Defining the Role of Language in Infants' Object Categorization with Eye-Tracking Paradigms

Alexander LaTourrette[1], **Sandra R. Waxman**[1,2]

[1]Department of Psychology, Northwestern University, Evanston, IL, USA

[2]Institute for Policy Research, Northwestern University, Evanston, IL, USA

Abstract

Assessing infant category learning is a challenging but vital aspect of studying infant cognition. By employing a familiarization-test paradigm, we straightforwardly measure infants' success in learning a novel category while relying only on their looking behavior. Moreover, the paradigm can directly measure the impact of different auditory signals on infant categorization across a range of ages. For instance, we assessed how 2-year-olds learn categories in a variety of labeling environments: in our task, 2-year-olds successfully learned categories when all exemplars were labeled or the first two exemplars were labeled, but they failed to categorize when no exemplars were labeled or only the final two exemplars were labeled. To determine infants' success in such tasks, researchers can examine both the overall preference displayed by infants in each condition and infants' pattern of looking over the course of the test phase, using an eye-tracker to provide fine-grained time-course data. Thus, we present a powerful paradigm for identifying the role of language, or any auditory signal, in infants' object category learning.

SUMMARY:

Here we present a protocol for familiarization-test paradigms which provide a direct test of infant categorization and help to define the role of language in early category learning.

Keywords

categorization; learning; infants; eye-tracking; language; labels; novelty preference; familiarization

INTRODUCTION:

Categorization is a fundamental building block of human cognition: infants' categorization abilities emerge early in infancy and become increasingly sophisticated with age.[1–3] Research has also revealed a powerful role for language in infant categorization: from 3 months of age, infants learn categories more successfully when category exemplars are paired with language.[4–6] Moreover, by the end of the first year, infants are attuned to the role

Corresponding Author: Alexander LaTourrette, (latourrette@u.northwestern.edu).

DISCLOSURES:
The authors have nothing to disclose.

of count noun labels in categorization. Pairing category exemplars with a consistent labeling phrase (“This is a vep!”) facilitates infants’ category learning relative to providing either a distinct label for each exemplar (“This is a vep,” “This is a dax,” etc.) or a non-labeling phrase (“Look at this.”).[7–9]

In infants’ everyday experiences, however, the vast majority of objects they encounter will likely remain unlabeled. No caregiver could label every object an infant sees, much less provide the many labels which apply to every object (e.g., “malamute,” “dog,” “pet,” “animal”). This presents a paradox: how can we reconcile the power of labels in infant categorization with their relative scarcity in infants’ daily lives?

To answer this question, we developed a protocol to assess how infants learn categories in a variety of different learning environments, including when they receive a mixture of labeled and unlabeled exemplars. Specifically, we propose that receiving even a few labeled exemplars at the beginning of learning can facilitate categorization—by enhancing infants’ ability to learn from subsequent, unlabeled exemplars as well. This strategy of using a small number of labeled exemplars as a foundation for learning from a larger number of unlabeled exemplars has been widely implemented in the field of machine learning, spawning a family of *semi-supervised learning* (SSL) algorithms[10–12]. Of course, the learning strategies implemented are not identical across different kinds of learners: in machine learning, algorithms typically are exposed to many more exemplars, make explicit guesses about each exemplar, and learn multiple categories simultaneously. Nevertheless, both machine and infant learners may benefit from successfully integrating both labeled and unlabeled exemplars to learn new categories in sparse labeling environments.

Our design focuses on whether 2-year-old children, in the throes of acquiring words for numerous new categories, are capable of this kind of semi-supervised learning. We employ a standard infant categorization measure: a familiarization-test task. In this paradigm, 2-year-olds were exposed to a series of exemplars from a novel category during a familiarization phase. Each exemplar was paired with a different auditory stimulus, depending on the condition (i.e., either a labeling or a non-labeling phrase). Then, at test, all 2-year-olds saw two new objects presented in silence: one object from the now-familiar category and one from a novel category.

If the 2-year-olds successfully form the category during the familiarization phase, then they should distinguish between the two exemplars presented at test. Importantly, because a systematic preference for either the novel or familiar test image reflects an ability to distinguish between them, both familiarity and novelty preferences are interpreted as evidence of successful categorization. Note that on a given task, the nature of this preference is a function of infants’ processing efficiency for the stimulus materials, with familiarity preferences associated with less efficient stimulus processing[4, 13–17]. Presenting the test phase in silence makes it possible to directly assess infants’ success in object categorization and how this success varies according to the information that accompanied the exemplars during familiarization. Thus, this paradigm provides a compelling test of how different types of linguistic environments affect category learning. If labeling enhances category learning in both semi-supervised and fully supervised environments, then 2-year-olds in these conditions should show stronger test preferences than infants in other environments.

PROTOCOL:

All methods described here have been approved by the Northwestern University Institutional Review Board.

1. Stimuli Creation

NOTE: The visual stimuli (see Figure 1) used in the representative design reported below were originally developed in Havy and Waxman (2016)[18] and are available for download at https://osf.io/n6uy8/.

1.1. To create a new continuous category, first design a pair of novel digital images. Next, morph the pair of images together, using software (see, e.g., Table of Materials) to form a continuum of exemplars between the two original images. Create at least two categories in this way so that one can serve as the category to be learned while the other provides the novel category exemplar for the test trial.

1.2. Select the familiarization exemplars at evenly spaced intervals from across each learned category's continuum (e.g., the 0%, 20%, 40%, 60%, 80%, and 100% exemplars). Select an appropriate number of exemplars (e.g., six) commensurate with the difficulty of the category and age of the participants.

1.3. To create the exemplars for the test phase, select the midpoints of the familiar category's continuum and the novel category's continuum (i.e., the 50% exemplar). Then match the color of the novel exemplar to that of the familiar exemplar using an image manipulation program (see, e.g., Table of Materials).

1.4. Record auditory stimuli produced by a female native English speaker in a soundproof booth. If possible, use the same speaker for both labeling phrases (i.e., "Look at the modi") and non-labeling phrases (i.e., "Look at that!").

1.4.1. Instruct the speaker to produce all utterances in infant- or child-directed speech.

1.4.2. Select utterances which are approximately the same length across conditions, likely around 1500 ms per phrase.

2. Apparatus

2.1. Use an appropriate eye-tracker. To collect adequate eye-tracking data for a familiarization-test measure, most widely available eye-trackers will suffice: the objects occupy large portions of the screen, and the data analysis investigates performance over a long window, rather than individual, rapidly occurring eye movements such as saccades.

2.2. Because this task requires eye-tracking infants, ensure that the system conforms to several requirements.

2.2.1. First, use an eye-tracker with a remote tracking mode, which does not require infants to place their heads on a chin-rest. Ensure that the eye-tracker can tolerate relatively large head movements or readjustments.

2.2.2. Second, use a relatively large screen to display the images to infants, (e.g., 57 × 45 cm).

2.2.3. Third, use an extendable arm mount for the eye-tracker to facilitate data collection by allowing the researcher to adjust the height of the eye-tracker to each infant.

2.2.4. Fourth, make the eye-tracking equipment unobtrusive, focusing infants’ attention solely on the display screen. For instance, some systems integrate the eye-tracking equipment with the display monitor or mount the equipment directly below the monitor.

2.3. Note that this task can also be completed by hand-coding high-quality video data of the infants’ looking behavior. While hand-coding techniques may pose some challenges for using the more fine-grained time-course analyses, hand-coded data are entirely sufficient for the aggregate looking analyses.

3. Task Design

3.1. In the eye-tracker’s associated software (see, e.g., Table of Materials), create four different conditions: Fully Supervised, Unsupervised, Semi-supervised, and Reversed Semi-supervised. Ensure these conditions are separate, so that each infant will see only one condition.

3.2. Generate at least two pseudo-random orders of the learning exemplars, with the constraint that no more than two exemplars from the same side of the continuum (0–40% or 60–100%) can be shown consecutively.

3.3. Create familiarization videos that pair the auditory stimuli with the visual stimuli as appropriate for each condition.

3.3.1. Combine the visual and auditory stimuli in video editing software (see, e.g., Table of Materials). Present all images on the same background. Set the onset of the auditory stimulus to an appropriate range, between 500 ms and 1500 ms after the onset of the visual stimulus. Use this short delay to ease infants’ processing load[19].

3.3.2. For instance, in the Fully Supervised condition, pair each familiarization exemplar with a labeling phrase.

3.3.3. In the Unsupervised condition, pair each familiarization exemplar with a non-labeling phrase.

3.3.4. In the Semi-supervised condition, pair only the first two exemplars in each order with labeling phrases but the rest with non-labeling phrases.

3.3.5. For the Reversed Semi-supervised condition, pair the final two exemplars with labeling phrases but the first four with non-labeling phrases (see Figure 1).

3.3.6. Upload these videos into the eye-tracker software, ordering the familiarization videos as determined by the pseudo-randomized order.

3.4. Upload a short (10 s or less) attention-grabbing animation displayed in the center of the screen after familiarization: this will ensure that most infants are looking to the center of the screen when the test phase begins.

3.5. Finally, for each learning category, design two test trials, each featuring two exemplars displayed side-by-side. Ensure that for both test trials, one exemplar will represent the midpoint of the now-familiar category while the other represents the midpoint of the novel category.

3.5.1. Counterbalance the trials so that the left/right positioning of the novel exemplar in the test trial is reversed across videos.

3.5.2. Upload these test trials to the eye-tracker software, positioning them after the post-familiarization attention-getter. Counterbalance these trials' presentation so each infant has an equal chance of seeing a left-novel or right-novel test trial.

3.5.2. Ensure that test trials last at least 5 s, and up to 20 s, in order for children initially looking away to accumulate sufficient looking.

3. Study Procedure

4.1. Before the infant arrives, set up the eye-tracker.

4.1.1. Randomly assign the infant to a condition and an order.

4.1.2. Open the eye-tracker software and select the assigned condition/ order pair.

4.1.3. Now enter the participant number for this recording.

4.2. After performing the consent process, bring the infant and the caregiver to the eye-tracking room. Ensure the room is moderately lit without any distracting decorations on the walls.

4.3. Place a chair in front of the eye-tracker at an appropriate distance for the model of eye-tracker being used. Seat the caregiver in this chair and the infant on the caregiver's lap. If the infant does not wish to sit in the caregiver's lap, they may sit on their own, or they may sit in a car seat.

4.4. If the infant is sitting on the caregiver's lap, instruct the caregiver not to bias infants' behavior in any way but to try to keep the infant centered on the caregiver's lap. Provide caregivers with a pair of blacked-out sunglasses to wear so they cannot see the stimuli.

4.5. Ask the infant to look at the eye-tracker screen; consider displaying an engaging image or video to attract their attention. Position the screen so that infants' eyes are within the calibration window.

4.6. Perform the eye-tracker's calibration procedure. Use a five-point calibration if possible, but less comprehensive calibrations are also likely to be adequate. Infants often respond better when the calibration image is an animation with auditory accompaniment.

4.7. If the infant passes calibration, then begin the experiment. If not, recalibrate until they are successful. Any infants who cannot be calibrated are excluded.

4.8. If multiple experiments are run consecutively, or if a single experiment is quite long, consider re-calibrating after each section.

5. Data Analysis

5.1. Use data analysis software to perform this analysis (e.g., see Table of Materials).

5.2. Create areas of interest (AOIs) around the exemplar positions on the left and right sides of the screen.

5.3. For familiarization trials, use the appropriate AOI to assess the time infants spent looking to the exemplar displayed on each trial. Exclude any infant who does not show sustained looking for a majority of the exemplars (e.g., require that infants attend to 4 of a possible 6 familiarization exemplars for at least 25% of those trials).

5.4. For the test trial, include only infants' first 5 s of accumulated looking. For younger infants, from 3 to 12 months of age, consider using a longer window such as 10 seconds of accumulated looking. Consider excluding infants who show insufficient sustained looking at test (e.g., accumulating less than 2.5 s of looking) or who fail to look to both of the exemplars.

5.5. Now create a preference score for each infant's test trial by dividing the amount of time spent looking to the novel exemplar by the total amount of time looking to both exemplars. To analyze these proportions, transform them first with an empirical logit or arc-sin square-root to make them suitable for analysis with linear models.

5.6. For a time-course analysis of infants' looking behavior at test, separate data into small bins (e.g., between 10 and 100 ms), and calculate a preference score within each bin for each infant.

5.7. Perform an analysis of the time-course data, testing whether infants' pattern of looking throughout the test trial varies by condition. Note that multiple forms of analysis may answer this question, including a cluster-based permutation analysis[20], as demonstrated here, and growth curve modeling.[21]

5.7.1. For a cluster-based permutations analysis, select a t-value threshold, corresponding to the desired alpha level (recommended alphas range from .01 to .20; note that this alpha value does not represent the overall test's alpha level, merely the level required for individual time-bins to exceed the threshold). Sum the t-statistics for every consecutive time-bin that surpasses the chosen t-threshold; these cumulative t-statistics indicate the size of the divergences between conditions in the data.

5.7.2. To determine if these divergences are greater than expected by chance, perform at least 1000 simulations with the condition labels randomly shuffled. Evaluate the unshuffled data's divergences against this chance-based distribution.

Note: It is this comparison of the original divergence against the chance-based distribution that determines the false-positive rate of the analysis, rather than the number of time-bins in which t-tests were conducted or even the t-value threshold selected for those initial t-tests. As a result, this analysis provides a conservative alternative to directly reporting the results from multiple t-tests across pre-specified time-bins (e.g., conducting tests every 500ms).

REPRESENTATIVE RESULTS:

Using the protocol above, we ran two experiments[22]. Analyses were conducted with the *eyetrackingR* package[23], and the data and code are available at https://github.com/sandylat/ssl-in-infancy. In the first experiment, we contrasted a fully supervised condition (n = 24, M_{age} = 26.8 mo), featuring only labeled exemplars, with an unsupervised condition (n = 24, M_{age} = 26.9 mo), featuring only unlabeled exemplars.

Fully Supervised vs. Unsupervised Environments

Infants in the Fully Supervised (M = 13.86 s, SD = 3.00) and Unsupervised (M = 14.94 s, SD = 1.91) conditions showed no difference in their attention to the exemplars during familiarization, t(46) = 1.48, p = .14, d = .43.

At test, 2-year-olds in the Fully Supervised condition (M = .59, SD = .15) displayed a significant preference for the novel category exemplar, t(23) = 3.05, p = .006, d = .62, indicating they had successfully formed the category. In contrast, 2-year-olds in the Unsupervised condition (M = .49, SD = .18) looked roughly equally between the objects at test, t(23) = .39, p = .70, d = .08. Performance differed significantly between these conditions, t(46) = 2.27, p = .028, d = .66 (see Figure 2). Finally, a cluster-based permutation analysis of the time-course of looking patterns at test revealed a significant divergence between the two conditions, p = .038, from 3450 ms to 3850 ms (see Figure 3).

Semi-supervised vs. Reversed Semi-supervised Environments

Next, we examined whether 2-year-olds could learn categories in semi-supervised environments by integrating labeled and unlabeled exemplars. We predicted that receiving labeled exemplars at the beginning of familiarization in a Semi-supervised condition (n = 24, M_{age} = 27.3, 12 female), where the labeled exemplars can provide a foundation for learning from the unlabeled exemplars, would facilitate category learning whereas receiving labeled exemplars at the end of familiarization in a Reversed Semi-supervised condition (n = 24,

M_{age} = 27.2, 13 female) would not. That is, receiving labeled exemplars first should enable 2-year-olds to learn more from the unlabeled exemplars than receiving those labeled exemplars after seeing the unlabeled exemplars.

Infants in the Semi-supervised condition (n = 24, *M* = 13.23 s, *SD* = 3.35) and Reversed Semi-supervised (n = 24, *M* = 12.58 s, *SD* = 2.78) conditions showed similar levels of attention to the exemplars during familiarization, *t*(46) = .73, *p* = .47, *d* = .21.

At test, however, infants in the Semi-supervised condition (*M* = .59, *SD* = .14), displayed a significant novelty preference, *t*(23) = 3.11, *p* = .005, *d* = .63, whereas infants in the Reversed Semi-supervised condition (*M* = .52, *SD* = .13) performed at chance levels, *t*(23) = .76, *p* = .45, *d* = .16. Infants' preferences were marginally different between the two conditions, *t*(46) = 1.80, *p* = .08, *d* = .52 (see Figure 2). Moreover, we also conducted a cluster-based permutation analysis of infants' looking behavior at test, revealing that the Semi-supervised condition showed a stronger novelty preference than the Reversed SSL condition between 3450ms and 3850ms, p = .047 (see Figure 3). This is exactly the same period of time during which the Fully Supervised condition diverged from the Unsupervised condition, suggesting infants were just as successful at learning the category in the Semi-supervised condition as in the Fully Supervised condition.

DISCUSSION:

Here, we present a procedure for evaluating the role of labeling in categorization. By presenting 2-year-olds with a realistic mix of labeled and unlabeled exemplars, we demonstrate that very young children are capable of learning in semi-supervised environments, extending work with adults and older children[24, 25]. Thus, this method offers a resolution to the paradox posed above: if even a few labeled exemplars can spark category learning, then labels can be both rare and powerful.

Critical aspects of this paradigm include the use of novel artificial stimuli and short trials, both of which make the task appropriately challenging and engaging for 2-year-olds. In addition, using an eye-tracker, rather than hand-coding infant looking behavior, provides richer and more precise data on participants' eye gaze; this richness and precision enables the implementation of time-course measures such as the cluster-based permutation analysis.

The central advantages of the familiarization-test paradigm are its straightforward assessment of category learning and its simplicity as a passive looking task. That is, the task directly tests category learning, rather than relying on more complex measures like naming behavior or inductive inferences[3, 26, 27]. Moreover, because familiarization-test tasks can be administered across a broad developmental range (e.g., from 3 months to 3 years), they offer an opportunity to identify developmental continuity and change.

Indeed, the familiarization-test paradigm presented here was designed for 2-year-olds, but similar designs have been widely used with infants in their first year of life[4, 6, 7, 9, 28]. For these younger infants, of course, the task must be simplified: longer exposure to the familiarization exemplars, more exemplars, simpler categories, and a longer window of looking at test may all improve the task's sensitivity for younger infants. More broadly, the familiarization-test paradigm employed here can be easily extended to evaluate the effect of any auditory signal on infant cognition, including silence, sine-wave tones, nonhuman primate vocalizations, and other non-linguistic sounds[5, 13, 29, 30].

Limitations of this task stem primarily from its use of a single outcome variable: infants' preference at test. This makes the task unsuitable for questions about, for instance, how each familiarization exemplar changes infants' category learning or the particular features infants use to learn the category. Time-course analyses, such as the cluster-based permutation analysis, can substantially enrich the insight offered by this paradigm. However, while these analyses enable us to draw stronger conclusions about when two conditions differ in performance, they also raise important questions about what factors drive infants' attentional patterns throughout the test phase, a promising area for future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

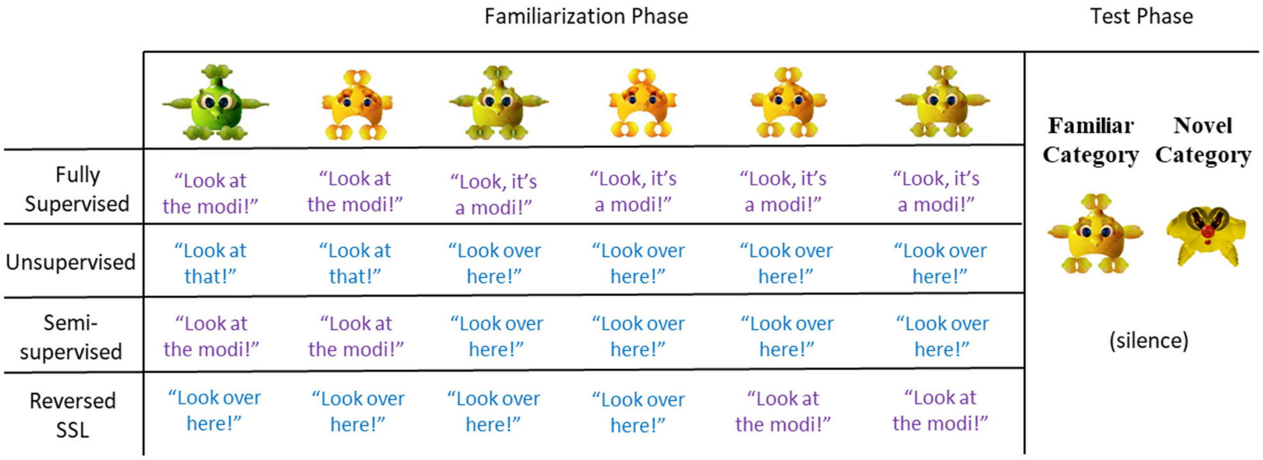
ACKNOWLEDGMENTS:

The research reported here was supported by the National Institute of Child Health and Human Development of the National Institutes of Health under award number R01HD083310 and a National Science Foundation Graduate Research Fellowship under grant no. DGE-1324585. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

REFERENCES:

1. Eimas PD, Quinn PC Studies on the Formation of Perceptually Based Basic-Level Categories in Young Infants. Child Development. 65 (3), 903–917, doi: 10.2307/1131427 (1994). [PubMed: 8045176]
2. Madole KL, Oakes LM Making sense of infant categorization: Stable processes and changing representations. Developmental Review. 19 (2), 263–296 (1999).
3. Gelman SA, Markman EM Categories and induction in young children. Cognition. 23 (3), 183–209, doi: 10.1016/0010-0277(86)90034-X (1986). [PubMed: 3791915]
4. Ferry AL, Hespos SJ, Waxman SR Categorization in 3- and 4-month-old infants: An advantage of words over tones. Child development. 81 (2), 472–479, doi: 10.1111/j.1467-8624.2009.01408.x (2010). [PubMed: 20438453]
5. Fulkerson AL, Waxman SR Words (but not Tones) Facilitate Object Categorization: Evidence From 6- and 12-Month-Olds. Cognition. 105 (1), 218–228, doi: 10.1016/j.cognition.2006.09.005 (2007). [PubMed: 17064677]
6. Balaban MT, Waxman SR Do words facilitate object categorization in 9-month-old infants? Journal of Experimental Child Psychology. 64 (1), 3–26, doi: 10.1006/jecp.1996.2332 (1997). [PubMed: 9126625]
7. Waxman SR, Braun I Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. Cognition. 95 (3), B59–68, doi: 10.1016/j.cognition.2004.09.003 (2005). [PubMed: 15788158]
8. Balaban MT, Waxman SR An examination of the factors underlying the facilitative effect of word phrases on object categorization in 9-month-old infants. Proceedings of the 20th Boston University Conference on Language Development. 1, 483–493 (1996).
9. Waxman SR, Markow DB Words as invitations to form categories: evidence from 12- to 13-month-old infants. Cognitive Psychology. 29 (3), 257–302, doi: 10.1006/cogp.1995.1016 (1995). [PubMed: 8556847]
10. Zhu X Semi-supervised learning literature survey (2005).

11. Chapelle O, Scholkopf B, Zien A Semi-supervised learning: Adaptive computation and machine learning. at <10.7551/mitpress/9780262033589.001.0001>. The MIT Press Cambridge, Mass., USA (2006).
12. Zhu X, Goldberg AB Introduction to semi-supervised learning. Synthesis lectures on artificial intelligence and machine learning. 3 (1), 1–130 (2009).
13. Ferry AL, Hespos SJ, Waxman SR Nonhuman primate vocalizations support categorization in very young human infants. Proceedings of the National Academy of Sciences of the United States of America. 110 (38), 15231–15235, doi: 10.1073/pnas.1221166110 (2013). [PubMed: 24003164]
14. Hunter MA, Ames EW A multifactor model of infant preferences for novel and familiar stimuli. Advances in infancy research (1988).
15. Rose SA, Feldman JF, Jankowski JJ Infant visual recognition memory. Developmental Review. 24 (1), 74–100, doi: 10.1016/j.dr.2003.09.004 (2004).
16. Wetherford MJ, Cohen LB Developmental changes in infant visual preferences for novelty and familiarity. Child Development. 416–424 (1973). [PubMed: 4730528]
17. Perone S, Spencer JP Autonomous visual exploration creates developmental change in familiarity and novelty seeking behaviors. Frontiers in psychology. 4, 648 (2013). [PubMed: 24065948]
18. Havy M, Waxman SR Naming influences 9-month-olds’ identification of discrete categories along a perceptual continuum. Cognition. 156, 41–51, doi: 10.1016/j.cognition.2016.07.011 (2016). [PubMed: 27501225]
19. Althaus N, Plunkett K Timing matters: The impact of label synchrony on infant categorisation. Cognition. 139, 1–9, doi: 10.1016/j.cognition.2015.02.004 (2015). [PubMed: 25781891]
20. Maris E, Oostenveld R Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods. 164 (1), 177–190, doi: 10.1016/j.jneumeth.2007.03.024 (2007). [PubMed: 17517438]
21. Raudenbush SW, Bryk AS Hierarchical Linear Models: Applications and Data Analysis Methods. SAGE. (2002).
22. LaTourrette A, Waxman SR A little labeling goes a long way: Semi-supervised learning in infancy. Developmental Science. e12736, doi: 10.1111/desc.12736.
23. Dink J, Ferguson B eyetrackingR: An R library for eyetracking data analysis (2015).
24. Kalish CW, Zhu X, Rogers TT Drift in children’s categories: When experienced distributions conflict with prior learning. Developmental Science. 18 (6), 940–956, doi: 10.1111/desc.12280 (2015). [PubMed: 25530185]
25. Gibson BR, Rogers TT, Zhu X Human semi-supervised learning. Topics in Cognitive Science. 5 (1), 132–172, doi: 10.1111/tops.12010 (2013). [PubMed: 23335577]
26. Keates J, Graham SA Category Markers or Attributes Why Do Labels Guide Infants’ Inductive Inferences? Psychological Science. 19 (12), 1287–1293, doi: 10.1111/j.1467-9280.2008.02237.x (2008). [PubMed: 19121139]
27. Booth AE, Waxman SR A horse of a different color: Specifying with precision infants’ mappings of novel nouns and adjectives. Child development. 80 (1), 15–22 (2009). [PubMed: 19236389]
28. Perszyk DR, Waxman SR Listening to the calls of the wild: The role of experience in linking language and cognition in young infants. Cognition. 153, 175–181, doi: 10.1016/j.cognition.2016.05.004 (2016). [PubMed: 27209387]
29. Althaus N, Mareschal D Labels direct infants’ attention to commonalities during novel category learning. PLoS ONE. 9 (7), e99670, doi: 10.1371/journal.pone.0099670 (2014). [PubMed: 25014254]
30. Fulkerson AL, Haaf RA The influence of labels, non-labeling sounds, and source of auditory input on 9- and 15-month-olds’ object categorization. Infancy. 4 (3), 349–369, doi: 10.1207/S15327078IN0403_03 (2003).



	Familiarization Phase						Test Phase
Fully Supervised	“Look at the modi!”	“Look at the modi!”	“Look, it’s a modi!”	“Look, it’s a modi!”	“Look, it’s a modi!”	“Look, it’s a modi!”	Familiar Category Novel Category (silence)
Unsupervised	“Look at that!”	“Look at that!”	“Look over here!”	“Look over here!”	“Look over here!”	“Look over here!”	
Semi-supervised	“Look at the modi!”	“Look at the modi!”	“Look over here!”	“Look over here!”	“Look over here!”	“Look over here!”	
Reversed SSL	“Look over here!”	“Look over here!”	“Look over here!”	“Look over here!”	“Look at the modi!”	“Look at the modi!”	

Figure 1. Sample task design.

The familiarization phase consists of 6 trials, each presenting one category member paired with either a labeling or a non-labeling phrase. The test phase simultaneously presents infants with one exemplar from the now-familiar category and one from a novel category. Conditions represent the four conditions presented in the representative results section. This figure has been modified from LaTourrette, A., Waxman, S.R. A little labeling goes a long way: Semi-supervised learning in infancy. *Dev. Sci.* e12736 (2018).

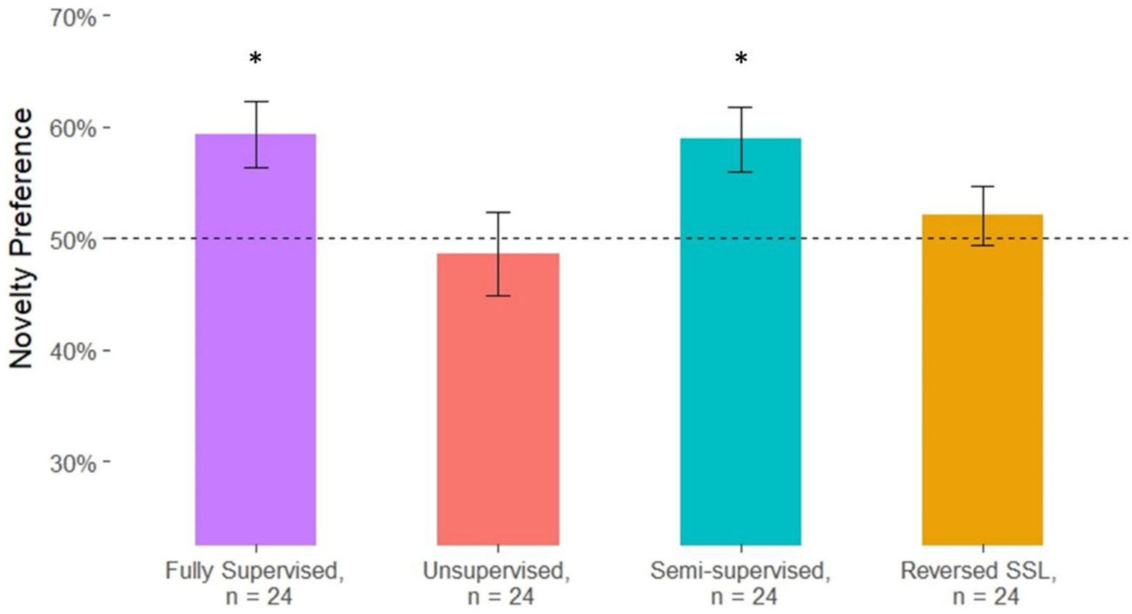


Figure 2. Mean preference scores across conditions.

Infants in the Fully Supervised and Semi-supervised conditions displayed novelty preferences significantly above chance, $p < .05$. Infants in the Unsupervised and Reversed SSL conditions performed at chance levels. Error bars represent standard errors of the mean. This figure has been modified from LaTourrette, A., Waxman, S.R. A little labeling goes a long way: Semi-supervised learning in infancy. *Dev. Sci.* e12736 (2018).

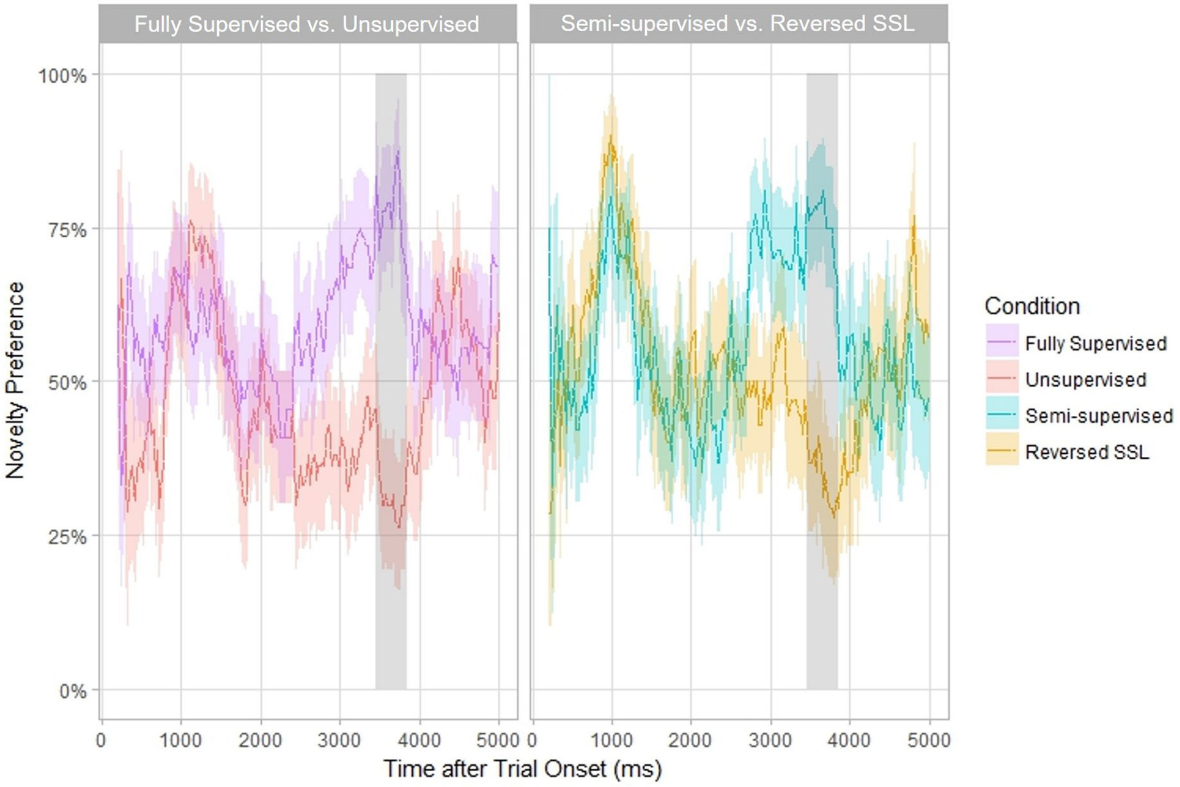


Figure 3. Infant's looking patterns during test.

In the Fully Supervised and Unsupervised conditions (at left) and in the Semi-supervised and Reversed Semi-supervised conditions (at right), infants' pattern of looking to the exemplars diverged between 3450ms and 3850ms. The grey shaded bar in each graph denotes this divergent period. The colored shaded regions around each condition indicate standard error of the mean. This figure has been modified from LaTourrette, A., Waxman, S.R. A little labeling goes a long way: Semi-supervised learning in infancy. *Dev. Sci*. e12736 (2018).