.,  
<sup>89</sup> 2013; Lew-Williams & Fernald, 2007; Marchman et al., 2018; Swingley & Aslin, 2002).

<sup>90</sup> While this research has been fruitful in advancing understanding of early word  
<sup>91</sup> knowledge, fundamental questions remain. One central question is how to accurately capture  
<sup>92</sup> developmental change in the speed and accuracy of word recognition. There is ample  
<sup>93</sup> evidence demonstrating that infants get faster and more accurate in word recognition over  
<sup>94</sup> the first few years of life (e.g., Fernald et al., 1998). However, precisely measuring  
<sup>95</sup> developmental increases in the speed and accuracy of word recognition remains challenging  
<sup>96</sup> due to the difficulty of distinguishing developmental changes in word recognition skill from  
<sup>97</sup> changes in knowledge of specific words. This problem is particularly thorny in studies with  
<sup>98</sup> young children, since the number of items that can be tested within a single session is  
<sup>99</sup> limited and items must be selected in an age-appropriate manner (Peter et al., 2019).  
<sup>100</sup> Another potential challenge are that differences in the design choices and analytic decisions  
<sup>101</sup> within single studies could obscure changes when comparing individual studies at different  
<sup>102</sup> developmental time points.

<sup>103</sup> One approach to addressing these challenges is to conduct meta-analyses  
<sup>104</sup> aggregating effects across studies while testing for heterogeneity due to researcher choices  
<sup>105</sup> (Bergmann et al., 2018; Lewis et al., 2016). However, meta-analyses typically lack the  
<sup>106</sup> granularity to estimate participant-level and item-level variation or to model behavior  
<sup>107</sup> beyond coarse-grained effect size estimates. An alternative way to approach this challenge is  
<sup>108</sup> to aggregate trial-level data from smaller studies measuring word recognition with a wide  
<sup>109</sup> range of items and design choices into a large-scale dataset that can be analyzed using a  
<sup>110</sup> unified modeling approach. A sufficiently large dataset would allow researchers to estimate  
<sup>111</sup> developmental change in word recognition speed and accuracy while generalizing across

112 changes related to specific words or the design features of particular studies.

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

## 126 Replicability and Reproducibility

127 A core challenge facing psychology in general, and the study of infant development in  
128 particular, are threats to the replicability and reproducibility of core empirical results (Frank  
129 et al., 2017; Nosek et al., 2021). In infant research, many studies are not adequately powered  
130 to detect the main effects of interest (Bergmann et al., 2018). These issues are often  
131 compounded by low reliability in infant measures, often due to limits on the number of trials  
132 that can be collected from an individual infant in an experimental session (Byers-Heinlein,  
133 Bergmann, & Savalei, 2021). One hurdle to improving the power in infant research is that it  
134 can often be difficult to develop a priori estimates of effect sizes, and how specific design  
135 decisions (e.g., the number of test trials) will impact power and reliability. Large-scale  
136 databases of infant behavior can aid researchers' in their decision-making by providing rich  
137 datasets that can help constrain expectations about possible effect sizes and can be used to

138 make data-driven design decisions. For example, if a researcher is interested in  
139 understanding how the number of test trials could impact the power and reliability of their  
140 looking-while-listening design, a large-scale database would allow them to simulate possible  
141 outcomes across a range of test trials, based on past eye-tracking data with infants.

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

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

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

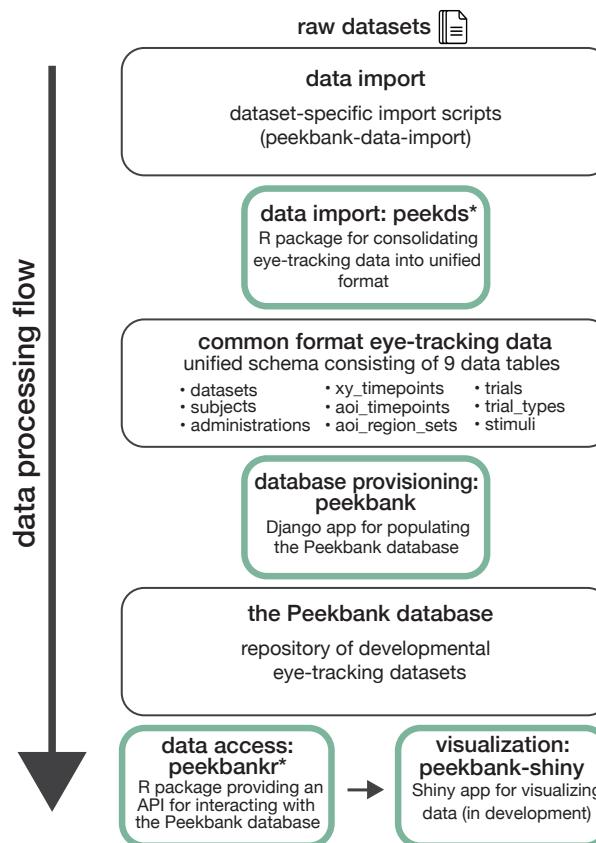
174 **Design and Technical Approach**

175 **Database Framework**

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

181 As illustrated in Figure 1, the Peekbank framework consists of four main components:  
182 (1) a set of tools to *convert* eye-tracking datasets into a unified format, (2) a relational  
183 database populated with data in this unified format, (3) a set of tools to *retrieve* data from  
184 this database, and (4) a web app (using the Shiny framework) for visualizing the data. These  
185 components are supported by three packages. The `peekds` package (for the R language, R  
186 Core Team, 2020) helps researchers convert existing datasets to use the standardized format

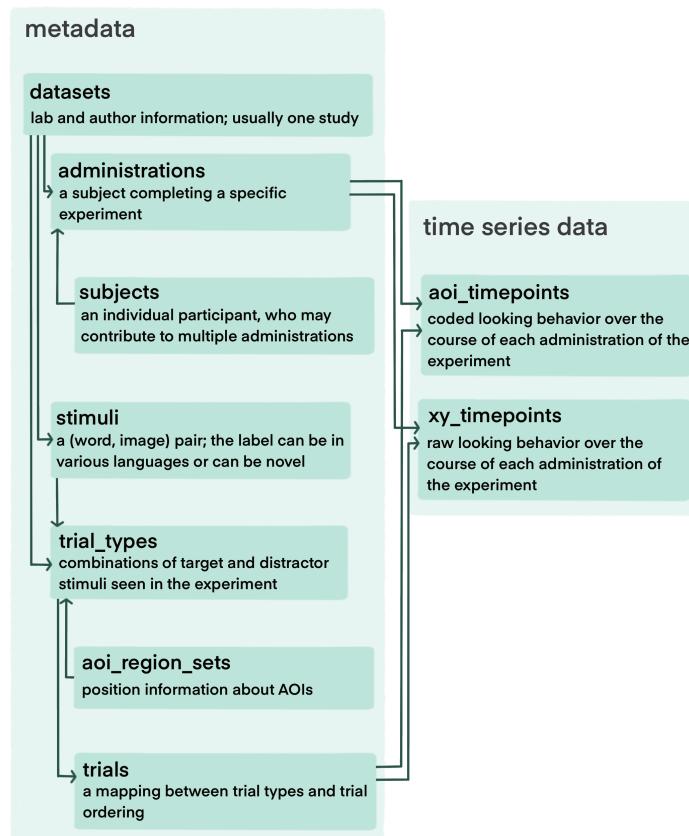
187 of the database. The `peekbank` module (Python) creates a database with the relational  
 188 schema and populates it with the standardized datasets produced by `peekds`. The database  
 189 is served through MySQL, an industry standard relational database server, which may be  
 190 accessed by a variety of programming languages, and can be hosted on one machine and  
 191 accessed by many others over the Internet. As is common in relational databases, records of  
 192 similar types (e.g., participants, trials, experiments, coded looks at each timepoint) are  
 193 grouped into tables, and records of various types are linked through numeric identifiers. The  
 194 `peekbankr` package (R) provides an application programming interface, or API, that offers  
 195 high-level abstractions for accessing the tabular data stored in Peekbank. Most users will  
 196 access data through this final package, in which case the details of data formatting,  
 197 processing, and the specifics of connecting to the database are abstracted away from the user.



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

<sup>198</sup> **Database Schema**

<sup>199</sup> The Peekbank database contains two major types of data: (1) metadata regarding  
<sup>200</sup> experiments, participants, and trials, and (2) time course looking data, detailing where on  
<sup>201</sup> the screen a child is looking at a given point in time (Fig. 2).



*Figure 2.* The Peekbank schema. Each square represents a table in the relational database.

<sup>202</sup> **Metadata.** Metadata can be separated into four parts: (1) participant-level  
<sup>203</sup> information (e.g., demographics) (2) experiment-level information (e.g., the type of eye  
<sup>204</sup> tracker used to collect the data) (3) session information (e.g. a participant's age for a specific  
<sup>205</sup> experimental session) and (4) trial information (e.g., what images or videos were presented  
<sup>206</sup> onscreen, and paired with which audio).

<sup>207</sup> ***Participant Information.***

<sup>208</sup> Invariant information about individuals who participate in one or more studies (e.g., a

209 subject's first language) is recorded in the `subjects` table, while the `administrations`  
210 table contains information about a subject's participation in a single session of a study (see  
211 Session Information, below). This division allows Peekbank to gracefully handle longitudinal  
212 designs: a single subject can be associated with many administrations.

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

216 ***Experiment Information.***

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

220 Information about the experimental design is split across the `trial_types` and  
221 `stimuli` tables. The `trial_types` table encodes information about each trial *in the design*  
222 *of the experiment*,<sup>1</sup> including the target stimulus and location (left vs. right), the distractor  
223 stimulus and location, and the point of disambiguation for that trial. If a dataset used  
224 automatic eye-tracking rather than manual coding, each trial type is additionally linked to a  
225 set of area of interest (x, y) coordinates, encoded in the `aoi_region_sets` table. The  
226 `trial_types` table links trial types to the `aoi_region_sets` table and the `trials` table.  
227 Each trial\_type record links to two records in the `stimuli` table, identified by the  
228 `distractor_id` and the `target_id` fields.

229 Each record in the `stimuli` table is a (word, image) pair. In most experiments, there is  
230 a one-to-one mapping between images and labels (e.g., each time an image of a dog appears  
231 it is referred to as *dog*). For studies in which there are multiple potential labels per image

---

<sup>1</sup> We note that the term *trial* is often overloaded, to refer to a particular combination of stimuli seen by many participants, vs. a participant seeing that particular combination at a particular point in the experiment. We track the latter in the 'trials' table.

232 (e.g., *dog* and *chien* are both used to refer to an image of a dog), images can have multiple  
233 rows in the `stimuli` table with unique labels as well as a row with no label to be used when  
234 the image appears solely as a distractor (and thus its label is ambiguous). This structure is  
235 useful for studies on synonymy or using multiple languages. For studies in which the same  
236 label refers to multiple images (e.g., the word *dog* refers to an image of a dalmatian and a  
237 poodle), the same label can have multiple rows in the `stimuli` table with unique images.

238        ***Session Information.***

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

243        ***Trial Information.***

244        The `trials` table includes information about a specific participant completing a  
245 specific instance of a trial type. This table links each record in the raw data (described  
246 below) to the trial type and specifies the order of the trials seen by a specific participant.

247        **Time course data.** Raw looking data is a series of looks to areas of interest (AOIs),  
248 such as looks to the left or right of the screen, or to (x, y) coordinates on the experiment  
249 screen, linked to points in time. For data generated by eye-trackers, we typically have (x, y)  
250 coordinates at each time point, which we encode in the `xy_timepoints` table. These looks  
251 are also recoded into AOIs according to the AOI coordinates in the `aoi_region_sets` table  
252 using the `add_aois()` function in `peekds`, and encoded in the `aoi_timepoints` table. For  
253 hand-coded data, we typically have a series of AOIs (i.e., looks to the left vs. right of the  
254 screen), but lack information about exact gaze positions on-screen; in these cases the AOIs  
255 are recoded into the categories in the Peekbank schema (target, distractor, other, and  
256 missing) and encoded in the `aoi_timepoints` table; however, these datasets do not have any  
257 corresponding data in the `xy_timepoints` table.

258        Typically, timepoints in the `xy_timepoints` table and `aoi_timepoints` table need to  
259      be regularized to center each trial's time around the point of disambiguation – such that 0 is  
260      the time of target word onset in the trial (i.e., the beginning of *dog* in *Can you find the*  
261      *dog?*). We re-centered timing information to the onset of the target label to facilitate  
262      comparison of target label processing across all datasets.<sup>2</sup> If time values run throughout the  
263      experiment rather than resetting to zero at the beginning of each trial, `rezero_times()` is  
264      used to reset the time at each trial. After this, each trial's times are centered around the  
265      point of disambiguation using `normalize_times()`. When these steps are complete, the  
266      time course is ready for resampling.

267        To facilitate time course analysis and visualization across datasets, time course data  
268      must be resampled to a uniform sampling rate (i.e., such that every trial in every dataset has  
269      observations at the same time points). All data in the database is resampled to 40 Hz  
270      (observations every 25 ms), which represents a compromise between retaining fine-grained  
271      timing information from datasets with dense sampling rates (maximum sampling rate among  
272      current datasets: 500 Hz) while minimizing the possibility of introducing artifacts via  
273      resampling for datasets with lower sampling rates (minimum sampling rate for current  
274      datasets: 30 Hz). Further, 25 ms is a mathematically convenient interval for ensuring  
275      consistent resampling; we found that using 33.333 ms (30 Hz) as our interval simply  
276      introduced a large number of technical complexities. The resampling operation is  
277      accomplished using the `resample_times()` function. During the resampling process, we  
278      interpolate using constant interpolation, selecting for each interpolated timepoint the looking  
279      location for the earlier-observed time point in the original data for both `aoi_timepoints`  
280      and `xy_timepoints` data. Compared to linear interpolation (see e.g. Wass et al., 2014),  
281      constant interpolation has the advantage that it is more conservative, in the sense that it

---

<sup>2</sup> 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, it has been conventional in the literature to use a standardized point of onset.

282 does not introduce new look locations beyond those measured in the original data.  
283 (Investigation of the effects of other interpolation functions would be an interesting  
284 application of the dataset).

285 **Processing, Validation, and Ingestion**

286 The `peekds` package offers functions to extract the above data. Once this data has  
287 been extracted in a tabular form, the package also offers a function to check whether all  
288 tables have the required fields and data types expected by the database. In an effort to  
289 double check the data quality and to make sure that no errors are made in the importing  
290 script, as part of the import procedure we create a time course plot based on our processed  
291 tables to replicate the results in the paper that first presented each dataset. Once this plot  
292 has been created and checked for consistency and all tables pass our validation functions, the  
293 processed dataset is ready for reprocessing into the database using the `peekbank` library.  
294 This library applies additional data checks, and adds the data to the MySQL database using  
295 the Django web framework.

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

301 **Current Data Sources**

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

Table 1  
*Overview of the datasets in the current database.*

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

307 span a wide age spectrum with participants ranging from 9 to 70 months of age, and are  
 308 balanced in terms of gender (47% female). The datasets vary across a number of  
 309 design-related dimensions, and include studies using manually coded video recordings and  
 310 automated eye-tracking methods (e.g., Tobii, EyeLink) to measure gaze behavior. All studies  
 311 tested familiar items, but the database also includes 5 datasets that tested novel  
 312 pseudo-words in addition to familiar words.

### 313 Versioning and Reproducibility

314 The content of Peekbank will change as we add additional datasets and revise previous  
 315 ones. To facilitate reproducibility of analyses, we use a versioning system where successive  
 316 releases are assigned a name reflecting the year and version, e.g., 2021.1. By default, users  
 317 will interact with the most recent version of the database available, though the `peekbankr`  
 318 API allows researchers to run analyses against any previous version of the database. For  
 319 users with intensive use-cases, each version of the database may be downloaded as a  
 320 compressed .sql file and installed on a local MySQL server.

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

331                          **Interfacing with Peekbank**

332                          **Peekbankr**

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

- 336      • `connect_to_peekbank()` opens a connection with the Peekbank database to allow  
 337        tables to be downloaded with the following functions
- 338      • `get_datasets()` gives each dataset name and its citation information
- 339      • `get_subjects()` gives information about persistent subject identifiers (e.g., native  
 340        languages, sex)
- 341      • `get_administrations()` gives information about specific experimental  
 342        administrations (e.g., subject age, monitor size, gaze coding method)
- 343      • `get_stimuli()` gives information about word–image pairings that appeared in  
 344        experiments
- 345      • `get_trial_types()` gives information about pairings of stimuli that appeared in the  
 346        experiment (e.g., point of disambiguation, target and distractor stimuli, condition,

347 language)

- 348 • `get_trials()` gives the trial orderings for each administration, linking trial types to  
the trial IDs used in time course data
- 349
- 350 • `get_aoi_region_sets()` gives coordinate regions for each area of interest (AOI)  
linked to trial type IDs
- 351
- 352 • `get_xy_timepoints()` gives time course data for each subject's looking behavior in  
each trial, as (x, y) coordinates on the experiment monitor
- 353
- 354 • `get_aoi_timepoints()` gives time course data for each subject's looking behavior in  
each trial, coded into areas of interest
- 355

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

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

## 367 Shiny App

368 One goal of the Peekbank project is to allow a wide range of users to easily explore and  
369 learn from the database. We therefore have created an interactive web application –  
370 `peekbank-shiny` – that allows users to quickly and easily create informative visualizations  
371 of individual datasets and aggregated data. `peekbank-shiny` is built using Shiny, a software

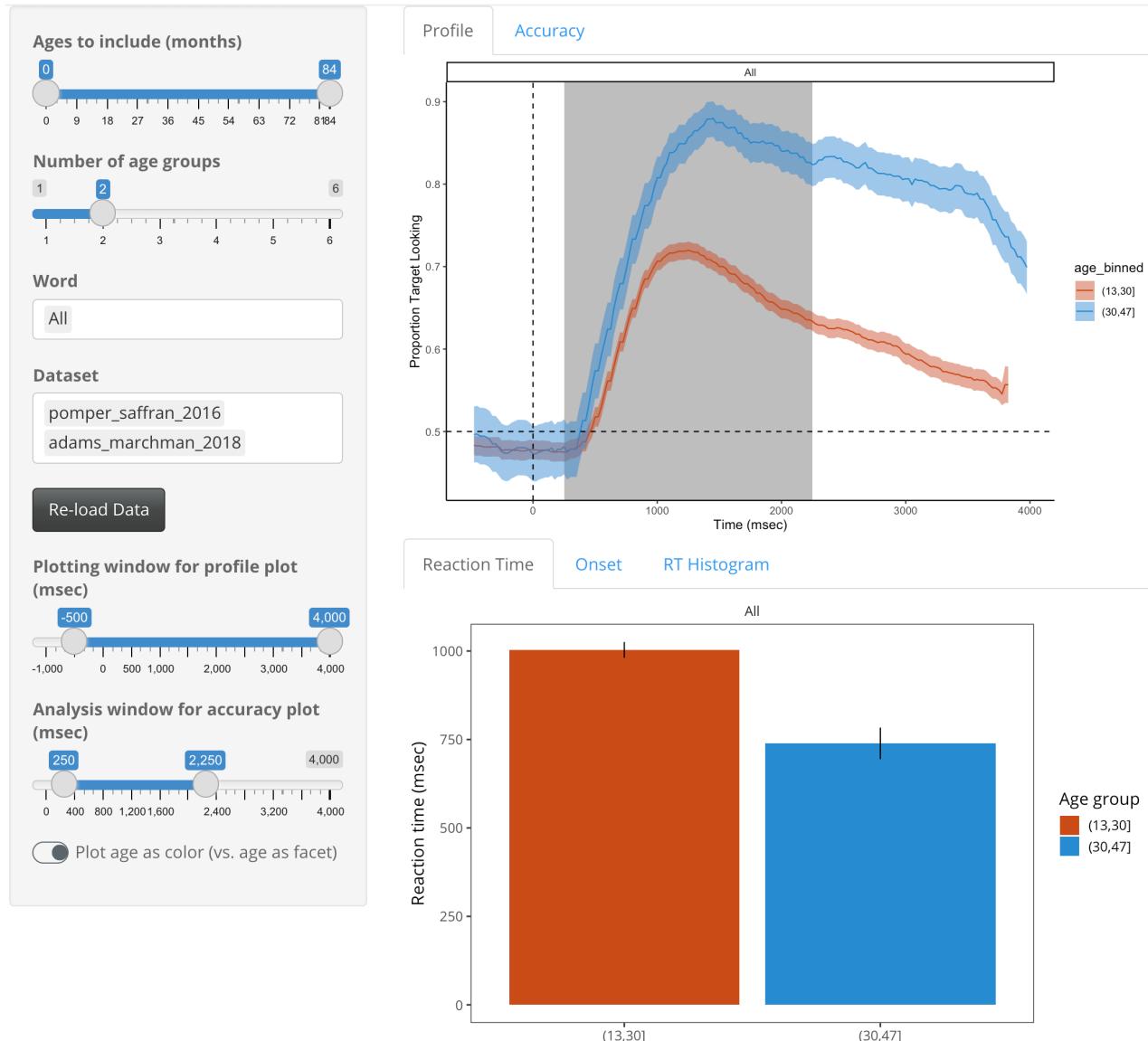
372 package for creating web apps for data exploration with R, as well as the `peekbankr` package.  
373 The Shiny app allows users to create commonly used visualizations of looking-while-listening  
374 data, based on data from the Peekbank database. Specifically, users can visualize:

- 375 1. the time course of looking data in a profile plot depicting infant target looking across  
376 trial time  
377 2. overall accuracy (proportion target looking) within a specified analysis window  
378 3. reaction times (speed of fixating the target image) in response to a target label  
379 4. an onset-contingent plot, which shows the time course of participant looking as a  
380 function of their look location at the onset of the target label

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

389 **OSF site**

390 In addition to the Peekbank database proper, all data is openly available on the  
391 Peekbank OSF webpage (<https://osf.io/pr6wu/>). The OSF site also includes the original raw  
392 data (both time series data and metadata, such as trial lists and participant logs) that was  
393 obtained for each study and subsequently processed into the standardized Peekbank format.  
394 Users who are interested in inspecting or reproducing the processing pipeline for a given  
395 dataset can use the respective import script (openly available on GitHub,  
396 <https://github.com/langcog/peekbank-data-import>) to download and process the raw data



*Figure 3.* Screenshot of the Peekbank visualization tool, 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.

<sup>397</sup> from OSF into its final standardized format. Where available, the OSF page also includes  
<sup>398</sup> additional information about the stimuli used in each dataset, including in some instances  
<sup>399</sup> the original stimulus sets (e.g., image and audio files).

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]

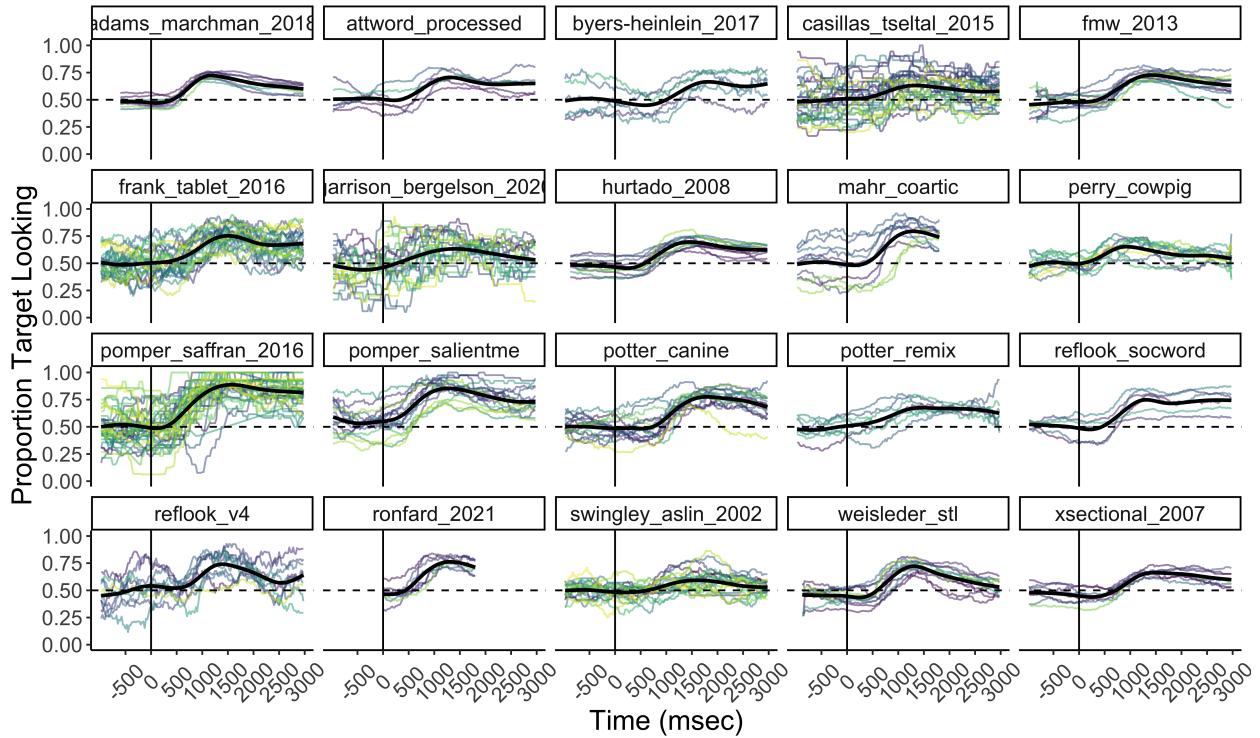
Table 2  
*Average proportion target looking in each dataset.*

400

## Peekbank: General Descriptives

<sup>401</sup> One of the values of the uniform data format we use in Peekbank is the ease of  
<sup>402</sup> providing cross-dataset descriptions that can give an overview of some of the general  
<sup>403</sup> patterns found in our data.

<sup>404</sup> A first broad question is about the degree of accuracy in word recognition found across  
<sup>405</sup> studies. In general, participants demonstrated robust, above-chance word recognition in each  
<sup>406</sup> dataset (chance=0.5). Table 2 shows the average proportion of target looking within a  
<sup>407</sup> standard critical window of 367-2000ms after the onset of the label for each dataset  
<sup>408</sup> (Swingley & Aslin, 2002). Proportion target looking was generally higher for familiar words  
<sup>409</sup> ( $M = 0.66$ , 95% CI = [0.65, 0.67],  $n = 1543$ ) than for novel words learned during the  
<sup>410</sup> experiment ( $M = 0.59$ , 95% CI = [0.58, 0.61],  $n = 822$ ).



*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.

411 A second question of interest is about the variability across items within specific

412 studies. While some studies use many, heterogeneous items, others focus on measuring a

413 much smaller and more homogeneous set. Figure 4 gives an overview of the variability in

414 accuracy for individual words in each dataset. Although all datasets show a gradual rise in

415 accuracy over chance performance, the number of unique target labels and their associated

416 accuracy vary widely across datasets.

417

## Peekbank in Action

418 We provide two potential use-cases for Peekbank data. In each case, we provide sample

419 code so as to model how easy it is to do simple analyses using data from the database. Our

420 first example shows how we can investigate the findings of a classic study. This type of

421 investigation can be a very useful exercise for teaching students about best practices for data

422 analysis (e.g., Hardwicke et al., 2018) and also provides an easy way to explore  
 423 looking-while-listening time course data in a standardized format. Our second example shows  
 424 an in-depth exploration of developmental changes in the recognition of particular words.  
 425 Besides its theoretical interest (which we will explore more fully in subsequent work), this  
 426 type of analysis could in principle be used for optimizing the stimuli for new experiments,  
 427 especially as the Peekbank dataset grows and gains coverage over a greater number of items.

428 **Investigating prior findings: Swingley and Aslin (2002)**

429 Swingley and Aslin (2002) investigated the specificity of 14-16 month-olds' word  
 430 representations using the looking-while-listening paradigm, asking whether recognition would  
 431 be slower and less accurate for mispronunciations, e.g. *oppel* (close mispronunciation) or *opel*  
 432 (distant mispronunciation) instead of *apple* (correct pronunciation). In this short vignette,  
 433 we show how easily the data in Peekbank can be used to visualize this result. Our goal here  
 434 is not to provide a precise analytical reproduction of the analyses reported in the original  
 435 paper, but rather to demonstrate the use of the Peekbank framework to analyze datasets of  
 436 this type. In particular, because Peekbank uses a uniform data import standard, it is likely  
 437 that there will be minor numerical discrepancies between analyses on Peekbank data and  
 438 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")
```

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

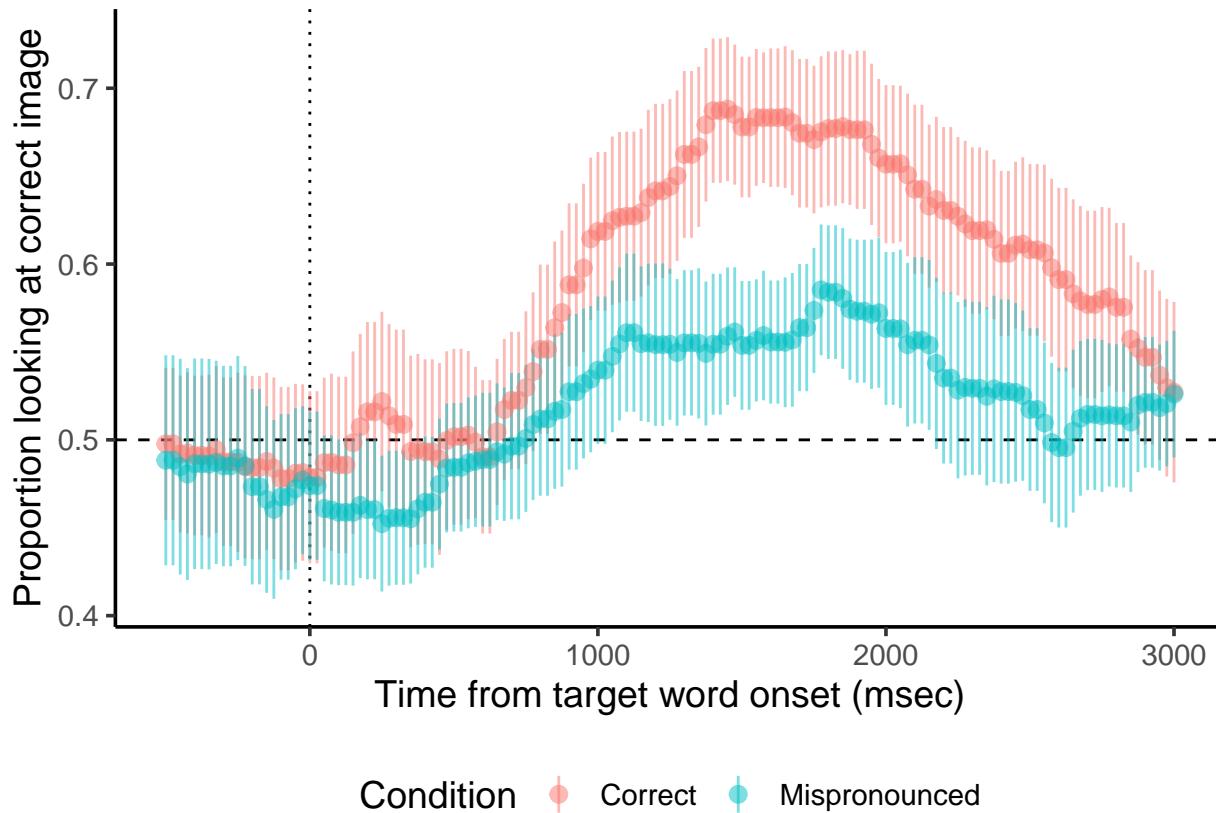
```
swingley_data <- aoi_timepoints |>
  left_join(administrations) |>
  left_join(trials) |>
  left_join(trial_types) |>
  filter(condition != "filler") |>
  mutate(condition = if_else(condition == "cp", "Correct", "Mispronounced"))
```

444 As the code above shows, once the data are joined, condition information for each  
 445 timepoint is present and so we can easily filter out filler trials and set up the conditions for  
 446 further analysis. For simplicity, here we combine both mispronunciation conditions since the  
 447 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()))
```

448 The final step in our analysis is to create a summary dataframe using `dplyr`  
 449 commands. We first group the data by timestep, participant, and condition and compute the  
 450 proportion looking at the correct image. We then summarize again, averaging across  
 451 participants, computing both means and 95% confidence intervals (via the approximation of  
 452 1.96 times the standard error of the mean). The resulting dataframe can be used for  
 453 visualization of the time course of looking.

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



*Figure 5.* Proportion looking at the correct referent by time from the point of disambiguation (the onset of the target noun) in Ssingley & 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.

#### 458 Item analyses

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

464        Our first step is to collect and join the data from the relevant tables including  
 465        timepoint data, trial and stimulus data, and administration data (for participant ages). We  
 466        join these into a single dataframe for easy manipulation; this dataframe is a common  
 467        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)

```

468        Next we select a set of four target words (chosen based on having more than XXX  
 469        children contributing data for each across several one-year age groups). We create age  
 470        groups, aggregate, and compute timepoint-by-timepoint confidence intervals using the  $z$   
 471        approximation.

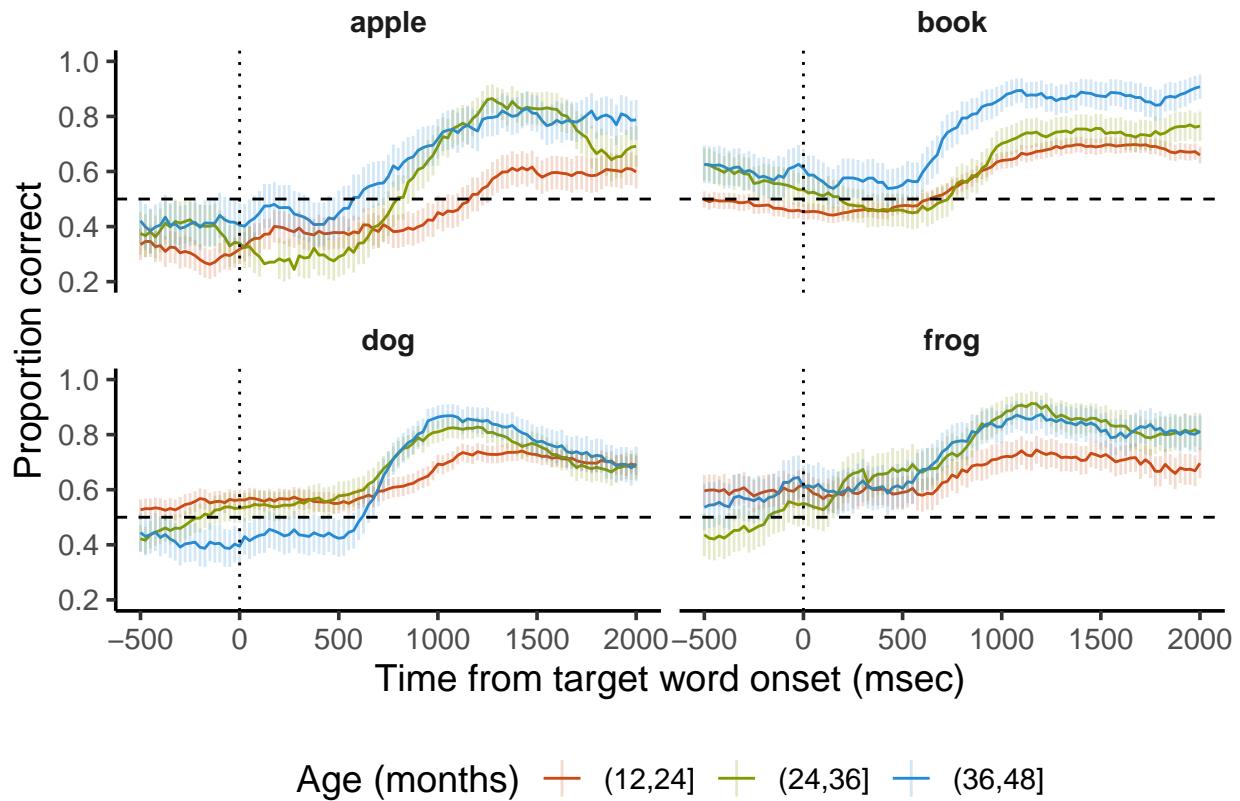
```

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())
```



*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.

Finally, we plot the data as time courses split by age. Our plotting code is shown

below (with styling commands again removed for clarity). Figure 6 shows the resulting plot, with time courses for each of three (rather coarse) age bins. Although some baseline effects are visible across items, we still see clear and consistent increases in looking to the target, with the increase appearing earlier and in many cases asymptoting at a higher level for older children. On the other hand, this simple averaging approach ignores study-to-study variation (perhaps responsible for the baseline effects we see in the *apple* and *frog* items especially). In

479 future work, we hope to introduce model-based analytic methods that use mixed effects  
 480 regression to factor out study-level and individual-level variance in order to recover  
 481 developmental effects more appropriately (see e.g., Zettersten et al., 2021 for a prototype of  
 482 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)
```

483

## Discussion

484 Theoretical progress in understanding child development requires rich datasets, but  
 485 collecting child data is expensive, difficult, and time-intensive. Recent years have seen a  
 486 growing effort to build open source tools and pool research efforts to meet the challenge of  
 487 building a cumulative developmental science (Bergmann et al., 2018; Frank et al., 2017;  
 488 Sanchez et al., 2019; The ManyBabies Consortium, 2020). The Peekbank project expands on  
 489 these efforts by building an infrastructure for aggregating eye-tracking data across studies,  
 490 with a specific focus on the looking-while-listening paradigm. This paper presents an  
 491 overview of the structure of the database, as well as how users can access the database and  
 492 some initial demonstrations of how it can be used both to investigate prior experiments and  
 493 to synthesize data across studies.

494 There are a number of limitations surrounding the current scope of the database. A  
 495 priority in future work will be to expand the size of the database. With 20 datasets currently  
 496 available in the database, idiosyncrasies of particular designs and condition manipulations  
 497 still have substantial influence on modeling results. Expanding the set of distinct datasets

498 will allow us to increase the number of observations per item across datasets, leading to more  
499 robust generalizations across item-level variability. The current database is also limited by  
500 the relatively homogeneous background of its participants, both with respect to language  
501 (almost entirely monolingual native English speakers) and cultural background (Henrich,  
502 Heine, & Norenzayan, 2010; Muthukrishna et al., 2020). Increasing the diversity of  
503 participant backgrounds and languages will expand the scope of the generalizations we can  
504 form about child word recognition.

505 Finally, while the current database is focused on studies of word recognition, the tools  
506 and infrastructure developed in the project can in principle be used to accommodate any  
507 eye-tracking paradigm, opening up new avenues for insights into cognitive development.  
508 Gaze behavior has been at the core of many of the key advances in our understanding of  
509 infant cognition (Bergelson & Swingley, 2012; Fernald et al., 1998; Lew-Williams & Fernald,  
510 2007; Weisleder & Fernald, 2013; Yurovsky & Frank, 2017). Aggregating large datasets of  
511 infant looking behavior in a single, openly-accessible format promises to bring a fuller picture  
512 of infant cognitive development into view.

513

### Acknowledgements

514 We would like to thank the labs and researchers that have made their data publicly  
515 available in the database.

**References**

- Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives*, 14(3), 142–149.
- Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *PNAS*, 109(9), 3253–3258.
- Bergelson, E., & Swingley, D. (2013). The acquisition of abstract words by young infants. *Cognition*, 127(3), 391–397.
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996–2009.
- Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, 37(6), 1461–1476.
- Byers-Heinlein, K., Bergmann, C., & Savalei, V. (2021). Six solutions for more reliable infant research. *PsyArXiv*. <https://doi.org/https://doi.org/10.31234/osf.io/ksfvq>
- DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. *Infancy*, 25(4), 393–419.  
<https://doi.org/10.1111/infa.12337>
- Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science*, 16(2), 234–248. <https://doi.org/10.1111/desc.12019>
- Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231.
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while

- 543 listening: Using eye movements to monitor spoken language comprehension by  
544 infants and young children. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen  
545 (Eds.), *Developmental psycholinguistics: On-line methods in children's language*  
546 *processing* (pp. 97–135). Amsterdam: John Benjamins.
- 547 Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ...  
548 Yurovsky, D. (2017). A Collaborative Approach to Infant Research: Promoting  
549 Reproducibility, Best Practices, and Theory-Building. *Infancy*, 22(4), 421–435.  
550 <https://doi.org/10.1111/infa.12182>
- 551 Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank:  
552 An open repository for developmental vocabulary data. *Journal of Child  
553 Language*, 44(3), 677–694.
- 554 Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability  
555 and Consistency in Early Language Learning: The Wordbank Project*. Cambridge,  
556 MA: MIT Press.
- 557 Gautheron, L., Rochat, N., & Cristia, A. (under review). Managing, storing, and  
558 sharing long-form recordings and their annotations. Retrieved from  
559 <https://doi.org/10.31234/osf.io/w8trm>
- 560 Golinkoff, R. M., Ma, W., Song, L., & Hirsh-Pasek, K. (2013). Twenty-five years  
561 using the intermodal preferential looking paradigm to study language acquisition:  
562 What have we learned? *Perspectives on Psychol. Science*, 8(3), 316–339.
- 563 Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P.,  
564 ... Poldrack, R. A. (2016). The brain imaging data structure, a format for  
565 organizing and describing outputs of neuroimaging experiments. *Scientific Data*,  
566 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- 567 Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C.,  
568 Kidwell, M. C., ... Frank, M. C. (2018). Data availability, reusability, and  
569 analytic reproducibility: Evaluating the impact of a mandatory open data policy

- 570 at the journal *Cognition*. *Royal Society Open Science*, 5(8).
- 571 <https://doi.org/10.1098/rsos.180448>
- 572 Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?
- 573 *The Behavioral and Brain Sciences*, 33(2-3), 61–83.
- 574 <https://doi.org/10.1017/S0140525X0999152X>
- 575 Hirsh-Pasek, K., Cauley, K. M., Golinkoff, R. M., & Gordon, L. (1987). The eyes
- 576 have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child*
- 577 *Language*, 14(1), 23–45.
- 578 Hurtado, N., Marchman, V. A., & Fernald, A. (2007). Spoken word recognition by
- 579 Latino children learning Spanish as their first language. *Journal of Child*
- 580 *Language*, 34(2), 227–249. <https://doi.org/10.1017/S0305000906007896>
- 581 Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake?
- 582 Links between maternal talk, processing speed and vocabulary size in
- 583 Spanish-learning children. *Developmental Science*, 11(6), 31–39.
- 584 <https://doi.org/10.1111/j.1467-7687.2008.00768.x>
- 585 Lewis, M., Braginsky, M., Tsuji, S., Bergmann, C., Piccinini, P. E., Cristia, A., &
- 586 Frank, M. C. (2016). *A Quantitative Synthesis of Early Language Acquisition*
- 587 *Using Meta-Analysis*. <https://doi.org/10.31234/osf.io/htsjm>
- 588 Lew-Williams, C., & Fernald, A. (2007). Young children learning Spanish make rapid
- 589 use of grammatical gender in spoken word recognition. *Psychological Science*,
- 590 18(3), 193–198.
- 591 Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015).
- 592 Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*,
- 593 142, 345–350. <https://doi.org/10.1016/j.cognition.2015.05.009>
- 594 Marchman, V. A., Loi, E. C., Adams, K. A., Ashland, M., Fernald, A., & Feldman, H.
- 595 M. (2018). Speed of language comprehension at 18 months old predicts
- 596 school-relevant outcomes at 54 months old in children born preterm. *Journal of*

- 597                   *Dev. & Behav. Pediatrics*, 39(3), 246–253.
- 598                   Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A.,  
599                   McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich,  
600                   and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of  
601                   Cultural and Psychological Distance. *Psychological Science*, 31(6), 678–701.
- 602                   Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A.,  
603                   ... Vazire, S. (2021). Replicability, Robustness, and Reproducibility in  
604                   Psychological Science. *PsyArXiv*.
- 605                   <https://doi.org/https://doi.org/10.31234/osf.io/ksfvq>
- 606                   Peter, M. S., Durrant, S., Jessop, A., Bidgood, A., Pine, J. M., & Rowland, C. F.  
607                   (2019). Does speed of processing or vocabulary size predict later language growth  
608                   in toddlers? *Cognitive Psychology*, 115, 101238.
- 609                   R Core Team. (2020). *R: A language and environment for statistical computing*.  
610                   Vienna, Austria: R Foundation for Statistical Computing. Retrieved from  
611                   <https://www.R-project.org/>
- 612                   Ronfard, S., Wei, R., & Rowe, M. L. (2021). Exploring the linguistic, cognitive, and  
613                   social skills underlying lexical processing efficiency as measured by the  
614                   looking-while-listening paradigm. *Journal of Child Language*, 1–24.  
615                   <https://doi.org/10.1017/S0305000921000106>
- 616                   Sanchez, A., Meylan, S. C., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank,  
617                   M. C. (2019). childe-db: A flexible and reproducible interface to the child  
618                   language data exchange system. *Behavior Research Methods*, 51(4), 1928–1941.  
619                   <https://doi.org/10.3758/s13428-018-1176-7>
- 620                   Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form  
621                   representations of 14-month-olds. *Psychological Science*, 13(5), 480–484.  
622                   <https://doi.org/10.1111/1467-9280.00485>
- 623                   The ManyBabies Consortium. (2020). Quantifying sources of variability in infancy

- 624 research using the infant-directed speech preference. *Advances in Methods and*  
625 *Practices in Psychological Science*, 3(1), 24–52.
- 626 Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language  
627 Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*,  
628 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- 629 Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: salience and  
630 social cues in early word learning. *Dev Sci*, 20(2).
- 631 Zettersten, M., Bergey, C., Bhatt, N., Boyce, V., Braginsky, M., Carstensen, A., ...  
632 others. (2021). Peekbank: Exploring children's word recognition through an open,  
633 large-scale repository for developmental eye-tracking data.