Received: 21 February 2018 | Accepted: 21 July 2018

DOI: 10.1111/desc.12736

SHORT REPORT

WILEY Developmental Science

A little labeling goes a long way: Semi-supervised learning in infancy

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

[1]Department of Psychology, Northwestern University, Evanston, Illinois

[2]Institute for Policy Research, Northwestern University, Evanston, Illinois

Correspondence
Alexander LaTourrette, Department of Psychology, Northwestern University, Evanston, Illinois.
Email: latourrette@u.northwestern.edu

Funding information
National Institute of Child Health and Human Development of the National Institutes of Health, Grant/Award Number: R01HD083310; National Science Foundation Graduate Research Fellowship, Grant/Award Number: DGE-1324585

Abstract

There is considerable evidence that labeling supports infants' object categorization. Yet in daily life, most of the category exemplars that infants encounter will remain unlabeled. Inspired by recent evidence from machine learning, we propose that infants successfully exploit this sparsely labeled input through "semi-supervised learning." Providing only a few labeled exemplars leads infants to initiate the process of categorization, after which they can integrate all subsequent exemplars, labeled or unlabeled, into their evolving category representations. Using a classic novelty preference task, we introduced 2-year-old infants (n = 96) to a novel object category, varying whether and when its exemplars were labeled. Infants were equally successful whether all exemplars were labeled (fully supervised condition) or only the first two exemplars were labeled (semi-supervised condition), but they failed when no exemplars were labeled (unsupervised condition). Furthermore, the timing of the labeling mattered: when the labeled exemplars were provided at the end, rather than the beginning, of familiarization (reversed semi-supervised condition), infants failed to learn the category. This provides the first evidence of semi-supervised learning in infancy, revealing that infants excel at learning from exactly the kind of input that they typically receive in acquiring real-world categories and their names.

KEYWORDS

category learning, conceptual development, language acquisition, language and thought, semi-supervised learning

1 | INTRODUCTION

The powerful connection between language and human cognition begins in infancy. Even before infants begin to produce their first words, language supports fundamental cognitive capacities (Feigenson & Halberda, 2008; Ferguson & Lew-Williams, 2016; Ferry, Hespos, & Waxman, 2010; Xu, 2002; for a review, see Perszyk & Waxman, 2018). Infant object categorization serves as a strong illustration of this link: decades of research reveal that naming facilitates infants' object categorization. When infants are introduced to a series of distinct objects belonging to the same object category, they successfully identify that category if the objects are each named with the same novel word (e.g., "Look at the *dax*!"), but they fail to do so if these objects remain unnamed (Balaban & Waxman, 1996; Waxman & Braun, 2005; Waxman & Markow, 1995). Naming also shapes the boundaries of the object categories infants form. When presented with a set of objects that vary along a continuous distribution, infants use labels to infer the underlying categories. If all the objects are named with the same novel word, infants form a single inclusive object category; in contrast, if the objects on opposite sides of the continuum receive different names, infants form two contrastive categories (Althaus & Westermann, 2016; Havy & Waxman, 2016; Plunkett, Hu, & Cohen, 2008). This link between naming and object categories, which develops within infants' first year, may play an important role in early category acquisition.

In infants' everyday experiences, however, their input differs dramatically from the input provided in carefully controlled experiments in the infant laboratory. Infants do not typically hear a name for every exemplar they encounter. Instead, the vast majority of exemplars remain unlabeled. Even a caregiver who readily labels novel objects as they appear (e.g., "Look, a grasshopper!") will not label all the objects within the infants' view (e.g., the leaf on which the grasshopper has alighted, the surrounding trees, the bird passing overhead). Nor will the caregiver label each grasshopper the infant later encounters. Complicating the task further, every object is a member of multiple categories that overlap in scope (e.g., grasshopper, bug, insect, animal; green; jumping), yet only a few such categories, if any, will be named in a given encounter. In addition, there is considerable variation, both within and across cultures, in how and how often adults label objects for infants (e.g., Cartmill et al., 2013; Gaskins, 1999; Lieven, 1994; Rogoff, Mistry, Göncü, & Mosier, 1993; Shneidman & Goldin-Meadow, 2012). Finally, even when a caregiver does name an object within the infant's view, identifying the intended referent is often still quite difficult (Cartmill et al., 2013; Gillette, Gleitman, Gleitman, & Lederer, 1999).

In sum, there can be little doubt that the clear and consistent naming episodes that have proven so advantageous in the infant laboratory do not characterize infants' experiences in the natural world. For many categories, of course, this poses no problem: infants can successfully learn certain categories exclusively from unlabeled exemplars (e.g., Quinn, Eimas, & Rosenkrantz, 1993). This sort of unsupervised learning may also support infants' ability to learn words referring to objects with which they have prior experience (Clerkin, Hart, Rehg, Yu, & Smith, 2017; Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). Particularly for more difficult categories, however, unambiguous labeling can play a powerful role in infant word learning (Tomasello & Farrar, 1986; Waxman, 1990, 1998; Woodard, Gleitman, & Trueswell, 2016). While rare, these moments in which infants successfully link labels and referents can dramatically boost word learning when they occur (Stevens, Gleitman, Trueswell, & Yang, 2016; Trueswell, Medina, Hafri, & Gleitman, 2013).

How, then, can we reconcile the power of labels in infants' categorization with their relative scarcity in infants' input? Can infants use the labels they *do* hear to learn new object categories? One possibility is that providing even a small set of labeled exemplars is sufficient to spark the acquisition of object categories: labeled exemplars may serve as a foundation for learning from subsequent, *unlabeled* exemplars. This strategy, known as "semi-supervised learning" (SSL), has been documented extensively in machine learning (for reviews, see Chapelle, Scholkopf, & Zien, 2006; Zhu, 2005; Zhu & Goldberg, 2009). Typically, SSL algorithms employ a two-step process. First, a small set of labeled exemplars is provided: this allows the algorithm to form initial estimates of the categories to be learned. Next, the algorithm is given a much larger set of unlabeled exemplars: this permits it to adjust the category boundaries to reflect the full distribution of exemplars. While there are many different mechanisms by which an algorithm might make this adjustment, they all share the goal of gradually refining categories to account for the unlabeled data (Zhu, 2005). For instance, one widely-used approach is expectation-maximization (Dempster, Laird, & Rubin, 1977). Essentially, the algorithm predicts each unlabeled exemplar's category by comparing it to the category estimates derived from the labeled exemplars. Next, the algorithm incorporates these unlabeled exemplars into their predicted categories, weighting each exemplar's impact on its category by the algorithm's confidence in that prediction. This process then repeats, using the new category representations, until the algorithm reaches an optimal set of category boundaries, informed by both labeled and unlabeled exemplars. Notably, this process typically relies on the algorithm processing the labeled exemplars first.

RESEARCH HIGHLIGHTS

- Two-year-old infants successfully learn novel object categories even when only the first few members are named.
- Labeled exemplars serve as a foundation for learning from subsequent, unlabeled exemplars
- Infants—like children, adults, and machines—are adept at learning in semi-supervised environments

Semi-supervised learning has proven successful in machine learning across myriad content domains including text classification (Goldberg & Zhu, 2006; Nigam, Mccallum, Thrun, & Mitchell, 2000; Xu, Fumera, Roli, & Zhou, 2009) and object/person identification (Balcan et al., 2005; Guillaumin, Verbeek, & Schmid, 2010). For example, Lu, Ting, Little, and Murphy (2013) compared fully supervised and semi-supervised algorithms tasked with identifying individual players in a video recording of the 2010 NBA Championship series. Although the fully supervised and semi-supervised algorithms ultimately achieved equivalent accuracy in the task, the number of labeling episodes required differed dramatically: the fully supervised algorithm required 10 times as many labeled exemplars (20,000) as the semi-supervised algorithm (2,000). Clearly, then, SSL offers a powerful solution to the challenge of learning when the available unlabeled exemplars far outnumber the available labeled exemplars.

Moreover, the benefits of SSL are evident not only in machines but also in adults (for a review, see Gibson, Rogers, & Zhu, 2013). Adults successfully learn novel categories by integrating a small set of labeled exemplars with a subsequent, larger set of unlabeled exemplars (Gibson et al., 2013; Kalish, Rogers, Lang, & Zhu, 2011; Zhu, Rogers, Qian, & Kalish, 2007). Indeed, Lake and McClelland (2011) estimated that when acquiring new categories, their adult participants weighted unlabeled exemplars at least 40% as heavily as labeled exemplars. Although the conditions under which adults most successfully use SSL are still under investigation (cf. McDonnell, Jew, & Gureckis, 2012; Rogers, Gibson, Harrison, & Zhu, 2010; Vandist, De Schryver, & Rosseel, 2009), adults appear to readily engage in SSL, drawing on both labeled and unlabeled exemplars to learn new categories.

Recall that, like these adults, most infants typically receive a few high-quality labeled exemplars amidst a larger set of unlabeled

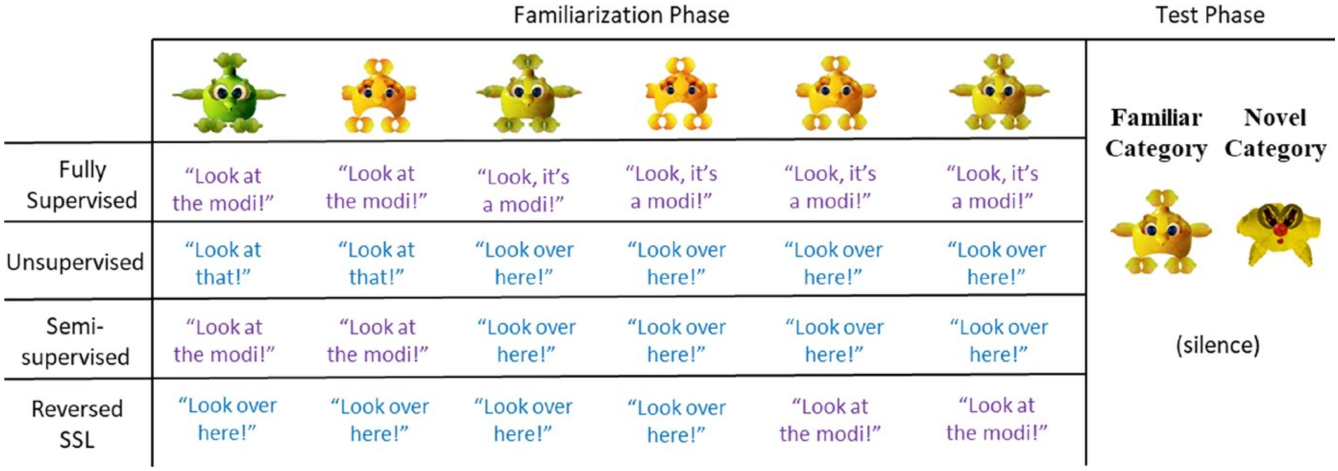

FIGURE 1 Design for Experiments 1–3. During the familiarization phase, all children viewed six category members, each paired with either a labeling or non-labeling phrase. At test, children simultaneously viewed one exemplar from the now-familiar category and one from a novel category

exemplars when learning object categories. In principle, then, SSL appears to be a natural fit for early category learning (cf., Kalish et al., 2011; McDonnell et al., 2012; Zhu, 2005). Unfortunately, however, developmental evidence is scant. We are aware of only one investigation of SSL in children: Kalish, Zhu, and Rogers (2015) documented successful SSL in 7- to 8-year-olds, with suggestive but more equivocal performance in 4- to 6-year-olds.

At issue, then, is whether infants take advantage of SSL in their first few years of life, as they acquire many new object categories and their names. There are several promising hints to suggest that they can. First, infants successfully learn object categories from a mixture of labeled and unlabeled exemplars; although it is unclear from this prior work whether infants integrated the unlabeled exemplars or simply ignored them (Balaban & Waxman, 1997; Waxman & Markow, 1995). Second, substantial evidence suggests the effect of naming extends beyond the exemplars that have been named: naming directs infants' attention to commonalities among objects and, moreover, influences their interpretation of as-yet-unnamed objects (Althaus & Mareschal, 2014; Althaus & Plunkett, 2015; Waxman & Braun, 2005; Waxman & Markow, 1995). Perhaps, then, exposure to labeled exemplars provides infants with a robust foundation for learning from subsequent, unlabeled exemplars. That is, perhaps infants, like SSL algorithms, use the labeled exemplars to classify subsequent unlabeled exemplars and incorporate them into an evolving category representation.

Here, we test this hypothesis directly. We focus on 2-year-olds, an age at which infants rapidly acquire object categories and their names. We adapt a classic object categorization task (Experiment 1) to explore SSL in infants (Experiments 2 and 3). We examine infants' object categorization in three distinct conditions: fully supervised learning (FSL: all exemplars labeled), unsupervised learning (USL: no exemplars labeled), and SSL (only the first few exemplars labeled). If infants benefit from SSL, then performance in the SSL and FSL conditions should be comparable. Alternatively, if SSL provides insufficient support for infant categorization, then performance in the SSL and USL conditions should be comparable.

2 | EXPERIMENT 1

Building upon previous work (e.g., Waxman & Markow, 1995), we engaged infants in a two-step categorization task. First, during familiarization, we introduced infants to six exemplars from one object category. Then, at test, we presented two new exemplars: one from the now-familiar category and one from a novel category. All infants viewed the same objects; what varied across conditions was whether the familiarization exemplars were labeled. For half the infants, all the exemplars were labeled (FSL); for the remaining infants, all exemplars were unlabeled (USL). We expected that infants in the FSL, but not the USL, condition would successfully form object categories.

2.1 | Method

2.1.1 | Participants

Forty-eight infants (23 female) between 25 and 30 months of age (M = 26.8, SD = 1.25) from predominantly college-educated, white families living in the Greater Chicago area participated. Five additional infants were excluded prior to analysis for technical issues (2), parental interference (1), or failing to accumulate at least 2,500 ms of looking during test (2). All infants attended to both exemplars during the test phase.

2.1.2 | Apparatus

A Tobii T60XL eyetracker was used for stimulus presentation and data collection. The eyetracker has a sampling rate of 60 Hz, and a display size of 57.3 × 45 cm.

2.1.3 | Materials

Auditory stimuli

Two naming phrases ("Look at the modi!"/"Look—it's a modi!") and two non-naming phrases ("Look at that!"/"Look over here!") were produced by a female using infant-directed speech and recorded in a sound isolation booth.

Visual stimuli

Two sets of novel objects, designed by Havy and Waxman (2016), served as visual stimuli (see Figure 1). First, two pairs of colorful, creature-like objects were created. The two objects in each pair were then morphed together, yielding two perceptual continua of objects. As a result, images varied along a variety of dimensions, including color, overall body shape, and feature details. By sampling from these continua at 20% intervals, we obtained 6 regularly distributed, continuously varied category exemplars for each of the two categories. These served as familiarization exemplars. At test, infants saw two new exemplars: a new member of the familiar category and a member of a novel category. For the familiar category exemplar, we selected the midpoint of the familiar continuum. For the novel category exemplar, we generated another two continuous categories and selected their midpoints. Each of these novel category exemplars was then paired with a familiar category test exemplar, and the novel exemplar's coloring was altered to match the familiar exemplar's.

Procedure

All infants saw the same images. During familiarization, infants viewed six different exemplars from one of the two continuous categories, counterbalanced across infants. Exemplars were presented in one of four pseudo-random orders; to prevent infants from forming spurious generalizations, the first two exemplars were always drawn from different sides of the continuum (cf. Gerken & Quam, 2017). Exemplars appeared on either the left or right side of the screen, with the initial side counterbalanced across infants, and were approximately 600 × 750 pixels. Each exemplar was presented once for 3 s.

Infants were randomly assigned to either the FSL or USL condition. In the FSL condition, each familiarization exemplar was paired with a labeling phrase, containing a novel noun (e.g., "Look at the modi!"); in the USL condition, each exemplar was paired with a non-labeling phrase (e.g., "Look at that!") (see Figure 1).

After familiarization, an attention-getter appeared at the center of the screen (10 s), followed immediately by the test trial (20 s). At test, all infants saw two exemplars: a new exemplar from the familiar category and an exemplar from a novel category. These were presented side-by-side and in silence; their side placement was counterbalanced. We analyzed each infant's first 5 seconds of looking to the test objects.[1]

Data preparation

We calculated each infant's novelty preference score (looking to the novel exemplar divided by looking to novel and familiar exemplars). Because this calculation yields a bounded proportion, we used an empirical logit transformation to ensure the data are suitable for analysis with linear models.

In addition to novelty preferences, we examined how infants' attention evolved over the course of the test phase. We predicted that their looking patterns would diverge over time, with infants in the FSL condition showing greater attention to the novel object than those in the USL condition. We employed a cluster-based permutations analysis (see Maris & Oostenveld, 2007) to identify when, if ever, performance in the two conditions diverged significantly and to do so without inflating the overall Type I error rate (for other examples, see Dautriche, Swingley, & Christophe, 2015; de Carvalho, Dautriche, & Christophe, 2016; Hahn, Snedeker, & Rabagliati, 2015). The analysis was implemented with the EYETRACKINGR package (Dink & Ferguson, 2015). To begin, we created 25 ms bins and compared performance across conditions within each bin using a *t*-test. For adjacent time-bins which yielded a significant result (based on alpha = 0.05), the *t*-statistics for those bins were summed together, creating a cumulative t-statistic representing the overall size of that divergence. The use of alpha = 0.05 as the threshold here represents a conservative choice, ensuring that any reported divergences will be large in scale. Finally, to evaluate the probability of observing these divergences by chance, we performed 1,000 simulations in which the condition labels were randomly shuffled. By evaluating the divergences in infant data against this chance-based distribution, we obtain a p-value, estimating the likelihood that each divergence might have occurred by chance.[2]

2.2 | Results

2.2.1 | Familiarization

Infants in both conditions were highly attentive to the visual stimuli: there were no differences in looking time between the USL (M = 14.94 s, SD = 1.91) and FSL (M = 13.86, SD = 3.00) conditions, $t(46) = 1.48$, $p = 0.14$, $d = 0.43$.

2.2.2 | Test

Preliminary analyses yielded no effect of age, sex, exemplar order, exemplar left/right position, whether the novel test exemplar occurred on the same side as the final familiarization exemplar, or the category learned, $ps > 0.10$ on infants' novelty preferences in any experiment. Therefore, here and in all subsequent analyses, we collapse across these factors.

As predicted, infants in the FSL condition (M = 0.59, SD = 0.15) successfully learned the object category, revealing a significant novelty preference, $t(23) = 3.05$, $p = 0.006$, $d = 0.62$, but those in the USL condition (M = 0.49, SD = 0.18) performed at chance levels, $t(23) = 0.39$, $p = 0.70$, $d = 0.08$ (see Figure 2). Moreover, performance in the two conditions was significantly different, $t(46) = 2.27$, $p = 0.028$, $d = 0.66$.

Infants' looking patterns over the course of the test phase also varied significantly as a function of condition (see Figure 3).

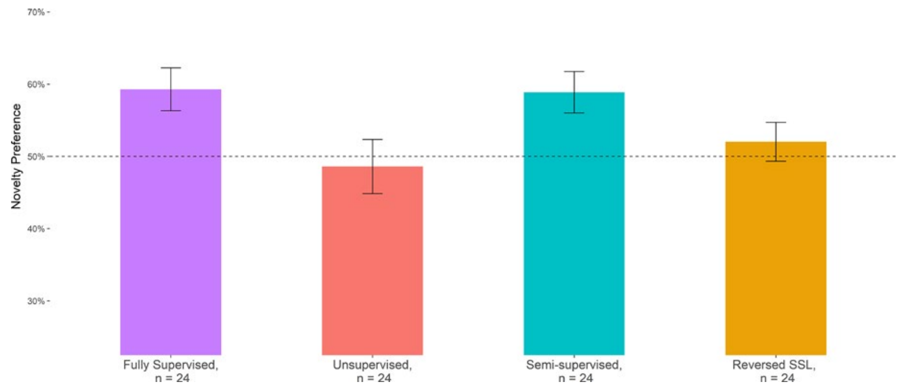

FIGURE 2 Mean novelty preferences by condition. Infants in the Fully Supervised (FSL) and Semi-supervised (SSL) conditions revealed reliable novelty preferences, $p < 0.05$. Infants in the Unsupervised (USL) or Reversed SSL conditions performed at chance levels. Error bars represent standard errors of the mean

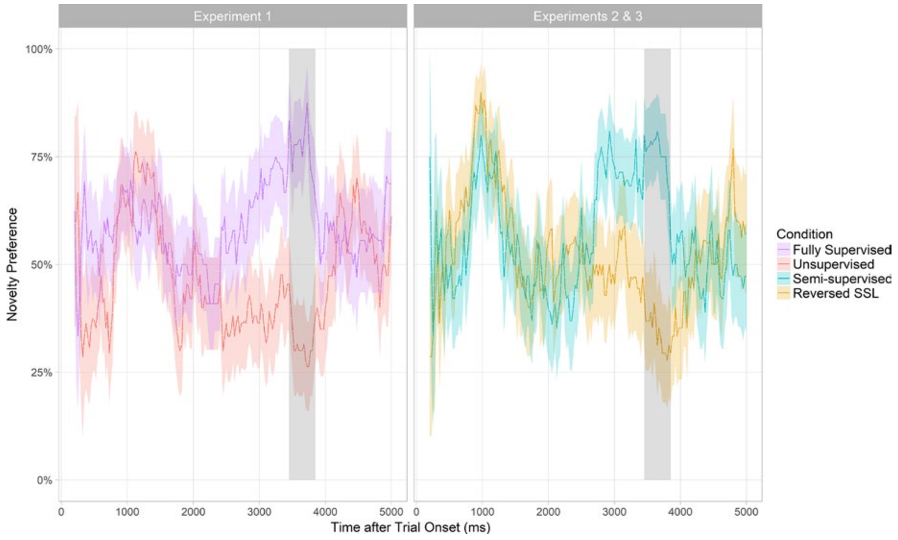

FIGURE 3 Looking patterns over time at test. In the Fully Supervised (FSL) and Unsupervised (USL) conditions (at left) and in the Semi-supervised (SSL) and Reversed SSL conditions (at right), infants' visual attention diverged between 3450ms and 3850ms. This divergent period is indicated in the grey shaded bar. The colored shaded regions around each condition indicate standard error of the mean

A cluster-based permutations analysis revealed that performance in the FSL and USL conditions diverged significantly, from 3,450 ms to 3,850 ms after test onset, $p = 0.038$, with stronger novelty preferences in the FSL condition throughout this window.

2.3 | Discussion

Infants in the FSL, but not USL, condition successfully acquired the novel category. This extends evidence from previous work (Balaban & Waxman, 1996; Waxman & Braun, 2005; Waxman & Markow, 1995) to a new age group and to entirely novel, continuous categories. This outcome, evident in both the novelty preference and time-course measures, provides a firm foundation for assessing infants' success in a semi-supervised condition.

3 | EXPERIMENT 2

Our next goal was to test infant categorization in a SSL condition. In the SSL condition, familiarization began with two labeled exemplars, followed by four unlabeled exemplars. If infants are capable of SSL, then they should successfully categorize in this condition, mirroring infants in the FSL condition (Experiment 1). But if a semi-supervised environment provides insufficient support, then infants in the SSL condition should mirror the chance-like performance of infants in the USL condition (Experiment 1).

3.1 | Methods

3.1.1 | Participants

Twenty-four infants (12 female) between 25 and 30 months of age ($M = 27.3$, $SD = 1.15$) served as participants. Two additional infants were excluded prior to analysis for technical issues, and one infant was excluded as an outlier.[3] All infants attended to both exemplars during the test phase.

3.1.2 | Procedure

The procedure was identical to Experiment 1, with one exception: only the first two familiarization exemplars were labeled (e.g., "Look at the modi!"), as in FSL (Experiment 1). The last four exemplars were accompanied by a non-labeling phrase (e.g., "Look at that!"), exactly as in USL (Experiment 1).

3.2 | Results

3.2.1 | Familiarization

Infants were attentive during familiarization (*M* = 13.23, *SD* = 3.35); their total looking time did not significantly differ from infants in Experiment 1, $t(70) = 1.64$, $p = 0.10$, $d = 0.41$.

3.2.2 | Test

Infants in the SSL condition showed a significant novelty preference (*M* = 0.59, *SD* = 0.14), $t(23) = 3.11$, $p = 0.005$, $d = 0.63$ (see Figure 2), mirroring the success of infants in the FSL condition (*M* = 0.59) in Experiment 1, $t(46) = 0.21$, $p = 0.84$, $d = 0.06$, and differing significantly from infants' in the USL condition (*M* = 0.49), $t(46) = 2.19$, $p = 0.034$, $d = 0.63$.

The cluster-based permutations analysis reveals a striking equivalence in infants' looking patterns in the SSL and FSL conditions: in every time bin, performance in these conditions was statistically indistinguishable. In contrast, performance in the SSL condition diverged significantly from that of the USL condition, doing so between 3450ms and 3825ms, $p = 0.042$.

3.3 | Discussion

Infants' success in the SSL condition is striking: they learned object categories just as successfully when only the first two exemplars were labeled as when all exemplars were labeled (FSL condition, Experiment 1). This is consistent with the proposal that infants integrated the first two named exemplars with subsequent unlabeled exemplars. However, it is also possible that infants did not rely on the unlabeled exemplars at all, and that their success derived from the two labeled exemplars alone. We test this possibility in Experiment 3.

4 | EXPERIMENT 3

If infants' success in the SSL condition was predicated only on the two labeled exemplars, then they should succeed whether the labeled exemplars occur at the beginning or end of the learning phase. To test this, we introduced a "Reversed SSL" condition. As in Experiment 2, only two exemplars were labeled. However, we presented the named exemplars at the end, rather than the beginning, of the familiarization phase. If providing two labeled exemplars is sufficient to support categorization, then infants should succeed here, just as they did in Experiment 2. Because these labeled exemplars occur last, however, there is no opportunity for infants to integrate subsequent unlabeled exemplars—a hallmark of SSL. Thus, if infants relied on SSL to succeed in Experiment 2, they should fail here.

4.1 | Methods

4.1.1 | Participants

Twenty-four infants (13 female) between 25 and 30 months of age (*M* = 27.2, *SD* = 1.27) served as participants. Four additional infants were excluded prior to analysis for technical issues. Another infant was excluded from analysis and replaced for failing to accumulate at least 2500ms of looking during test. All infants attended to both exemplars during the test phase.

4.1.2 | Procedure

Identical to Experiment 2, with one exception: the last two familiarization exemplars were labeled, whereas the first four remained unnamed.

4.2 | Results

4.2.1 | Familiarization

Infants' attention during familiarization (*M* = 12.58, *SD* = 2.78) did not differ from that of Experiment 2, $t(46) = 0.73$, $p = 0.47$, $d = 0.21$.

4.2.2 | Test

Infants in the Reversed SSL condition failed to form the object category (*M* = 0.52, *SD* = 0.13), $t(23) = 0.76$, $p = 0.45$, $d = 0.16$ (see Figure 2), resembling infants in the USL condition (Experiment 1), $t(46) = 0.75$, $p = 0.45$, $d = 0.22$. Performance in the Reversed SSL condition was marginally different from both the SSL, $t(46) = 1.80$, $p = 0.08$, $d = 0.52$, and FSL conditions, $t(46) = 1.90$, $p = 0.06$, $d = 0.55$.

Most strikingly, infants' patterns of attention over the course of the test trial paralleled those in the USL condition (see Figure 3). Infants' looking in Reversed SSL diverged significantly from the SSL condition from 3,450 to 3,850 ms, $p = 0.047$. This is precisely the same window in which the FSL and USL conditions diverged in Experiment 1. Furthermore, performance in the Reversed SSL condition diverged significantly from that in the FSL condition (3,450–3,875 ms, $p = 0.032$), but not the USL condition ($ps > 0.7$).

4.3 | Discussion

Infants' failure to form an object category in the Reversed SSL condition rules out the possibility that two labeled exemplars are, in and of themselves, sufficient for infants to successfully form the novel object category in this task. Instead, infants' failure in the Reversed SSL condition, considered in conjunction with their success in the SSL condition, suggests that infants do engage in SSL: providing infants with two labeled exemplars at the outset of learning provides them with a robust foundation for learning from subsequent, unlabeled exemplars.

5 | GENERAL DISCUSSION

These experiments offer new evidence to reconcile the conceptual power of object labels with their relative scarcity in infants' input. We demonstrate that 2-year-old infants, in the throes of building their lexicons, learn new object categories successfully in a semi-supervised environment. Providing just two labeled exemplars is sufficient for infants to initiate the process of object categorization and to begin incorporating subsequent, unlabeled exemplars into their category representation. Indeed, infants categorized just as successfully in this semi-supervised environment as they did in a fully supervised environment. Moreover, the order of the labeling events matters: infants only benefited when labeled exemplars preceded the unlabeled exemplars (as in SSL), rather than following them (as in Reversed SSL). Intriguingly, this reflects standard practice in machine learning, in which algorithms are traditionally provided with labeled exemplars first.

Critically, our results suggest that infants can take advantage of exactly the kind of input that they receive in acquiring real-world categories. Because caregivers cannot label every new category exemplar a child sees—but do accurately label a subset of them—infants will inevitably receive a mix of labeled and unlabeled exemplars when acquiring most object categories. In principle, infants might respond to this input in different ways: for instance, by ignoring the unlabeled exemplars (supervised learning) or by learning from all exemplars but disregarding the labels (unsupervised learning). Instead, we find that infants deftly exploit all aspects of their input, using labeled exemplars to initiate category formation and incorporating subsequent labeled *and* unlabeled exemplars into those emerging category representations. This outcome is consistent with the view of words as "invitations to form categories" (Waxman & Gelman, 2009; Waxman & Markow, 1995) but takes it one step further by demonstrating that these invitations, while instrumental for initiating category formation, need not be constantly re-issued thereafter.

These findings reveal remarkable continuity across development: success in SSL environments is evident from infancy through adulthood. This work also suggests several intriguing questions. First, although this new evidence from two-year-old infants is striking, it raises the question of how early infants begin to engage in SSL. Second, it remains an open question whether there are indeed some conditions under which infants, or learners of any age, successfully integrate unlabeled exemplars to which they have been exposed before seeing labeled exemplars. Third, it will be important in future work to assess whether and how infants use SSL to identify categories that may be more challenging than the ones we presented here (e.g., Christie & Gentner, 2014; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Waxman, 1998). In particular, future work should assess the power of SSL as infants move beyond identifying single object categories to constructing the many nested and overlapping object categories to which any given exemplar belongs. Discovering how infants rise to such challenges will require that we take into account both their limitations and their tremendous capacity for learning.

Finally, future research should strive to identify with greater precision the mechanism by which labeled and unlabeled exemplars contribute to category learning. Here, we have shown that infants do incorporate unlabeled exemplars after a few labeled exemplars have been presented. What remains to be seen is whether and how these unlabeled exemplars prompt a *shift* in the content of the categories, as they do in adult and machine learners (Gibson et al., 2013). By examining how infants' categories change to reflect each new exemplar, we can also begin to more precisely model infants' learning process. Semi-supervised learning represents a family of algorithms with a common goal—the integration of labeled and unlabeled data—but vastly different computational approaches. Subsequent work should examine more specific algorithms that might underlie infants' learning. In doing so, researchers might also consider how different category learning challenges may alter the value of labels. For instance, while labels may generally be beneficial, their impact is likely to be reduced for more transparent categories (Vong, Navarro, & Perfors, 2015) or for extremely high-dimensional spaces (Hinton & Salakhutdinov, 2006).

In sum, our findings provide the first evidence that infants are capable of SSL. Drawing on research from the machine, adult, and child learning literatures, these results suggest a new framework for infant category learning: infants take advantage of both labeled and unlabeled exemplars to learn novel object categories but rely on labeled exemplars to initiate category formation. In this way, even a small number of labeled exemplars can have an impressive developmental impact.

ACKNOWLEDGEMENTS

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.

ENDNOTES

[1]Although test trials were 20 s in duration to permit infants who were initially looking elsewhere to accrue enough looking time to the test objects, we analyzed only infants' first 5 s of accumulated looking. However, here and in all subsequent experiments, the same overall pattern of results emerges whether analyses include only the first 5 s of the trial or each infant's first 5 s of accumulated looking.

[2]All data and code are available at available at https://github.com/sandylat/ssl-in-infancy.

[3]The outlier was identified using the interquartile range rule, a method of outlier detection suitable for asymmetric data. The overall pattern of results for this experiment is unchanged by this outlier's exclusion.

ORCID

Alexander LaTourrette http://orcid.org/0000-0003-3414-8647

Sandra R. Waxman http://orcid.org/0000-0002-5009-2068

REFERENCES

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PLoS ONE*, *9*(7), e99670. https://doi.org/10.1371/journal.pone.0099670

Althaus, N., & Plunkett, K. (2015). Categorization in infancy: Labeling induces a persisting focus on commonalities. *Developmental Science*, *19*, 770–780. https://doi.org/10.1111/desc.12358

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Balaban, M. T., & Waxman, S. R. (1996). An examination of the factors underlying the facilitative effect of word phrases on object categorization in 9-month-old infants. In A. Stringfellow, D. Cahana-Amitay, E. Hughes, & A. Zukowski, (Eds.), *Proceedings of the 20th Boston University Conference on Language Development* (Vol. 1, pp. 483–493). Somerville, MA: Cascadilla Press.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26. https://doi.org/10.1006/jecp.1996.2332

Balcan, M.-F., Blum, A., Choi, P., Lafferty, J., Pantano, B., Rwebangira, M., & Zhu, X. (2005). *Person identification in webcam images: An application of semi-supervised learning*. Proc. of the 22nd ICML Workshop on Learning with Partially Classified Training Data, Bonn, Germany.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences of the USA*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Chapelle, O., Scholkopf, B., & Zien, A. (2006). *Semi-supervised learning: Adaptive computation and machine learning* (1st edn.). Cambridge, MA: The MIT Press. Retrieved from https://doi.org/10.7551/mitpress/9780262033589.001.0001 https://doi.org/10.7551/mitpress/9780262033589.001.0001

Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science*, *38*(2), 383–397. https://doi.org/10.1111/cogs.12099

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86. https://doi.org/10.1016/j.cognition.2015.06.003

de Carvalho, A., Dautriche, I., & Christophe, A. (2016). Preschoolers use phrasal prosody online to constrain syntactic analysis. *Developmental Science*, *19*(2), 235–250. https://doi.org/10.1111/desc.12300

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B (Methodological)*, *39*, 1–38.

Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis. Available from http://www.eyetrackingr.com.

Feigenson, L., & Halberda, J. (2008). Conceptual knowledge increases infants' memory capacity. *Proceedings of the National Academy of Sciences of the USA*, *105*(29), 9926–9930. https://doi.org/10.1073/pnas.0709884105

Ferguson, B., & Lew-Williams, C. (2016). Communicative signals support abstract rule learning by 7-month-old infants. *Scientific Reports*, *6*, 25434. https://doi.org/10.1038/srep25434

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Gaskins, S. (1999). Children's daily lives in a Mayan village: A case study of culturally constructed roles and activities. *Children's engagement in the world: Sociocultural perspectives* (pp. 25–60). New York, NY: Cambridge University Press.

Gerken, L., & Quam, C. (2017). Infant learning is influenced by local spurious generalizations. *Developmental Science*, *20*(3), e12410. https://doi.org/10.1111/desc.12410

Gibson, B. R., Rogers, T. T., & Zhu, X. (2013). Human semi-supervised learning. *Topics in Cognitive Science*, *5*(1), 132–172. https://doi.org/10.1111/tops.12010

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, *1*(1), 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goldberg, A. B., & Zhu, X. (2006). Seeing stars when there aren't many stars: Graph-based semi-supervised learning for sentiment categorization. In *Proceedings of the first workshop on graph based methods for natural language processing* (pp. 45–52). Stroudsburg, PA: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1654758.1654769

Guillaumin, M., Verbeek, J., & Schmid, C. (2010). *Multimodal semi-supervised learning for image classification*. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 902–909). https://doi.org/10.1109/cvpr.2010.5540120

Hahn, N., Snedeker, J., & Rabagliati, H. (2015). Rapid linguistic ambiguity resolution in young children with autism spectrum disorder: Eye tracking evidence for the limits of weak central coherence. *Autism Research: Official Journal of the International Society for Autism Research*, *8*(6), 717–726. https://doi.org/10.1002/aur.1487

Havy, M., & Waxman, S. R. (2016). Naming influences 9-month-olds' identification of discrete categories along a perceptual continuum. *Cognition*, *156*, 41–51. https://doi.org/10.1016/j.cognition.2016.07.011

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Kalish, C. W., Rogers, T. T., Lang, J., & Zhu, X. (2011). Can semi-supervised learning explain incorrect beliefs about categories? *Cognition*, *120*(1), 106–118. https://doi.org/10.1016/j.cognition.2011.03.002

Kalish, C. W., Zhu, X., & Rogers, T. T. (2015). Drift in children's categories: When experienced distributions conflict with prior learning. *Developmental Science*, *18*(6), 940–956. https://doi.org/10.1111/desc.12280

Lake, B., & McClelland, J. (2011). Estimating the strength of unlabeled information during semi-supervised learning. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1400–1405). Austin, TX: Cognitive Science Society. Retrieved from http://web.mit.edu/brenden/www/LakeMcClelland2011CogSci.pdf

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In *Input and interaction in language acquisition* (pp. 56–73). New York, NY: Cambridge University Press. https://doi.org/10.1017/cbo9780511620690.005 https://doi.org/10.1017/CBO9780511620690

Lu, W.-L., Ting, J.-A., Little, J. J., & Murphy, K. P. (2013). Learning to track and identify players from broadcast sports videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(7), 1704–1716. https://doi.org/10.1109/TPAMI.2012.242

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McDonnell, J., Jew, C., & Gureckis, T. (2012). Sparse category labels obstruct generalization of category membership. *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 749–754.

Nigam, K., Mccallum, A. K., Thrun, S., & Mitchell, T. (2000). Text classification from labeled and unlabeled Documents using EM. *Machine Learning*, *39*(2–3), 103–134. https://doi.org/10.1023/A:1007692713085

Perszyk, D. R., & Waxman, S. R. (2018). Linking language and cognition in infancy. *Annual Review of Psychology*, *69*(1), 231–250. https://doi.org/10.1146/annurev-psych-122216-011701

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. *Perception*, *22*(4), 463–475. https://doi.org/10.1068/p220463

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*(6), 909–957. https://doi.org/10.1111/j.1551-6709.2009.01092.x

Rogers, K. C., Gibson, B. R., Harrison, J., & Zhu, X. (2010). Semi-supervised learning is observed in a speeded but not an unspeeded 2D categorization task. *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 2320–2325.

Rogoff, B., Mistry, J., Göncü, A., & Mosier, C. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*, *58*(8, Serial No. 236).

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2016). The pursuit of word meanings. *Cognitive Science*, *41*, 638–676. https://doi.org/10.1111/cogs.12416

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development*, *57*(6), 1454–1463. https://doi.org/10.2307/1130423

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Vandist, K., De Schryver, M., & Rosseel, Y. (2009). Semisupervised category learning: The impact of feedback in learning the information-integration task. *Attention, Perception, & Psychophysics*, *71*(2), 328–341. https://doi.org/10.3758/APP.71.2.328

Vong, W. K., Navarro, D. J., & Perfors, A. (2015). The helpfulness of category labels in semi-supervised learning depends on category structure. *Psychonomic Bulletin & Review*, *23*(1), 230–238. https://doi.org/10.3758/s13423-015-0857-9

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development*, *5*(2), 123–150. https://doi.org/10.1016/0885-2014(90)90023-M

Waxman, S. R. (1998). Linking object categorization and naming: Early expectations and the shaping role of language. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 38, pp. 249–291). San Diego, CA: Academic Press.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition*, *95*(3), B59–B68. https://doi.org/10.1016/j.cognition.2004.09.003

Waxman, S.R., & Gelman, S.A. (2009). Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*, *13*(6), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302. https://doi.org/10.1006/cogp.1995.1016

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two- and three-year-olds track a single meaning during word learning: Evidence for Propose-but-verify. *Language Learning and Development: The Official Journal of the Society for Language Development*, *12*(3), 252–261. https://doi.org/10.1080/15475441.2016.1140581

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition*, *85*(3), 223–250. https://doi.org/10.1016/S0010-0277(02)00109-9

Xu, J.-M., Fumera, G., Roli, F., & Zhou, Z.-H. (2009). Training spamassassin with active semi-supervised learning. In Proceedings of the 6th Conference on Email and Anti-Spam (CEAS'09). Retrieved from http://pages.cs.wisc.edu/~xujm/pub/ceas09.pdf

Zhu, X. (2005). *Semi-supervised learning literature survey*. Madison, WI: University of Wisconsin-Madison.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *3*(1), 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006

Zhu, X., Rogers, T., Qian, R., & Kalish, C. (2007). Humans perform semi-supervised classification too. In *Proceedings of the National Conference on Artificial Intelligence* (Vol. 22, p. 864). Menlo Park, CA; Cambridge, MA; London: AAAI Press; MIT Press; 1999. Retrieved from http://www.aaai.org/Papers/AAAI/2007/AAAI07-137.pdf

How to cite this article: LaTourrette A, Waxman SR. A little labeling goes a long way: Semi-supervised learning in infancy. *Dev Sci*. 2018;e12736. https://doi.org/10.1111/desc.12736