NORTHWESTERN UNIVERSITY

Perceptual and Linguistic Factors in Infants’ Relational Learning

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology

By

Erin M. Anderson

EVANSTON, ILLINOIS

June 2020

Abstract

Humans are prodigious learners. One of our most powerful learning tools is the ability to detect and transfer relational similarities between items and events, despite their perceptual differences. Previous research has found that the roots of this ability extend into infancy. As early as three months of age, infants can recognize the abstract relations *same* and *different*, (Anderson, et al., 2018). Additionally, infants already show two patterns that characterize relational learning in older children and adults: 1) learning improves when they have greater opportunities to compare multiple exemplars of a relation, and 2) learning is hindered when surface differences are highlighted instead. These results indicate consistencies in this learning process across the lifespan.

However, there are numerous differences between infants, children and adults. Older learners benefit from mature perceptual and attentional processes, improved working memory, and from experience with abstract and relational language. This dissertation focuses on how these factors interact with relational learning in the first year of life. Chapter 1 asks how the maturation of perceptual processes and short term memory in the first year affects recognition and generalization. Chapter 2 examines the impact of labeling the relations at 12 months, compared to labeling the objects involved. Finally Chapter 3 asks if infants can transition from context-specific learning to more generalizable representations of the relation in a short period of time. Ultimately, these studies highlight that what improves relational alignment depends on the age of the learner, by pointing to a protracted period over which the relational learning process becomes linked to factors that will later benefit it.

Acknowledgements

Thank you to past and present members of Infant Cognition lab at Northwestern University, especially to my advisor Sue Hespos, who has always (gently) pushed me out of my comfort zone while making it clear that she has my back, and who has recently devoted innumerable hours to catching my mistakes. Thanks, too, to Julie Hoather and Shelley Powers for contacting parents, scheduling babies, conducting last-minute stimuli repairs, handling my brooding about cancellations for revise and resubmit or dissertation studies with good humor, and for generally acting as emotional ballasts. Thanks to the parents and infants of Chicagoland who came from near and far to participate and without whom I would have no dissertation.

Thank you to Dedre Gentner and Lance Rips for being invaluable collaborators, and to the many members of their labs for countless insights. I've been lucky to benefit from your expertise and critical eyes, and I am touched by how you handle my over-enthusiastic caroling every winter (I'm still working on that high note in Silent Night). Thanks.

Thanks to my cohort and "extended" cohort. The best thing for someone with social anxiety leaving their hometown family and friends for the first time was, apparently, a big bunch of smart, considerate extroverts who had organized a regular calendar of social events within the first six months. Some of these events, like Wednesday night pub trivia and Podcast Club, are still going even during this shelter-in-place (albeit virtually). In these strange times, I am that much grateful for the stability these events provide and for their attendees, who are among my closest friends. Special thanks to Francisco Maravilla, my former office-mate, current band-mate and best friend here, for the memes, for the sad songs and for always listening.

Finally, thanks to my family for all the support they have given me through this program. To my parents, thank you for your love and unconditional positive regard. To Robert: you're the person I rely on the most. Thank you for moving north with me without any reservations, and thank you for making me breakfast almost every morning of this shelter-in-place dissertation marathon so that I'm not living off of coffee.

Table of Contents

List of Figures

List of Tables

Chapter 1: Infant perceptual processes and relational learning

Analogical ability is the ability to recognize common relations between events and generalize them beyond the current context. This has been argued to be a key component of higher-order cognition (Christie & Gentner, 2010; Gentner, 2003; Holyoak & Thagard, 1995). By permitting learners to make deep structural connections between events (Gentner & Medina, 1998), analogical ability is implicated in scientific discovery and creative problem-solving (e.g., modeling an atom's structure on that of the solar system or deriving self-driving rideshare routes from the foraging patterns of ants). However, it is not always obvious to a learner that a relation is shared between events. The challenge of perceiving these shared relations is increased when individual differences between events (like the objects involved) are made more salient (Markman & Gentner, 1993; Richland & McDonough, 2010). Young children, who might benefit the most from the insights offered by analogical learning, are particularly susceptible to focusing on the object features of individual exemplars at the cost of perceiving the relation (Gentner & Rattermann, 1991; Gentner & Toupin, 1986; Paik & Mix, 2006; Rattermann & Gentner, 1998; Richland et al., 2006).

How do young children learn a relation, given the draw of individual objects? A reliable answer has been comparison. Comparison has been theorized to promote a process of structural alignment, in which the relational structures that are shared between otherwise distinct examples are put into alignment with each other and become more apparent (Falkenhainer et al., 1989; Gentner & Markman, 1997). Supporting this, children are more likely to generalize new relations when they have an opportunity to compare two examples than when they see only one (Christie & Gentner, 2010; Gentner et al., 2011; Gentner & Namy, 1999; Kotovsky & Gentner, 1996;

Richland et al., 2017). These learning patterns – that comparing items helps relational generalization while focusing on individual objects hinders it – have been identified as signatures of analogical learning.

Recent work has shown that analogical ability is present early. Infants as young as three months can generalize relations by comparing across serial examples in visual and linguistic domains (Anderson et al., 2018a; Ferry et al., 2015; Gerken & Bollt, 2008; Gervain et al., 2008, 2012; Gómez, 2002; Hochmann et al., 2018; Johnson et al., 2010; Marcus et al., 1999) For example, Ferry, Hespos & Gentner (2015) found that infants show the same signatures of analogical learning as older children, specifically: 1) comparing multiples exemplars of a relation facilitates learning, and 2) rendering objects individually salient hinders learning. In support of the first signature, infants at 7 and 9 months could abstract *same* or *different* after habituating to a series of exemplars (e.g., AA, BB, CC, DD for *same*) but not after seeing only one exemplar. In contrast, these same infants failed to generalize on test trials that featured objects that had been made individually salient, showcasing the second signature. Finding these patterns in the first year suggests that the analogical learning mechanism found in infancy may be continuous with the one seen at other ages.

This is not to argue that analogical ability is already present in a mature form in infancy. Certainly, infants have not demonstrated the higher-order relational abilities that adults are capable of (see Gentner, 2003 for examples). These differences in ability could be due to other factors which interact with analogical learning. The theory of developmental cascades highlights how even when a single learning mechanism is continuous from infancy into adulthood, the interaction with the many other abilities that are changing and maturing will converge to produce variation in learning outcomes based on age and individual differences (Oakes & Rakison, 2020). A major change that has not been investigated in regards to infants' relational learning is the dramatic maturation in infants' encoding abilities. In the first months of life, vision and attentional processes are changing rapidly, becoming increasingly stable throughout the first year (for reviews see Arterberry & Kellman, 2016; Colombo, 2001). Visual acuity improves steadily throughout the first several months. While newborns have an acuity of 20/400, by 8 months, infants have acuity near adult levels (Arterberry & Kellman, 2016; Norcia & Tyler, 1985). Infants' attention also improves over this time, with habituation and fixation periods decreasing dramatically during the first 6 months (Bornstein et al., 1988; Colombo & Mitchell, 2009). Furthermore, infant's working memory is improving. While infants as young as 3 months can remember that a single object is present through periods of occlusion (Baillargeon & DeVos, 1991), infants become capable of tracking the presence of two occluded objects by 6 months (Kibbe & Leslie, 2016), and up to three objects between 8 months and 14 months (Feigenson & Carey, 2005; Kibbe & Leslie, 2013). Even when infants can remember the number of objects that are present, they still may not encode all the features. In fact, when objects are occluded and then revealed, four-month-old infants show failures to recognize changes in color, texture, and / or location (Mareschal & Johnson, 2003; Ross-Sheehy et al., 2003; Simon et al., 1995; Wilcox, 1999), and six-month-olds only remember the features of one object when two are occluded (Káldy & Leslie, 2005; Kibbe & Leslie, 2019; Oakes et al., 2013). Only as infants reach 8 to 12 months do they show short-term memory for multiple objects and their features (Kibbe & Feigenson, 2016; Ross-Sheehy et al., 2003)

In sum, infants demonstrate recognition failures during the first few months of life, but reveal increasingly stable encoding towards the end of the first year. This pattern of development implies that there is a considerable amount of variability between what young infants are presented with and what they remember. For infants 6 months or younger, object features may not be fully retained from one presentation to the next. Given this high likelihood of recognition failure, even one example that is presented again and again could be perceived as multiple distinct items (though perhaps highly similar ones). In this paper, we investigate what these changes mean for relational learning and the role of comparison. We make the counterintuitive proposal that, for young infants, the opportunity to compare “multiple” exemplars of a relation could occur even when infants are only presented with repetitions of one exemplar. If this is the case, infants might be able to generalize relations from a single pair presented repeatedly. This ability would be dependent on the age of the infant. Infants nearing the end of the first year should be able to recognize that repetitions are just that: a single exemplar, whereas younger infants would be more likely to generalize from them. In sum, developmental changes due to still-maturing perceptual process and visual short-term memory could interact with the infants’ relational learning in a visual habituation paradigm.

Our experiment tests this prediction by presenting 3-month-old and 7-to-9-month-old infants with a series of repetitions of a single example (e.g., a pair of identical objects for *same*), and then testing whether the infants form a relational abstraction (e.g., *same* (X, X)). If neither age group generalizes from this sequence in our experiment, this will be consistent with the possibility that relational abstraction depends on multiple exemplars. However, if we find generalization in 3-month-old infants, but not in 7-to-9-month-old infants, this will support our hypothesis that developmental changes in acuity and short-term memory creates the perception of a set of distinct exemplars, allowing for comparison and relational abstraction.

Experiment 1

This experiment asks two questions. First, is it possible for infants to abstract and generalize a relation like *same* or *different* from just one exemplar? Second, is there a developmental difference linked to the maturity of infants' encoding, such that younger infants will be more likely to generalize from one exemplar than will older infants? On a larger scale, these questions can ascertain whether relational learning interacts with the development of basic vision and memory processes in the first year.

To test whether 3-month-old infants and 7- or 9-month-old infants could abstract the *same* or *different* relation from repetition to a single exemplar, we used a habituation/ dishabituation looking time paradigm that had successfully captured relational learning in 3-month-olds (Anderson et al., 2018). This paradigm uses three test trial types to look for particular patterns of learning (see Figure 1). The critical test of whether infants could abstract the relation was generalization to new pairs of toys, discriminating those that made up a novel relation from those that composed a familiar relation (*New*). Additionally, we included test trials that included the exact pair seen during the habituation period, as a manipulation check of whether infants were discriminating at all (*Memory Check*). Finally, we looked for a consistent error pattern indicated by the relational abstraction literature. Previous research shows that infants who can generalize *same* or *different* to new pairs look equally between pairs of items shown individually before the experiment (Ferry et al., 2015; Anderson et al., 2018). We likewise predict that if infants can generalize from one exemplar, they will still fail to discriminate the familiar relation from the novel one in a pair that is made of objects made salient before test (*Salient Object*). If the predicted patterns of looking are found for each test trial (looking longer at the novel relation in the *Memory Check* and *New* pairs, but looking equally between relations in the *Salient Object* pair), it will suggest that infants are using relational learning processes to succeed, even with one exemplar. We also consider habituation condition in our analyses. While none of the studies using our paradigms have found an advantage for *same* over *different*, it remains an open question about whether this will be the case when infants are only given a single exemplar. Additionally, infants are able to recall the features of two identical objects earlier than they can with two distinct objects (see Oakes, et al., 2013).

Finally, a habituation/ dishabituation paradigm assumes that learning takes plan during the habituation phase. Therefore, it is important to measure the decline in looking in both age groups. We predict that both age groups should show a habituation decline, but because 7- and 9-month-olds have better visual acuity and short-term memory, we might expect an age-based difference in the rate of this decline, with the older infants showing a steeper decline.

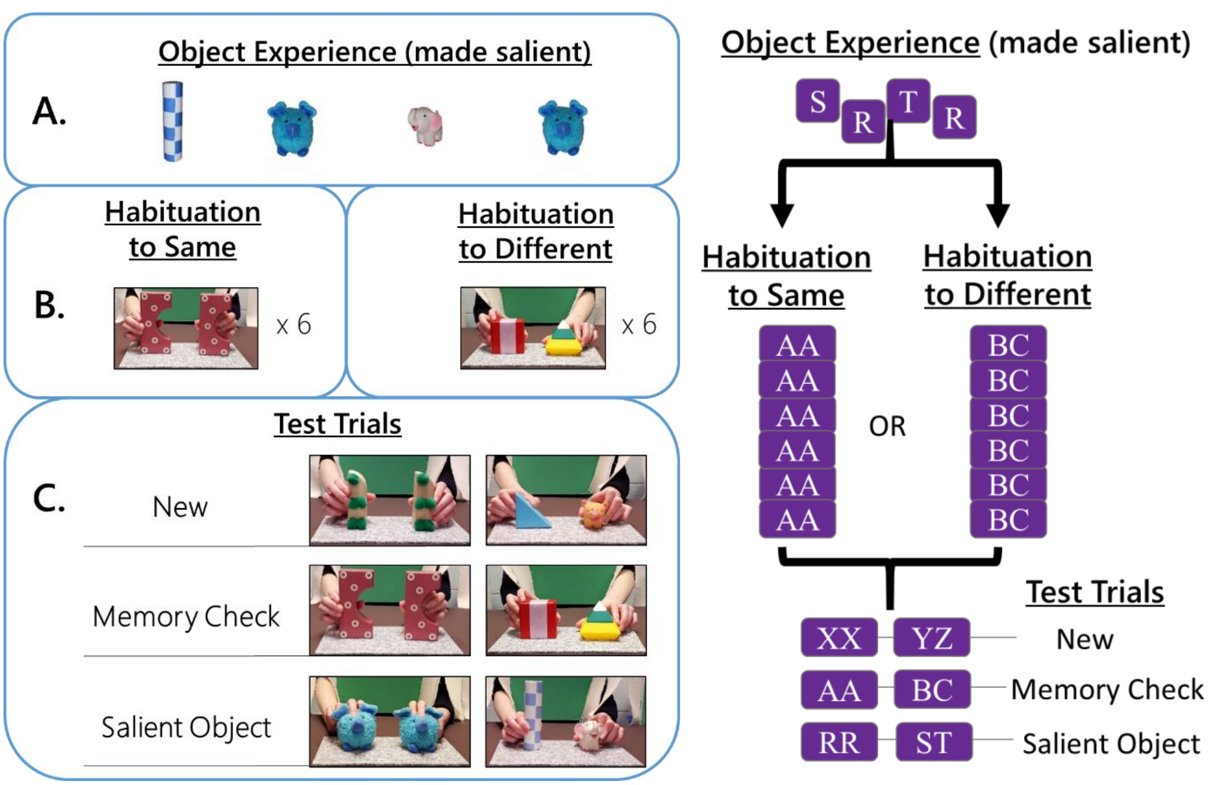

Figure 1. Pictures of the stimuli used in Experiment 1 and a schematic of the procedure. (A) Infants saw a subset of individual toys before the experiment. (B) Infants were habituated to one pair of objects, either same or different. (C) Infants saw six sequential test trials alternating between novel and familiar relational pairs, where the dependent measure was looking time. There were three types of test trials: *New*, *Memory Check*, and *Salient Object*. The order of these test trial types were counterbalanced across participants.

Methods

Participants

The participants were 47 healthy, full-term 3-month-old infants (23 female, mean age 3 months, 14 days), and 46 7- to 9-month-olds (25 female, mean age 8 months, 1 days). The target sample size (*N* = 48 for each age group) was determined based on power analyses of 3-month-olds who demonstrated relational learning on *New* trials in Anderson et al. (2018) and 7- and 9-month-olds who did so in Ferry et al. (2015). In both cases, the partial eta-squared was .18, and a sample size of 48 provided a .89 power level. Half of the infants in each age group were assigned to the *same* condition, and the other half to the *different* condition. Twenty-two additional infants participated but were excluded. Seventeen of these were in the younger age group: 10 were eliminated because of fussiness (defined as two independent coders judging the infant's state as fussy or crying for at least half the test trials), 3 for having bowel movements during the test trials, and 4 because their looking time was more than two standard deviations above the mean during test. Five of these were in the older age group: 2 were eliminated because of fussiness, 2 their looking time was more than two standard deviations above the mean during test, and 1 because coders could not reach agreement on the length of the infant's looks after multiple recodes.

Parents of infants were recruited through online ads and word of mouth. Parents who agreed to their infants' participation were provided informed consent before the experiment and given $20 as compensation. The race, ethnicity and parental education level of the sample is described in the table below.

Table 1.

Participant demographics for Experiment 1

Ethnicity	*%*	*Race*	*%*	*Education*	*% Maternal*	*% Paternal*
Hispanic	22	African American	9	Some high school	0	0
Non-**Hispanic**	72	Asian/ Hawaiian/ Pacific Islander	1	High school diploma	1	3
		American Indian	0	Some college	4	9
		White	79	College degree	92	85
		Multiracial	5			
No response	6	No response	6	No response	3	3

Apparatus

Parents sat in a chair with infants on their lap facing a wooden puppet stage that displayed all stimuli. The parents were asked to refrain from interacting with the infant during the experiment and to close their eyes during the test trials. The stage measured 243.5 cm high, 128 cm wide, and 61 cm deep. The opening in the front of the stage that displayed the objects was 93 cm above the floor, 61 cm high, and 106 cm wide. The back wall had two rectangular openings with cloth fringe over the openings that allowed the experimenter to manipulate the objects between trials. A screen covered the infants' view of the stage between trials. The MATLAB program Baby Looking Time (BLT), was used to record looking times for habituation and test trials during the experiment (Chang et al., 2018).

The stimuli consisted of 12 objects (see Figure 1). During the experiment, each pair of objects was placed on the puppet stage on a 26.5 x 15.5 cm cardboard tray that was covered with contact paper. Four of the objects were presented individually before the experiment began. A single pair of objects was seen in the habituation phase within *same* or *different* pairs: either two pink dotted arches together or a red striped square with a yellow and green pyramid. Finally, all 12 objects appeared in pairs during the test phase. Of these, four objects had been seen individually, two had been seen in habituation, and four objects had not been seen before test.

Procedure

The experiment consisted of three sequential parts:

(A) Objects made salient. We manipulated infants' attention on some individual objects by showing four of the test objects to the infant in the waiting room prior to the experiments, during naturalistic play interactions. Showing the objects one at a time, the experimenter held each object between the infant and themselves for 5 seconds while they jointly attended to it and made comments such as "Look!" and "See this one?". The two identical objects were never shown in immediate succession.

(B) Pair habituation trials. When the screen was raised at the start of every trial, a pair of objects rested on the cardboard tray on the stage. To engage infants' attention, in both habituation and test trials, the pairs of objects were moved during the trial. The experimenter grasped one object in each hand and raised the objects straight up (1 s), tilted them to the left (1 s), returned them to the center (1 s), tilted them to the right (1 s), returned them to the center (1s), returned them to the tray (1 s), and paused on the tray (2 s). This 8-s cycle repeated continuously until the trial ended. In the *same* condition, infants saw habituation trials in which the pairs of objects were the *same* (see Figure 1). In the *different* condition, infants saw habituation trials in which the pairs of objects were *different*. The number of habituation trials was infant-controlled (see procedure section for the criterion), ranging from 6 to 9 trials.

(C) Test trials. For each test trial type, infants were presented with a novel relation, followed by a familiar relation (or vice versa), summing to six total test trials. In each test trial, infants viewed one pair of objects, presented in the same motion pattern as in the habituation trials, while their looking time was recorded. The three kinds of test trials were (a) objects that had not been seen before test (*New*); (b) objects that the infant had experienced individually in the waiting room (*Salient Object*); and (c) objects that the infant had seen presented as a pair in the habituation trials (*Memory Check*). There were three trial orders (abc, cab, bca) counterbalanced across infants.

Coding

There was a small hole in the front face of the stage containing a camera that captured a video image of the infant's face. While the experimenter conducted habituation and test trials in the room with the infants, two research assistants in a separate room viewed the video and coded infants' visual fixations online as either on target or off. Each researcher depressed a computer button when the infant attended to the events on stage and released the button when the infant looked away. Each trial ended when the infant either looked away for 2 consecutive seconds after having looked at the event for at least 2 s or looked at the event for 60 cumulative seconds without looking away for 2 consecutive seconds. The BLT program determined the end of the trial and beeped, signaling to the experimenter to lower the screen and move to the next trial. After each test trial, research assistants checked one or more boxes to indicate the behavioral state of the infant on the preceding trial: sleepy, quiet and alert, active, fussy or crying. Coders also noted any breaks and their length. As noted above, if two coders independently judged the infant's state as fussy, crying, or falling asleep for more than half the test trials, the infant's data was excluded from the analysis. The coders were blind to the condition and the trial order.

Inter-observer agreement was measured for all infants and averaged 91%. The Intraclass Correlation Coefficient for a fixed set of raters (ICC3) was .998 with a 95% confidence interval from .996 to .998, $F(3, 95) = 899.92$, $p < .001$. Our looking time data significantly deviated from a normal distribution per the Shapiro-Wilks test. Therefore, we performed parametric tests on log-transformed data, following recommendations outlined by Csibra, Hernik, Mascaro, Tatone, and Lengyel (2016).

Results

The two main questions were whether infants could learn *same* / *different* relations from repetitions of one exemplar, and whether this interacted with age group. The gold standard for relational learning is generalization to never-before-seen objects, so we conducted a repeated-measures ANOVA on the *New* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Age Group (3 months vs. 7 and 9 months), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). After seeing only one exemplar, did infants look longer at novel relations in the *New* pair? The answer was yes: averaging across age groups and conditions, infants looked 28.68 s (*SD* = 19.94) at novel relations and 17.70 s (*SD* = 16.26) at familiar ones. The ANOVA revealed a main effect of relation, such that infants looked longer at pairs instantiating novel relations than familiar ones, $F(1, 77) = 4.84$, $p = .031$, $\eta^2_p = .06$. The non-parametric tests suggested different rates of generalization across age groups: For *New* trials, 29 of the 46 3-month-olds looked longer at the novel relation[1], $p = .052$, while only 25 of 46 7- and 9-month-olds did the same, $p = .33$. However, the ANOVA showed no main effect of age

[1] The final sample include 47 3-month-olds, but only 46 had usable *New* pairs.

group on looking times during this test trial type, $F(1, 89) < 1$, $p = .525$, nor an interaction between age group and looking between relations, $F(1, 77) = 1.074$, $p = .303$ (see Figure 2). There was a unpredicted interaction between sex and condition on looking time, such that male infants showed longer looking in the habituation to same condition compared to the habituation to different condition longer, while female infants looked equally in both conditions, $F(1, 77) = 4.096$, $p = .046$, $\eta^2_p = .05$. However, this interaction did not depend on whether infants were looking at the novel or familiar relation, $F(1, 77) < 1$, $p = .998$. There was also a four-way interaction between the between-subjects factors of sex, order, condition, and age group, $F(1, 77) = 5.984$, $p = .017$, $\eta^2_p = .07$. This four-way interaction did not interact significantly with relation, $F(1, 77) = 1.189$, $p = 0.279$. No other effects or interactions were significant.

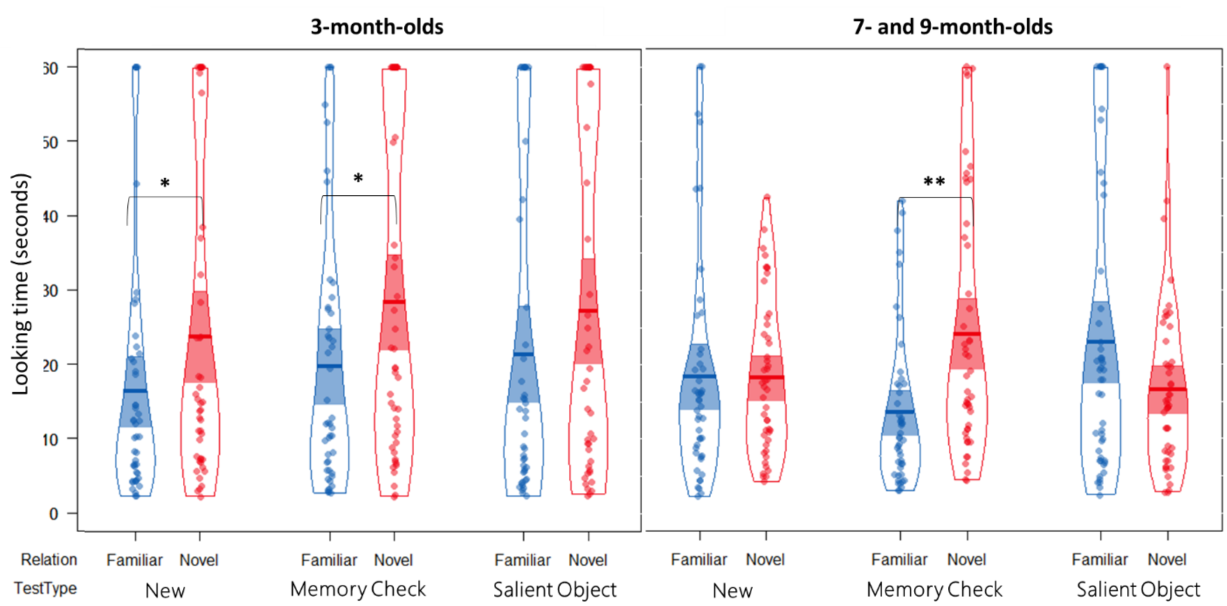

Figure 2. Experiment 1 looking times at novel and familiar pairs for each test type separated by age group. The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value. * indicates $p < .05$ and ** indicates $p < .001$.

Next, we conducted a manipulation check to see whether infants could identify the exact pair seen in habituation. Here, we conducted a repeated-measures ANOVA on the *Memory Check* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Age Group (3 months vs. 7 and 9 months), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). For the Memory check trial, infants also showed discrimination: averaging across groups, infants looked 26.19 s (*SD* = 19.39) at novel relations and 16.56 s (*SD* = 14.43) at familiar ones. Supporting this, there was a main effect of looking at the novel pair compared to the familiar one from habituation, $F(1, 77) = 26.55$, $p < .001$, $\eta^2_p = .26$. Examining the non-parametric data, 40 of 48 7- and 9-month-olds preferred the novel pair to the one seen in habituation, $p < .001$, while only 29 of 46 3-month-olds did, $p = .052$. However, the ANOVA showed no main effect of age group on looking times during this test trial type, $F(1, 77) = 1.25$, $p = .27$, nor an interaction between age group and looking between relations, $F(1, 77) < 1$, $p = .44$. No other effects or interactions were significant.

Finally, we wanted to know whether infants failed to generalize to pairs containing objects that had been made salient, a signature of relational learning. We conducted a repeated-measures ANOVA on the *Salient Object* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Age Group (3 months vs. 7 and 9 months), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). For *Salient Object* test trials, infants looked near equally at the novel relation (*M* = 21.57 s; *SD* = 18.60) and the familiar one (*M* = 22.21 s; *SD* = 20.08), $F(1, 75) < 1$, and there was no main effect of relation, $p = .735$. Examining the non-parametric data, there were not significant numbers of infants in either age group that preferred the novel relation: only 18 of 48 7- and 9-month-olds did, p = .97, and 24 of 43 3-month-olds did, p = .27. Similarly the ANOVA found no main effect of age group on looking times during this test trial type, $F(1, 75) < 1$, p = .923. However, there was an interaction between age group and relation, $F(1, 75) = 6.062$, p = .016, η^2_p = .08. That is, while we expected both age groups to look similar on the *Salient Object* test trials, we instead found a slight divergence: 3-month-olds preferred the novel relation (M_{Novel} = 27.21 s; $M_{Familiar}$ = 21.39 s), but 7- and 9-month-olds preferred the familiar relation (M_{Novel} = 18.53 s; $M_{Familiar}$ = 22.94 s), and neither reflected significant looking differences (see next section). There was also an interaction between relation and test order, such that infants who saw the novel relation first during each test pair looked longer at the familiar relation, but infants who saw the familiar relation first looked longer at the novel relation, $F(1, 75) = 4.114$, p = .046, η^2_p = .05. No other effects or interactions were significant.

Generalization by test type and age

Next, we conducted a series of planned comparisons within each age group to examine which test trials yielded looking differences between the novel and familiar relations. These comparisons help answer whether one age group was driving the effect of relation in *New* and *Memory Check* trials where there was no interaction with age group, and for the *Salient Object* trial where there was an interaction with age, whether this was due to significant discrimination at each age or a slight divergence in preferences. Again, the critical question was whether infants generalized the relation to *New* pairs within each age group. For 3-month-olds, this was true: they looked an average of 23.72 s at the novel relations but 16.44 s at the familiar relation, $t(45) = 2.337$, p = .024, Cohen's d = .35. In contrast, 7- and 9-month-olds looked an average of 18.11 s at the novel relations but 18.35 s at the familiar relation, *t(46*) < 1, $p = .37$. We also predicted that both age groups would look longer at the novel relation in the *Memory Check*: this was the case for 3-month-olds ($M_{Novel} = 28.38$, $M_{Familiar} = 19.72$), $t(45) = 2.792$, $p = .008$, $d = .41$, as well as for 7- and 9-month-olds ($M_{Novel} = 24.05$, $M_{Familiar} = 13.48$), $t(46) = 4.653$, $p < .001$, $d = .68$. Finally we predicted that neither age group would discriminate the relations with salient objects, and neither 3-month-olds ($M_{Novel} = 27.21$ s, $M_{Familiar} = 21.39$ s), $t(42) = 1.485$, $p = .145$, nor 7- and 9-month-olds did ($M_{Novel} = 16.52$ s, $M_{Fam} = 22.94$ s), $t(47) = -1.601$, $p = .116$.

Habituation patterns by age

Finally, we examined looking time during the habituation phase. Because 3-month-olds have limited visual acuity and memory compared to 7- and 9-month-olds, they tend to look longer at stimuli in general, and we predicted that there might be differences between age groups in the habituation decline. A repeated measures ANOVA examined looking time by the within-subjects factor of habituation trial (first through last trial) and the between-subjects factor of age group (3 vs. 7 and 9 months). Collapsing across age groups, this revealed a significant decline in looking between habituation trials, $F(5, 450) = 58.35$, $p < .001$, $\eta^2_p = .35$ (see Figure 3). On average, infants looked 45.74 s on their first habituation trial and 15.85 s on their last trial. While 7- and 9-month-olds showed a more dramatic decline between the first and last three trials (M = 45.80 s vs. M = 12.78 s) than 3-month-olds did, (M = 45.67 s vs. M = 18.80 s), there was no significant interaction between habituation trial and age group, $F(5, 450) = 2.02$, $p = .072$.

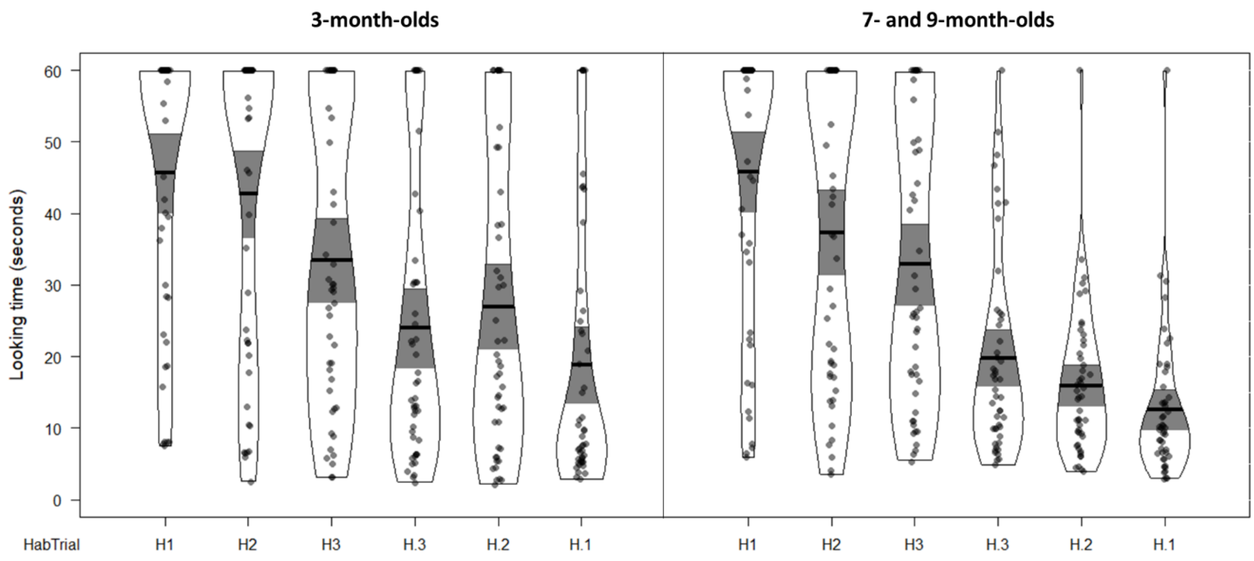

Figure 3. Looking decline over the course of 6 to 9 habituation trials, by age group. Because the number of habituation trials depended on the individual infant, the graphs represent the first three (H1, H2, H3) and the last three trials (H-3, H-2, H-1). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

While there was not a between-group difference on average looking time during habituation, $F(1, 90) = 2.47$, $p = .12$, there was a difference in total looking time during habituation, $F(1, 88) = 6.52$, $p = .012$, $\eta^2_p = .07$, such that younger infants looked longer over the course of habituation than older infants did (199.04 s vs. 54.74 s). Younger infants also took slightly longer to reach the habituation criterion, with an average of 7.60 trials compared to older infants' 7.06 trials, $F(1, 88) = 4.08$, $p = .046$, $\eta^2_p = .04$. Additionally, there was an effect of age group on whether infants reached the habituation criterion or not (65% of 3-month-olds compared to 88% of 7- and 9-month-olds), X^2 $(1, N = 95) = 7.146$, $p = .008$. There were also no main effects or interactions with habituation condition (*same* vs. *different*) in any of these analyses.

Table 2.

Habituation Details by Age Group

Age group	Habituated? (Y/N)		Number of trials		Total habituation time	
	M	*SD*	*M*	*SD*	*M*	*SD*
3 months	65%	48%	7.60	1.33	254.75 s	109
7 & 9 months	88%	33%	7.06	1.28	199.04 s	99

Discussion

Overall, infants in our experiment learned the *same-different* relations, replicating previous work with these age groups (Anderson et al., 2018a; Ferry et al., 2015). Furthermore, infants showed the second signature of relational learning, consistent with prior findings on analogical processing (Christie & Gentner, 2010; Rattermann & Gentner, 1998; Richland et al., 2006). Specifically, infants failed to detect the relations between objects that had been made individually salient, despite the fact that they generalized relations to *New* trials and discriminated objects from the habituation phase (*Memory Check*). The evidence of relational learning is surprising given that infants were only presented with one exemplar, and that the vast majority of evidence on relational learning suggests that learners are considerably more likely to abstract a new relation when they can compare two or more exemplars (Christie & Gentner, 2010; Gentner et al., 2011; Gick & Holyoak, 1983; Richland et al., 2017).

The second part of our research questions was whether infants showed different patterns of learning depending on their age. We predicted that younger infants with less developed visual acuity and memory would be less likely to perceive repetitions as a single exemplar and therefore more likely to broadly generalize the relation. Our results showed only slight support for this hypothesis. The within-age group analyses revealed that three-month-old infants generalized *same* or *different* to *New* trials, while 7- and 9-month-olds failed to do so. This suggests that although 3-month-olds were only presented with one repeated example, these younger infants failed to recognize it as a single pair throughout. In contrast, 7- and 9-month-olds learned only the exact pair of toys, which they demonstrated by successfully passing the *Memory Check* (recognizing the pair from habituation) but not generalizing to *New*. Despite these differences, however, our ANOVA did not yield a significant interaction with age group in our analysis of variance, indicating that within-age-group variation was greater than between-group variation. This lack of interaction could be due to our limited age span across participants. Additionally this experiment focused on a cluster of encoding-related processes that show dramatic improvements in the first several months, but it is unclear which of these (acuity, scanning and attention, or short-term memory capacity) is the most reliable predictor of the recognition failures seen in the literature. One the one hand, visual acuity nears adult levels by 8 months (Norcia & Tyler, 1985) but visual short-term memory continues to improve through the first twelve months and beyond (Kibbe & Feigenson, 2016), and even 11-month-olds may fail to individuate hidden objects depending on the context (Stavans et al., 2019).

Our findings that recognition failures may be linked to greater generalization has implications beyond our paradigm. Gerken et al., (2015) found generalization from a one repeated linguistic example at 9 months. The authors argued that this overgeneralization was due to the rarity of the linguistic structures that they included. However, we might ask how short-term memory interacts with linguistic stimuli as well. Even in adults, auditory memory is less robust than visual memory (Cohen et al., 2009). The 9-month-old infants may have experienced recognition failures with these spoken stimuli, similar to what we theorize occurred with the younger infants in our paradigm. If this interpretation is correct, it would point to how perceptual and memory processes might be involved in relational learning outcomes on a broader level.

Regardless of age or modality, there is likely to be a sweet spot to when recognition failures can facilitate relational learning and when they would hinder it. With identical or highly similar items that are easy to compare and align, recognition failures could increase generalization. With less similar stimuli, however, more recognition failures could make it harder to align the shared relations. Future studies could examine whether making it more or less challenging to recognize that visual presentations are repetitions of one exemplar would result in broader or narrower abstractions. This could be done with older infants – whose perceptual processes and memory are more developed – by changing the context in which the pair appears, such as the motion path that the objects take or the background against which they move.

Taken together, our findings suggest that perceptual and memory development may interact with relational learning to increase young infants' ability to generalize from one exemplar. This result showcases humans' remarkable relational ability, even in infancy. Ultimately, at three months, infants were able to generalize from presentations of only one example. This stands in bold contrast to the previous literature arguing that comparison is necessary to overcome the allure of salient object features and abstract a relation. In young infants, recognition failures may offer a temporary solution, as object features are forgotten between presentations. It may even be that this source of overgeneralization in the first months is one of the bases of infants' incredible conceptual growth. By capturing both this achievement and pointing to its decline as infants mature, this study reveals how the development of basic processes can interact with structural alignment mechanisms to support analogical ability from the earliest days.

Chapter 2: Impact of labels on infant relational learning

Humans are prodigious learners. Not only can we learn by association and by perceptual similarity, but unlike most species, we can learn by relational comparison. That is, humans can detect and transfer relational similarities between situations or ideas, despite perceptual differences. This relational ability has been argued to be the root of higher-order cognition (Gentner, 2003; Gentner & Medina, 1998). Indeed, there are numerous links between recognizing common relations and higher-order reasoning. Identifying shared structure is critical in categorizing items together or separately, as well as in tracking the hierarchical relationships between categories (e.g., peppers and carrots are both types of vegetable, and vegetables and meat are both types of food). By generalizing relational structure between scenarios that differ in scale and appearance, we can also make astounding inferences (e.g. that an atom has an orbital structure similar to the solar system). We can even reason about events that have never occurred, by generalizing causal relationships to hypothetical and counter-factual scenarios (e.g., "If I hadn't put my cup so close to the edge, it wouldn't have fallen when I bumped the table.")

There is a dramatic difference between human's relational ability and that of other animals. Bees, pigeons and rhesus monkeys can all recognize when objects or events are identical (A matches A), yet there is no evidence that they can discriminate when the relation *same* is shared between pairs of objects (AA matches BB) (Flemming et al., 2007; Giurfa et al., 2001; Wasserman & Young, 2010). While baboons and some monkeys can learn to match *same-different* relations, they need thousands of trials to do so(Fagot & Thompson, 2011; Katz & Wright, 2006). Even chimpanzees, our nearest evolutionary neighbors, require hundreds of trials to match *same-different* relations (Premack, 1983; Robinson, 1955). In contrast, five-year-old humans can match *same* and *different* pairs in less than ten training trials (Hochmann et al., 2017). This raises the question of whether these between-species' differences in relational ability are an expanding gap driven by the affordances of human culture, or whether they are an insurmountable evolutionary gulf.

To investigate this, recent studies have tested for whether relational ability is present in infancy. Ferry et al. (2015) first demonstrated *same-different* learning in 7- and 9-month-olds, by habituating infants to pairs that instantiated either *same* or *different* relations. After a learning period that lasted less than ten minutes, infants succeeded in abstracting the relation they had habituated to and discriminating it from a novel relation. Infants also made a telling failure. Prior to seeing the habituation pairs, the infants had been shown a set of individual toys. Despite the fact that infants generalized the *same* and *different* relations to new pairs at test, they failed to discriminate these relations in test pairs that contained the individually-salient objects. Thus, infants showed two key signatures of learning that also characterize relational learning in children and adults. The first is that facilitating comparison across exemplars (here, the habituation pairs) promotes relational abstraction. The second is that rendering individual objects salient disrupts the ability to align and abstract shared relations (Gentner & Rattermann, 1991; Paik & Mix, 2006). More recently, this paradigm has been extended to 3-month-olds, and the results show that both relational ability and the signatures of relational learning are already present at this young age (Anderson et al., 2018b). This work provides compelling evidence that relational learning is continuous across the lifespan.

The infant findings support the idea that there is already a gulf between human relational ability and that of other species. However, this is not to say that infants are capable of the complex reasoning that is the pinnacle of adult relational reasoning. Adults benefit from culture, knowledge and experience of the world, from mature executive functioning and memory, and from another uniquely human talent – language. Words complement the first signature of relational learning: the importance of comparison for aligning and abstracting a shared relation. When a common word encapsulates a structure shared by multiple events, hearing it applied across these instances can invite comparison and highlight the common structures. In turn, this allows familiar words to carry abstract concepts, not tied to any specific exemplars (e.g., “cat” applies to my short-haired tabby and your long-haired Himalayan and even to a two-dimensional drawing of a cat). These aspects of language can be especially powerful as we acquire words that not only represent objects, but also the abstract relations between them (e.g., the cup *on* the table). In fact, Gentner (2003) argues that we reach the full extent of our relational reasoning when “our analogical prowess is multiplied by the possession of relational language.”

Evidence with children supports the theory that language amplifies relational learning (Christie et al., 2016; Christie & Gentner, 2014; Loewenstein & Gentner, 2005). Preschool-age children are particularly susceptible to focusing on individual object differences at the cost of perceiving the relation (Gentner & Rattermann, 1991; Gentner & Toupin, 1986; Paik & Mix, 2006; Rattermann & Gentner, 1998; Richland et al., 2006), and giving children a label for the relation has been shown to help them overcome this object focus. For example, Rattermann and Gentner (1998) presented 3-year-olds with two sets of three objects, where the objects monotically increased in size. The challenge was that the biggest object in one set was the same size as the medium object in the other set. When 3-year-olds were asked to match the objects based on their relative size, they chose correctly only 32% of the time. However, when children were taught to label the objects as “daddy”, “mommy”, and baby” objects, they made successful relational matches 89% of the time. The benefits of a common relational label have been found for other familiar words and sets of words, such as “same”, “tiny/little/big” and “top/middle/bottom” (Casasola, 2005a; Christie & Gentner, 2014; Loewenstein & Gentner, 2005; Son et al., 2012). Common label effects have also been found when a novel word is used, such as “dax” or “blicket” (Christie & Gentner, 2010; Gentner & Namy, 1999; Namy & Gentner, 2002). The findings with novel words strengthen the idea that a major part of how a label improves relational learning is by inviting the learner to compare. In turn, if increasing comparison is the main way language improves relational learning, then it raises the question of whether learners could benefit from hearing relational labels even before they have acquired these words. In this paper, we explore whether language plays a role in infant relational learning in the first year.

Finding that even a novel label can increase comparison is a reason to suspect that common labels might also benefit relational learning even in infancy. In fact, common label effects have been found for infants as young as 3 months, but only for object categories. As with relations, hearing a common label applied to exemplars of an object category can highlight their shared features and can lead to the formation of categories (Althaus & Westermann, 2016; Balaban & Waxman, 1997; Ferry et al., 2010; Waxman & Booth, 2003). Additionally, this effect holds whether the labels are words in the learner’s native language like “rabbit” or “pig” (Balaban & Waxman, 1997), or whether they are novel words like “geepee” or “boota” (Althaus & Westermann, 2016; Waxman & Braun, 2005). This could suggest that that common labels could also improve infants’ relational learning. However, it is worth noting the gap in participants’ ages between the object category literature and the relational language literature. At

18 months, even the youngest participants to succeed with a common relational label ("on"; Casasola, 2005) are far older than the 3-month-olds who can form object categories with a common label. Furthermore, object categories are often more perceptually apparent than relations, because they can be based on clusters of shape and color features. Thus is it an open question as to whether a common label could improve relational learning in the first year.

Summary

While relational ability is present in infants before they begin to produce language, it is an open question as to when these two mechanisms begin to interact. As discussed above, a common label may be an invitation to compare (Gentner & Namy, 1999), raising the possibility that there is no lower age limit on language benefits for relational learning. However, while the research on object labels extends into infancy, the literature investigating the benefits of relational labels focus on preschoolers. The following experiments bridge this gap by examining the impact of relational labels on learning in the first year. We test infants at 12 months of age, around the time that they are beginning to acquire their first abstract words (Bergelson & Swingley, 2013). This is the first study to test the effects of relational labels in the first year of life, and it has implications for understanding how language shapes learners' attention during the early phases of language acquisition.

To test for an impact of language on relational learning, we use a habituation / dishabituation paradigm based on Ferry et al., (2015). Recall that, even though infants in this study could discriminate *same-different* relations with new pairs, they failed to do so with objects that had been made salient before the experiment by being shown individually. In our current experiment, we use this *Salient Object* trial type to test the impact of common relational labels, by manipulating the context in which infants see these objects. Experiment 1 shows these items in pairs with a common label (*same* or *different*) applied across exemplars. If language and relational learning interact in pre-verbal infants as they do in older children, then giving these pairs a common relational label should allow 12-month-olds to detect the relation at test.

A final consideration is that infants might show differences depending on whether they hear *same* or *different* labeled. Some researchers have theorized that infants conceptualize *same* before they conceptualize *different* (Addyman & Mareschal, 2010; Hochmann et al., 2016), and MCDI data supports this possibility, in that *same* is produced earlier (Frank et al., 2016). Counter to this, none of the previous relational learning studies in our lab have differences in generalization based on whether infants habituate to *same* or to *different* (Anderson et al., 2018; Ferry et al., 2015). To address this, we test for an interaction with habituation condition, where 12-month-old infants might be more likely to learn the label for *same* than for *different*.

Experiment 1

Experiment 1 examined the influence of a common relational label on 12-month-old infants. Infants were habituated to either the *same* or *different* relation. Prior to the study, the infants saw three pairs of objects that all shared a common relation and were labeled accordingly. For example, infants in the *same* condition would see three *same* pairs that were each labeled "These are *same*" as they were presented. In the study room, infants were then habituated to another four pairs with the target relation (e.g., *same* if they had heard the label *same*), which were presented silently over the course of 6 to 9 habituation trials. Finally, infants saw three types of test trials: *New* trials with previously unseen objects, *Habituation* trials with objects that appeared in the habituation phase, and *Relational Label* trials with the objects that had been given a relational label in the beginning. Each of these test trial types were composed of two back-to-back trials: one featuring the familiar relation (e.g., *same* if they had heard the label *same*) and one featuring the novel relation (e.g., *different* if they had heard *same*).

If a common relational label facilitates comparison and supports relational learning, then infants should be able to generalize the relation to never-before-seen objects and look longer at the novel relation in the *New* trials. The critical test for an interaction with language is the *Relational Label* trial type. If common labels help infants detect the relation, then infants should look longer at the novel relation compared to the familiar (labeled) one. The third test trial is *Habituation*, where infants only need to recognize an exact pair from habituation. Thus infants should also look longer at a novel pair made of habituation objects in *Habituation* test trials.

Additionally, as mentioned earlier, there may be an interaction with habituation condition, such that *same* is easier to learn than *different* (Hochmann et al., 2018). If this is the case, infants in the *same* condition should discriminate with the *Relational Label* trials, while infants in the *different* condition may not. An advantage for same could also appear with *New* trials. Infants should be able to discriminate between novel and familiar *Habituation* pairs no matter their habituation condition, because these trials only requires learning the exact object pair.

Methods

Participants

The participants were 64 12-month-old infants, ranging from 10 months, 29 days to 13 months 3 days (33 female, (M_{Age} = 12 months, 0 days). Half were assigned to the *same* condition and half to the *different* condition. Seven additional infants were excluded: six for fussiness and one for a bowel movement during test (see Coding section for how fussiness was defined).

Parents of infants were recruited through online ads and word of mouth. Parents who agreed to their infants' participation were provided informed consent before the experiment and given $20 as compensation. The race, ethnicity and parental education level of the sample is described in the table below.

Table 1.

Participant demographic information

Ethnicity	%	Race	%	Education	Maternal %	Paternal %
Hispanic	10	African American	8	Some high school	0	0
Non-Hispanic	83	Asian/ Hawaiian/ Pacific Islander	1	High school diploma	0	0
		American Indian	0	Some college	5	10
		White	75	College degree	91	84
		Multiracial	10			
No response	7	No response	6	No response	4	6

Apparatus

Parents sat in a chair with infants on their lap facing a wooden puppet stage that displayed all stimuli. The parents were asked to refrain from interacting with the infant during the experiment and to close their eyes during the test trials. The stage measured 243.5 cm high, 128 cm wide, and 61 cm deep. The opening in the front of the stage that displayed the objects was 93 cm above the floor, 61 cm high, and 106 cm wide. The back wall had two rectangular openings with cloth fringe over the openings that allowed the experimenter to manipulate the objects. A screen that covered the infants' view of the stage was raised and lowered between trials. The MATLAB program Baby Looking Time (BLT), was used to record looking times for habituation and test trials during the experiment (Chang et al., 2018).

The stimuli consisted of 18 three-dimensional objects (see Figure 1). During the experiment, each pair of objects was placed on the puppet stage on a 26.5 x 15.5 cm cardboard tray that was covered with contact paper. Six of the objects were presented in the waiting room beforehand in the context of *same* or *different* pairs. Four additional *same* or *different* pairs were shown in the habituation phase. Finally, 12 objects appeared in pairs during the test phase. Of these, four objects had been labeled in pairs in the waiting room, four objects had been shown in pairs in habituation, and four objects had not been seen before test.

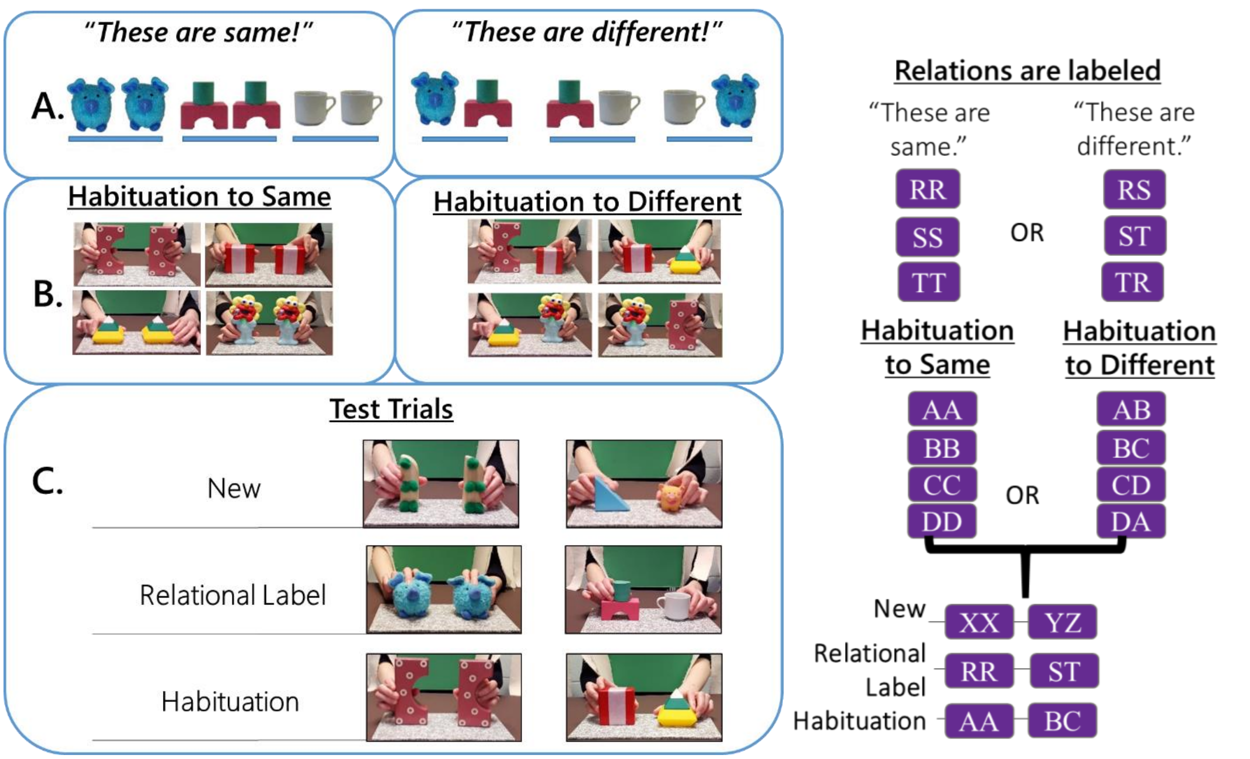

Figure 1. Pictures of the stimuli used in Experiment 1 and a schematic of the procedure. (A) Infants were presented with three pairs before the experiment, labeled appropriately as same or different. (B) Infants were habituated to four same or different pairs. (C) Infants saw three types of test trials: *New*, *Relational Label*, and *Habituation*. The order of trials was counterbalanced across participants.

Procedure

The experiment consisted of three sequential parts:

(D) Relations are labeled. The experimented showed three pairs of objects to the infant in the waiting room prior to the experiments, while labeling the relations they instantiated. These were either three pairs of two identical objects (for the *same* condition) or three pairs of two non-identical objects (for the *different* condition). During each pair presentation, the experimenter held the object in front of the infant for 5 seconds and said to the infant, "These are *same*!" or "These are *different*!" The six objects used were the same for both conditions: two blue aardvarks, two white espresso cups and two pink and green foam towers (see Figure 1-A for configurations).

(E) Pair habituation trials. When the screen was raised at the start of every trial, a pair of objects rested on the cardboard tray on the stage. To engage infants' attention, the experimenter held and moved the objects throughout the trial. The experimenter grasped one object in each hand and raised the objects straight up (1 s), tilted them to the left (1 s), returned them to the center (1 s), tilted them to the right (1 s), returned them to the center (1s), returned them to the tray (1 s), and paused on the tray (2 s). This 8-s cycle repeated continuously until the trial ended. In the *same* condition, infants saw habituation trials in which the pairs of objects were the *same* (see Figure 1-B). In the *different* condition, infants saw habituation trials in which the pairs of objects were *different*. The number of habituation trials was infant-controlled (see procedure section for the criterion), and ranged from 6 to 9 trials.

(F) Test trials. For each test trial type, infants were presented with a novel relation, followed by a familiar relation (or vice versa), summing to six test trials (see Figure 1-C). In each test trial, infants viewed one pair of objects, presented in the same motion pattern as in the habituation trials, while their looking time was recorded. The three kinds of test trials were (a) objects that had not been seen before test (*New*); (b) objects that the infant had heard labeled in the waiting room (*Relational Label*); and (c) objects that the infant had seen presented in the pair habituation trials (*Habituation*). There were three trial orders (abc, cab, bca) which were counterbalanced across infants.

Coding

There was a small hole in the front face of the stage containing a camera that captured a video image of the infant's face. While the experimenter conducted habituation and test trials in the room with the infants, two research assistants in a separate room viewed a video of the infants' face and coded the infants' visual fixations online as either on- or off-target. Each researcher depressed a computer button when the infant attended to the events on stage and released the button when the infant looked away. Each trial ended when the infant either looked away for 2 consecutive seconds after having looked at the event for at least 2 s or looked at the event for 60 cumulative seconds without looking away for 2 consecutive seconds. The BLT program indicated the end of the trial with a beep, signaling to the experimenter to lower the screen and move to the next trial. After each test trial, research assistants also checked one or more boxes to indicate the behavioral state of the infant on the preceding trial: sleepy, quiet and alert, active, fussy or crying. Coders also noted any breaks and their length. As noted above, if two coders independently judged the infant's state as fussy, crying, or falling asleep for more than half the test trials, the infant's data was excluded from the analysis. The coders were blind to the condition and the trial order. Interobserver agreement was measured for all infants and averaged 91%. The Intraclass Correlation Coefficient for a fixed set of raters (ICC3) was .91

with a 95% confidence interval from .88 to .93, $F(40, 233) = 11$, $p < .001$. Our looking time data were skewed towards the lower bound and significantly deviated from a normal distribution per the Shapiro-Wilks test. Therefore, we performed parametric tests on log-transformed data, following recommendations outlined by Csibra, Hernik, Mascaro, Tatone, and Lengyel (2016).

Results

The key question was whether relational labels would facilitate learning in 12-month-olds. The answer is no: there was no clear evidence of learning for any of the test trial types. We first conducted a repeated-measures ANOVA on the *New* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). Averaging across groups, infants looked 21.10 s (SD = 16.16) at novel relations and 16.95 s (SD = 11.85) at familiar ones, and there was no main effect of relation, $F(1, 56) = 2.81$, $p = .099$. A binomial comparison also reflects the lack of discrimination: only 37 infants out of 64 looked longer at the novel relation in the *New* trials, $p = .130$. However, there was an interaction between relation and habituation condition: infants in the *same* condition looked longer at the novel relation than the familiar one (M_{Novel} = 24.13 s; $M_{Familiar}$ = 15.49 s), while infants in the *different* condition looked equally between them (M_{Novel} = 17.89 s; $M_{Familiar}$ = 18.51 s), $F(1, 56) = 7.61$, $p = .008$, $\eta_p^2 = .12$ (see Figure 2). A binomial comparison also suggested that infants' relational generalization fared better in the *same* condition: 22 of the 33 in the *same* condition looked longer at the novel relation than the familiar one, $p = .04$, while only 16 of 31 infants in the *different* condition did so, $p = .64$. There was also an interaction between relation and test order, such that infants who saw novel relations first looked during test longer at the novel relation than the familiar one (M_{Novel} = 21.24 s; $M_{Familiar}$ = 14.49 s), while infants who saw the familiar relation first during test looked more equally between them one (M_{Novel} = 20.97 s; $M_{Familiar}$ = 19.42 s), $F(1, 56) = 4.95$, $p = .030$, $\eta_p^2 = .08$. Finally, there was a three-way interaction between relation, condition and test order. The infants who saw the novel relation first in the *same* condition showed greater looking differences between novel and familiar relation than did infants in any of the other cells, $F(1, 56) = 5.12$, $p = .028$, $\eta_p^2 = .08$. There was no main effect of condition on overall looking, $F(1, 56) < 1$, $p = .63$, nor a main effect of order, $F(1, 56) < 1$, $p = .71$. There were no other significant main effects or interactions.

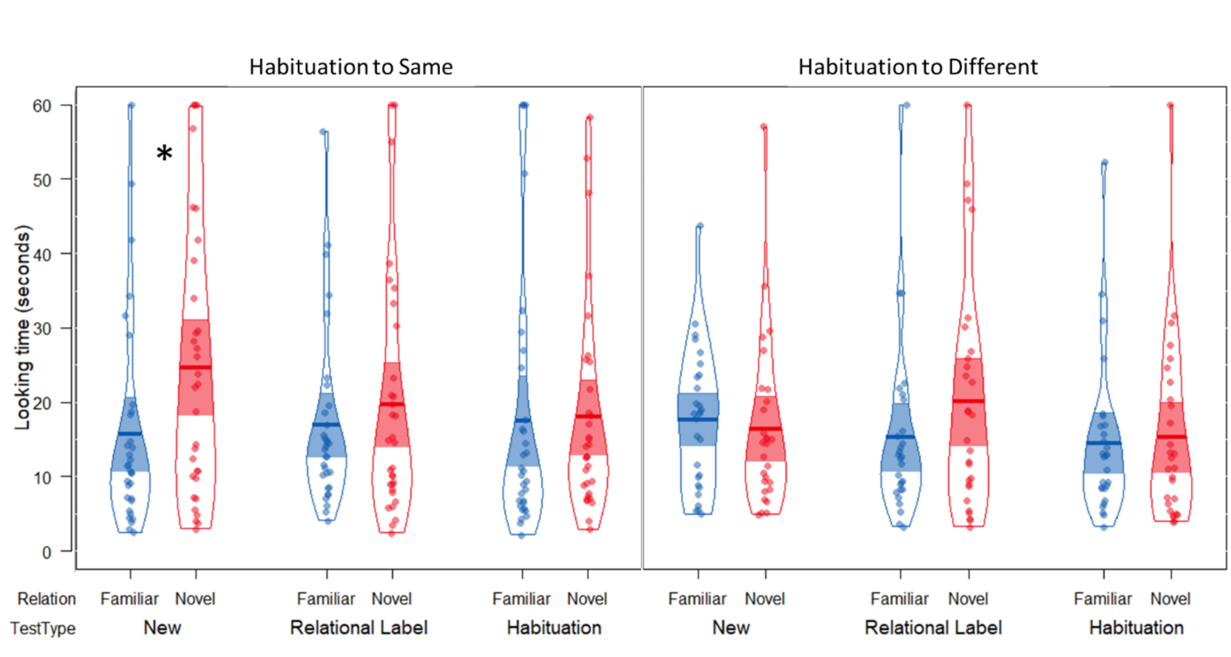

Figure 2. Looking times for 12-month-olds who heard relational labels, separated by habituation condition (*same* vs. *different*) and by test trial type (*Habituation*, *New*, and *Relational Label*). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value. * indicates p < .05 and ** indicates p < .001.

Next, we conducted a repeated-measures ANOVA on the *Relational Label* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). Averaging across groups, infants looked 20.32 s (SD = 15.31) at novel relations and 17.02 s (SD = 12.33) at familiar ones, and there was no main effect of relation, $F(1, 56) < 1$, $p = .427$. Again, the binomial comparison reflects this pattern: only 34 infants out of 64 looked longer at the novel relation in the *New* trials, $p = .354$. There was no interaction with habituation condition, $F(1, 56) < 1$, $p = .376$. A binomial comparison also suggested that infants performed similarly in both conditions: 18 of the 33 in the *same* condition looked longer at the novel relation than the familiar one, $p = .364$, and only 16 of 31 infants in the *different* condition did so, $p = .50$. There were no other significant main effects or interactions.

Finally, we conducted a third repeated-measures ANOVA on the *Habituation* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). Infants looked near equally at novel (M = 16.60 s, SD = 13.14) and familiar relations (M =16.05 s, SD = 14.34), and there was no main effect of relation, $F(1, 52) < 1$, $p = .427$. Again, the binomial comparison reflects this: only 33 of 61 infants looked longer at the novel relation in the *Habituation* trials, $p = .304$. There was also no interaction with habituation condition, $F(1, 52) < 1$, $p = .376$. A binomial comparison also suggested that infants' performed similarly in both conditions: 21 of the 33 in the *same* condition looked longer at the novel relation than the familiar one, $p = .081$, and only 12 of 28 infants in the *different* condition did so, p = .828. There were no other significant main effects or interactions.[2]

Comparison to Ferry et al. (2015)

Did a relational label have a positive impact on the *Relational Label* trials, compared to the unlabeled *Salient Object* trials in Ferry et al., (2015)? The answer is no. We conducted a repeated measures ANOVA on the *Relational Label / Salient Object* trials, examining the within subjects factor of Relation and the between-subjects factors of Experiment (Ferry et al. 2015 vs. current study) and Condition (habituation to *same* vs. *different*). This ANOVA showed no main effect of Relation, $F(1, 124) < 1$, $p = .820$. The lack of generalization did not interact with Experiment. That is, infants in Ferry et al., (2015) looked near equally between relations (M_{Novel} = 17.74 s; M_{Fam} = 17.23 s) as well as with labeled pairs in the current experiment (M_{Novel} = 19.90 s; M_{Fam} = 16.61), $F(1, 124) = 1.69$, $p = .197$. There was no interaction between Relation and Condition, $F(1, 124) = 1.13$, $p = .290$, nor was there a three-way interaction between these factors and Experiment, $F(1, 124) < 1$, $p = .739$. There were no main effects.

We also conducted a repeated measures ANOVA on the *New* trials, using the within subjects factor of Relation and the between-subjects factors of Experiment and Condition, which did show a main effect of Relation, $F(1, 124) = 10.84$, $p = .001$, $\eta_p^2 = .08$. This effect of Relation did not depend on Experiment: even though infants looked significantly longer at the novel relation in Ferry et al., 2015 (M_{Novel} = 18.91 s; M_{Fam} = 13.09 s) and not in this experiment (M_{Novel} = 21.10 s; M_{Fam} = 16.95), this was no significant interaction, $F(1, 124) = 1.16$, $p = .284$. The effect of Relation did not interact with Condition, either, $F(1, 124) = 2.53$, $p = .114$. However,

[2] A pilot experiment implemented this procedure with 7- and 9-month-olds (n = 53) and found null results for all test trials: *New*, $t(52) = 3.472$, $p = .068$; *Relational Label*, $t(52) < 1$, $p = .987$; and *Habituation*, $t(52) < 1$, p = .368.

there was a three-way Relation x Experiment x Condition interaction, in that infants in Ferry et al., (2015) generalized with both relations, while infants in the current experiment only generalized same $F(1, 124) = 5.45$, $p = .021$, $\eta_p^2 = .04$. There were no other main effects. We could not do a comparison to the *Habituation* trial in the current study because Ferry et al., (2015) had a different third test trial type[3].

Habituation patterns

Finally, we examined looking time during the habituation phase. Some theories that argue that *same* should be easier learn that *different* (Addyman & Mareschal, 2010; Hochmann et al., 2018). In the context of this experiment, this would suggest that the habituation decline between conditions should be commensurate to the ease of learning *same* compared to *different*. A repeated measures ANOVA examined looking time by the within-subjects factor of Habituation Trial (first three through last three) and the between-subjects factor of Condition (*same* and *different*), and revealed a significant decline in looking across conditions, $F(5, 310) = 57.30$, $p < .001$, $\eta_p^2 = .43$ (see Figure 3). On average, infants looked 46.96 s on their first habituation trial and 15.66 s on their last trial. There was not a significant interaction between Habituation Trial and Condition, $F(5, 310) = 1.042$, $p < .382$, nor a main effect of Condition, $F(1, 62) < 1$, $p < .536$. Additionally, there were no between-condition differences in terms of total looking time during habituation, $F(1, 62) < 1$, $p = .743$, in how many trials it took to reach the habituation

[3] Infants in Ferry et al. (2015) saw objects presented in both the waiting room and then in habituation pairs before they finally reappeared in test trials, while infants in Experiment 1 only saw the objects in habituation before they reappeared at test.

criterion, $F(1, 62) < 1$, $p = .811$, or on whether infants reached the habituation criterion, X^2 (1, $N = 63$) = .522, $p = .470$.

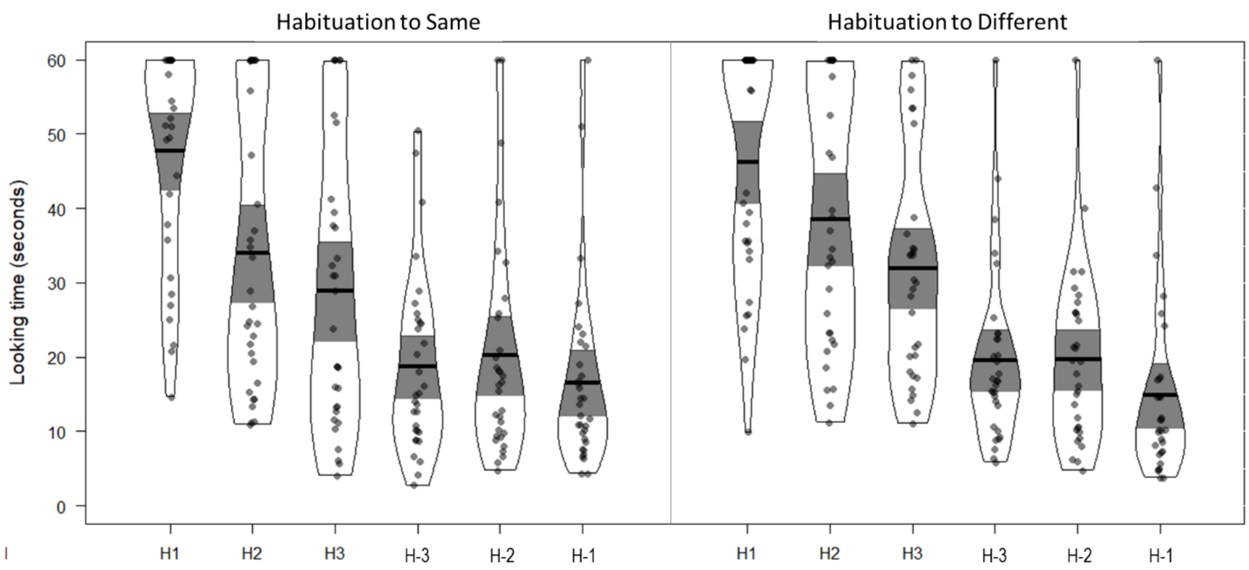

Figure 3. Looking decline over the course of 6 to 9 habituation trials, separated by habituation condition. Because the number of habituation trials was infant-controlled, the graphs represent the first three (H1, H2, H3) and the last three trials (H-3, H-2, H-1). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Discussion

The results show catastrophic failure for 12-month-olds learning *same* or *different* from a relational label. Not only did the labels not help infants to discriminating between relations on the *Relational Label* pairs, but across conditions, infants failed to recognize the novel and familiar relations during *New* and *Habituation* test trials as well. Infants' failure to discriminate the *Habituation* test trials is especially concerning, because differentiating an exact pair seen in habituation (e.g. AA) from a novel pair made of habituation objects (e.g., BC from BB and CC) should not require any relational learning, only remembering what was presented in the Pair Habituation phase. One possibility is that these null results reflect 12-month-olds' inability to learn *same-different* relations, regardless of whether a label is involved. However, this explanation is surprising in the face of younger infants' success (Anderson et al., 2018b; Ferry et al., 2015). Instead, these results suggest that the common relational label prevented infants' ability to detect the relation on any trials.

Infants who habituated to *same* fared slightly better, generalizing this relation to *New* test trials. We had considered the possibility that a *same* label might be easier to learn than a *different* label, found in other infant studies (Addyman & Mareschal, 2010; Hochmann et al., 2018) and on data of children's productive vocabulary showing that *same* is acquired earlier (Frank et al., 2016). However, if the relative ease of the *same* label is what allowed infants to generalize in this condition, then infants should also have discriminated between relations in the *Relational Label* trials and the *Habituation* test trials. Instead, infants in the *same* condition also looked equally between relations on both of these trial types. This lack of discrimination suggests that infants in the *same* condition generalized to *New* despite the detrimental effect of the relational label, rather than because its benefits.

Previously, we asked if common labels might improve relational learning even in infancy through an implicit invitation to compare. The dismal failures of Experiment 1 argue that this was not the case. What could have had such a severe impact on infants' ability to learn from these relational labels? One answer is that infants' early lexicon is so devoted to concrete objects, that infants may not grasp a more abstract label, such as *same* or *different*. Highly concrete referents overwhelmingly dominate early vocabularies (Bornstein et al., 2004; Frank et al., 2016; Gentner, 1982; Gentner & Boroditsky, 2001), putting English speakers in the midst of a “nominal explosion” (Hoyos et al., 2016). For example, the first words that infants typically produce are names for specific individuals (e.g., “mama” and “dada”), as well as words for food, body parts, clothing, animals, and household objects (Clark, 1979). In contrast, more abstract relational terms are acquired significantly later (e.g., only 2% of English-speaking 18-month-olds produce the word “same”) (Frank et al., 2016; Gentner, 1982). The pattern also appears in the comprehension timelines: infants can understand some concrete object words like “bottle” and “apple” at 9 months but do not understand more abstract terms like “all gone” until 12 months (Bergelson & Swingley, 2012, 2013). With early language experience that is based almost entirely around concrete objects, infants might not recognize that labels can have abstract referents, like relations. Thus, even if a common label increased the chance of comparison in infancy, early language experience could still limit infants’ ability to map the label onto such an abstract referent as a relation.

To understand whether infants are able to learn relations in the context of more concrete labels, we turn from relational labels to object labels. Unlike relations and most other abstract terms, there is ample evidence that object words are a part of infants’ early vocabulary (Bergelson & Swingley, 2012). There is also evidence that drawing attention to objects interferes with relational learning (Anderson et al., 2018b; Ferry et al., 2015). In particular, highlighting the differences in object features between relational exemplars interferes with aligning the exemplars and detecting their relational commonalities. In Experiment 1, we had hypothesized that a common relational label might aid relational abstraction by emphasizing the shared structure. In Experiment 2, we ask whether individual objects labels should hinder learning by emphasizing the distinct objects in the pairs and thus making alignment across them more difficult. In contrast to common labels, which lead to groupings and an emphasis on shared features, giving exemplars individual labels makes differences more salient. By 9 months, infants are more likely to distinguish between exemplars of a perceptual category when each is given an individual label than when a common label is applied across them (Best et al., 2010; LaTourrette & Waxman, submitted; Pickron et al., 2018). In Experiment 1, we found that common labels could not yet help relational learning at 12 months. In Experiment 2, we ask if, in contrast, individual object labels may already be able to disrupt relational learning in predictable way.

In the context of our paradigm, infants in Experiment 2 will again see objects and hear labels, but unlike in Experiment 1, these will be individual object labels. Here, we predict that object labels should negatively impact relational learning, but that this effect might be limited to the labeled items. This is based on the fact that, even though infants in Ferry et al. (2015) and in Anderson et al. (2018) did not discriminate between pair of objects that were individually salient, they could still generalize to *New* trials.

Experiment 2

Experiment 2 examined the influence of individual object labels on 12-month-old infants' relational learning. Prior to the study, the infants saw three pairs of objects. In contrast to Experiment 1, these objects were shown individually instead of in pairs, and each object received its own individual label (see Figure 4). The labels were English words whose content matched the items presented (e.g., *cup, tower*). The sequence of objects and their labels was presented twice. Next, in the study room, infants were habituated to four pairs that shared a common relation (e.g., all *same* or all *different*). These were presented silently over the course of 6 to 9 habituation trials. Finally, infants saw four types of test trials: *New* trials with previously unseen objects, *Object Label* trials with the objects that had been given a label in the beginning, *Habituation* trials with objects that appeared in the habituation phase, and a *Memory Check* trial. The *Memory Check* trial type tests infants on a familiar pair from habituation compared to a pair that is novel in terms of both the objects and the relation. The *Memory Check* trial is similar to the *Habituation* trial type in that it measures recognition rather than generalization. The difference is that, in *Habituation* trials, infants have seen all of the objects but must recognize the relation they were in (i.e., *same* or *different*), whereas in the *Memory Check*, infants can succeed by recognizing that either the relation or the objects are novel. The *Memory Check* trial type was added as an easier version of the *Habituation* trial in response to the widespread failures to discriminate relations in Experiment 1. Each of the four test trial types included two sequential trials: one featuring the familiar relation (e.g., *same* if they had learned *same*) and one featuring the novel relation (e.g., *different* if they had learned *same*).

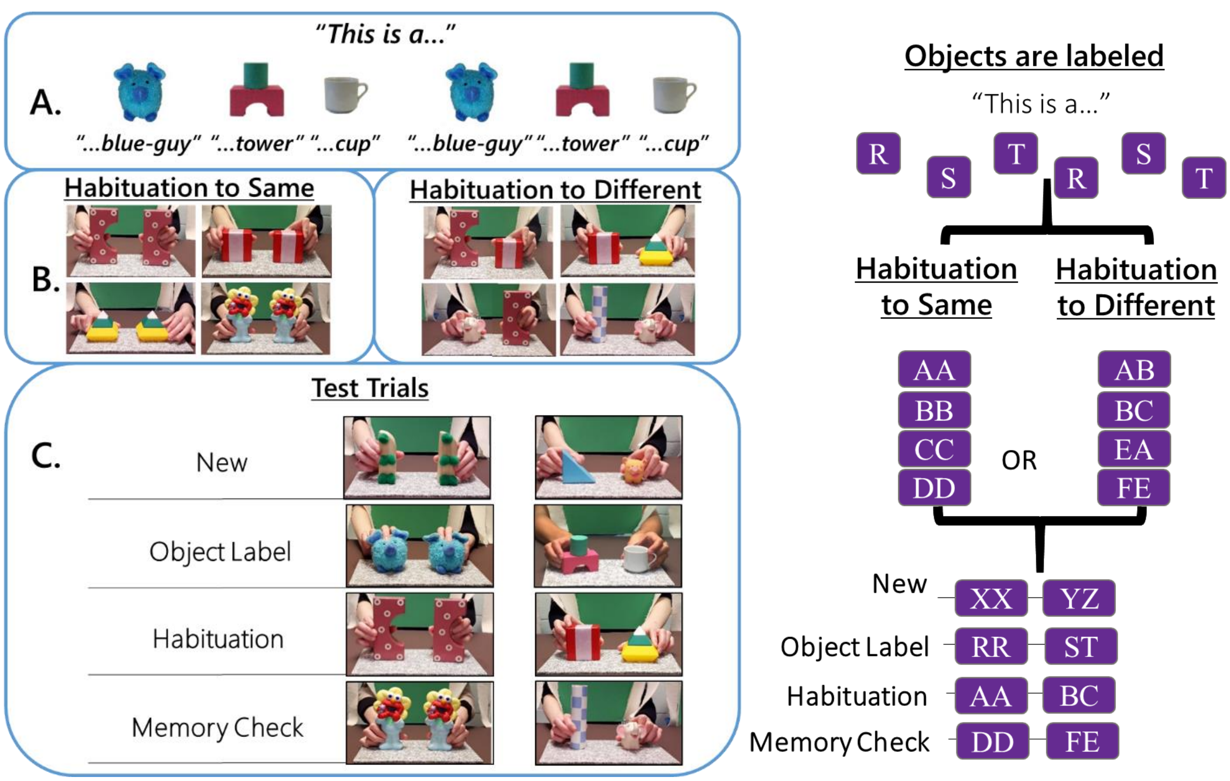

Figure 4. (A) Infants heard individual objects labeled before the experiment. (B) Infants habituated to four *same* or *different* pairs. (C) Infants saw three types of test trials: *New*, *Object*

Label, *Habituation*, and *Memory Check*. The order of trials was counterbalanced across participants.

If labeling individual objects disrupts relational learning, for only the labeled objects, then object labels should not impact generalization to never-before-seen objects. Infants should still look longer at the novel relation in the *New* trials. The critical test for an interaction with these individual object labels is the *Object Label* trial type. Here, infants should look equally at novel relation and familiar relations. The third and fourth test trials are *Habituation* and the *Memory Check*, where infants only need to remember the exact pairs that they saw in the habituation phase. Infants should discriminate the relation in both *Habituation* and *Memory Check* trials, regardless of whether relational learning occurs.

Methods

Participants

The target sample is 48 twelve-month-old infants. The current sample includes 45 infants, ranging from 11 months, 13 days to 12 months 20 days (17 female, M_{Age} = 12 months, 2 days). Half of the infants were assigned to the *same* condition, and half to the *different* condition. An additional three infants were excluded because of fussiness for over half of the test trials. Exclusion criteria was the same as in Experiment 1, as were recruitment and compensation. The demographic details of the sample are listed below.

Table 2.

Participant demographic information

Ethnicity	%	Race	%	Education	Maternal %	Paternal %
Hispanic	10	African American	12	Some high school	0	0

Non-Hispanic	83	Asian/ Hawaiian/ Pacific Islander	6	High school diploma	0	2
		American Indian	0	Some college	10	12
		White	65	College degree	90	87
		Multiracial	17			
No response	0	No response	0	No response	0	0

Apparatus

The stage was the same as in Experiment 1.The stimuli were 32 three-dimensional objects. The objects were the same as those in Experiment 1, except that six of the objects were presented in the waiting room beforehand, four of which reappeared at test.

Design & Procedure

The design and procedure were modeled on Experiment 1 with the following differences. The object experience phase differed in that the objects were labeled individually. Specifically, the experimenter held up each individual object and said to the infant, “This is a (cup / tower / blueguy)!” using the appropriate label. The test trials were the same as in Experiment 1, except the *Object Label* pairs at test were composed of objects that have been labeled individually. Additionally, a fourth test trial type was added: a *Memory Check* trial, where a habituation pair (familiar objects and familiar relation) was compared to a novel relation made of new objects.

Coding

Coding was the same as in Experiment 1. The coders were blind to the condition and the trial order. Interobserver agreement was measured for all infants and averaged 93%. The Intraclass Correlation Coefficient for a fixed set of raters (ICC3k) was .99 with a 95% confidence interval from .98 to .99, $F(47, 383) = 103$, $p < .001$. Our looking time data skewed towards the lower bound and significantly deviated from a normal distribution per the Shapiro-Wilks test, so we performed parametric tests on log-transformed data, as in Experiment 1.

Experiment 2 Results

Current n = 45. The last three infants will be tested after COVID-19 suspensions end.

The critical test for relational learning is generalization to never-before-seen objects, so we conducted a repeated-measures ANOVA on the *New* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). Averaging across groups, infants looked 22.34 s (*SD* = 16.07) at novel relations and 17.97 s (*SD* = 16.30) at familiar ones, resulting in a main effect of relation, $F(1, 37) = 5.04$, $p = .031$, $\eta_p^2 = .12$ (see Figure 5). A binomial comparison reflected the effect of relation, with 29 of the 45 infants looking longer at the novel relation on this trial type, $p = .036$. There were no other significant main effects nor any significant interactions.

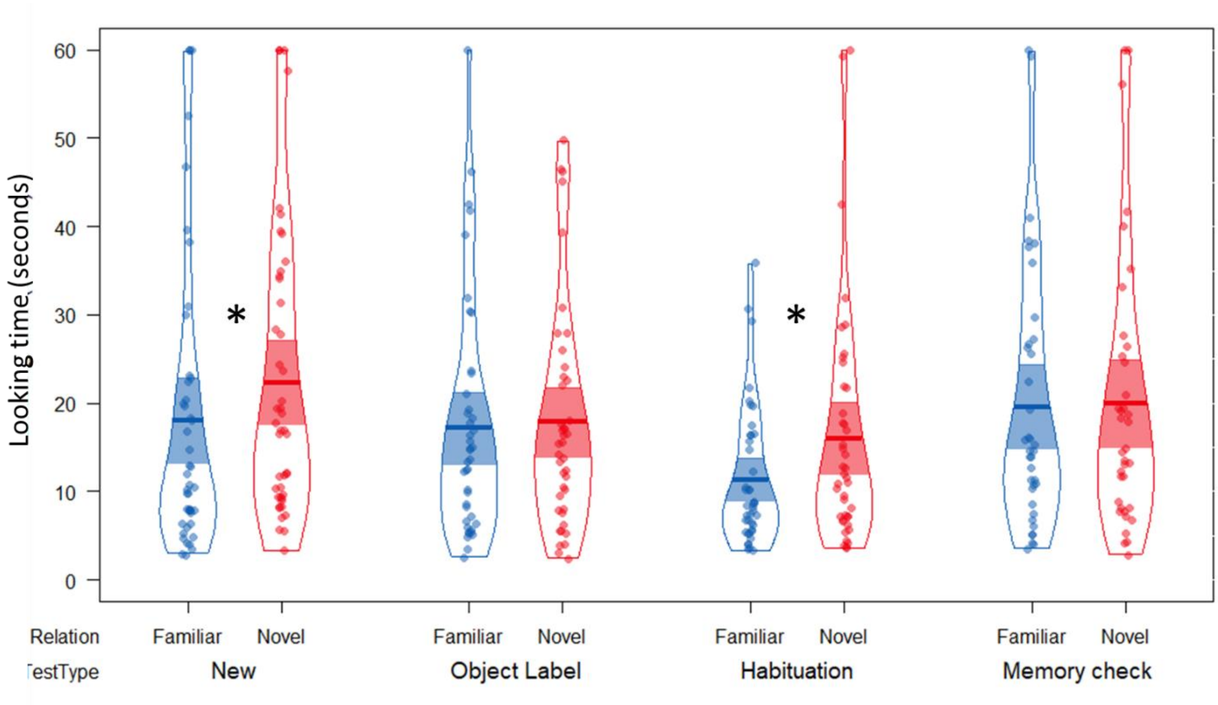

Figure 5. Experiment 2 looking times for 12-month-olds who heard individual object labels, separated by test trial type (*New*, *Object Label*, *Habituation* and *Memory Check*). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Next, we conducted a repeated-measures ANOVA on the *Object Label* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). As predicted, infants looked near equally at novel relations (M = 17.83 s; SD = 12.58) and at familiar ones (M = 17.21 s; SD = 13.20), and there was no main effect of relation, $F(1, 34) < 1$, $p = .914$. A binomial comparison reflected this lack of discrimination, with 20 of the 42 infants with usable data for this trial type looking longer at the novel relation on this trial type, $p = .678$. There was no interaction with condition indicating an advantage for *same* over *different*, $F(1, 34) = 1.99$, $p = .168$. There were no other significant main effects or interactions.

We conducted a repeated-measures ANOVA on the *Habituation* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female). Infants discriminate between relations on this test trial type, looking an average of 16.02 s ($SD = 13.30$) at novel relations and 11.35 s ($SD = 7.78$) at familiar relations, and there was a significant effect of relation, $F(1, 34) = 4.14$, $p = .05$, $\eta_p^2 = .11$. A binomial comparison reflected this effect of relation, with 29 of the 42 infants with usable data for this trial type looking longer at the novel relation on this trial type, $p = .01$. Here there was a significant interaction with condition, reflecting an advantage for *same* over for *different*: $F(1, 34) = 4.89$, $p = .034$, $\eta_p^2 = .13$. Similarly, a binomial comparison showed that 17 of 20 infants discriminated the novel and familiar pairs in the *same* condition, $p = .001$, but only 12 of 22 discriminated in the *different* condition, $p = .42$. There were no other significant main effects or interactions.[4]

Experiment 2 also included a *Memory Check* test trial. However, due to experimenter error, 8 of 23 infants in the *different* condition had their test pairs dropped for this trial type). We conducted a repeated-measures ANOVA on the *Habituation* test trial type, looking at the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Test Order (novel relation or familiar relation first), and Sex (male vs. female), but did not include

[4] A pilot experiment implemented a similar procedure with 7- and 9-month-olds ($n = 32$) and, in contrast to Experiment 2, found null results on all trials: *New*, $F(1,29) < 1$, $p = .331$; *Object Label*, $F(1,30) = 2.58$, $p = .119$, and *Habituation*, $F(1,29) = 3.961$, $p = .056$. As in the current experiment, there was a Relation and Condition interaction for Habituation trials, where infants in *same* performed better, $F(1,29) = 4.707$, $p = .039$.

Condition (habituation to *same* vs. *different*) because of the imbalanced cell sizes between conditions. Counter to our predictions, infants looked equally between these trials, an average of 19.98 s (*SD* = 15.26) at novel relations and 19.60 s (*SD* = 14.56) at familiar relations, and there was no significant effect of relation, $F(1, 36) < 1$, $p = .901$. A binomial comparison reflected this lack of discrimination, with 21 of the 37 infants with usable data for this trial type looking longer at the novel relation on this trial type, $p = .256$. There were no other significant main effects or interactions.

Comparison to Experiment 1

The main question was whether infants would show differences between the *Relational Label* trial in Experiment 1 and the *Object Label* trial in Experiment 2, indicating that the different types of labels had distinct effects. A repeated-measures ANOVA[5] on the *Relational / Object Label* test trial type, with the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Experiment (1 vs. 2), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female), revealed no main effect of relation, $F(1, 90) < 1$, $p = .556$. There was no interaction of Relation and Experiment, $F(1, 90) = 2.53$, $p = .115$, nor any other main effects or interactions.

Next we asked whether these different types of labels impacted generalization to new objects. A repeated-measures ANOVA on the *New* test trial type, with the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Experiment (1 vs. 2), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female), revealed that there was a main effect of relation, $F(1, 93) =$

---

[5] To compensate for the unequal sample sizes in the ANOVA, JASP software uses weighted means and Type I sum of squares.

8.89, $p = .004$, $\eta_p^2 = .09$. There was no interaction of Relation and Experiment, $F(1, 85) < 1$, $p = .373$, nor an interaction of Relation and Condition, $F(1, 85) < 1$, $p = .855$. However, there was a three-way interaction of Relation by Experiment by Condition, where infants who habituated to *same* showed stronger novelty preferences in Experiment 1, but infants who habituated to *different* showed stronger novelty preferences in Experiment 2, $F(1, 93) = 8.21$, $p = .005$, $\eta_p^2 = .08$.

Finally, a repeated-measures ANOVA on the *Habituation* test trial type, with the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Experiment (1 vs. 2), Condition (habituation to *same* vs. *different*), Test Order (novel relation or familiar relation first), and Sex (male vs. female) revealed that there was a main effect of relation, $F(1, 85) = 4.70$, $p = .033$, $\eta_p^2 = .05$. There was no interaction of Relation and Experiment, $F(1, 85) = 1.07$, $p = .304$. There was an interaction between Relation and Condition, such that, collapsing across experiments, infants habituated to *same* looked longer at the novel pair ($M = 18.69$) than at the familiar pair ($M = 13.96$), but infants looked equally between pairs in the *different* condition, ($M_{Novel} = 14.25$; $M_{Familiar} = 14.21$), $F(1, 85) = 5.77$, $p = .018$, $\eta_p^2 = .06$. Finally, there was a three-way interaction of Relation, Order and Sex, where male infants showed greater discrimination between relations when the novel relation was first in the test pair, and female infants showed the opposite pattern, $F(1, 85) = 4.63$, $p = .034$, $\eta_p^2 = .05$. There were no other main effects or interactions.

Habituation patterns

Lastly, we examined looking time during the habituation phase. A repeated measures ANOVA examined looking time by the within-subjects factor of Habituation Trial (first three and last three) and the between-subjects factor of Condition (*same* and *different*). The ANOVA revealed a significant decline in looking across conditions, $F(5, 215) = 33.64$, $p < .001$, $\eta_p^2 = .44$ (see Figure 6). On average, infants looked 47.42 s on their first habituation trial and 17.70 s on their last trial. There was no interaction between Habituation Trial and Condition, $F(5, 215) < 1$, $p = .849$, nor a main effect of Condition, $F(1, 43) < 1$, $p = .345$. There were also no between-conditions differences in terms of total looking time during habituation, $F(1, 43) = 11.92$, $p = .281$, in how many trials it took reach the habituation criterion $F(1, 43) < 1$, $p = .946$, or on whether infants reached the habituation criterion or not, X^2 (1, $N = 45$) = .012, $p = .914$.

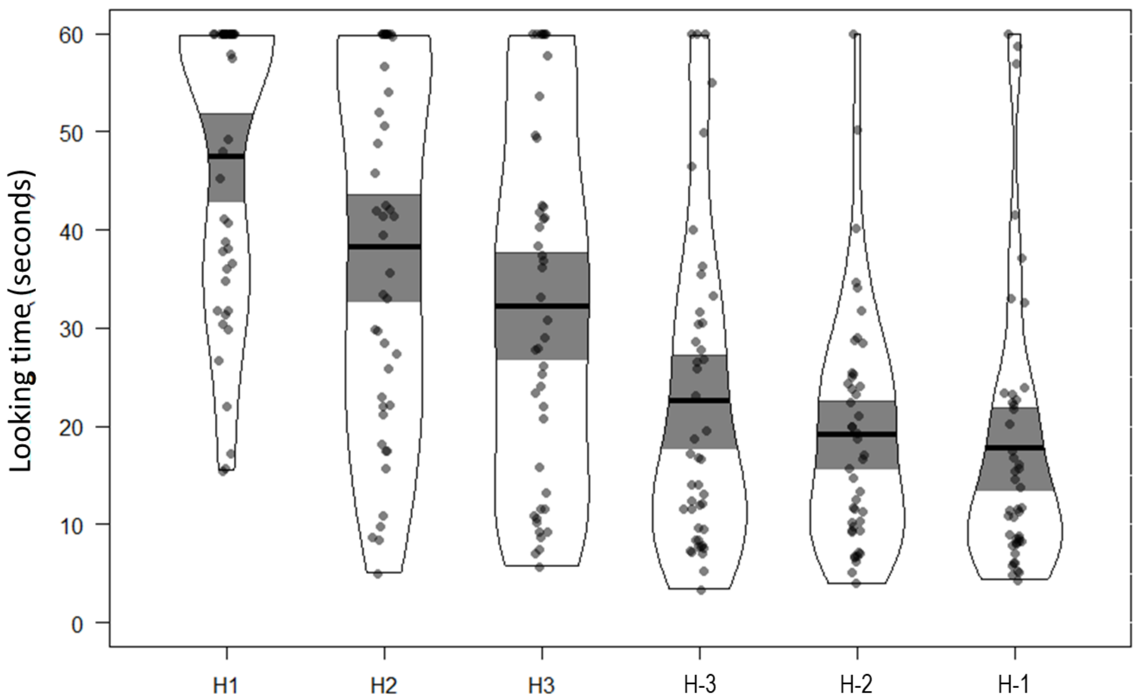

Figure 6. Looking decline over the course of 6 to 9 habituation trials. Because the number of habituation trials depended on the individual infant, the graphs represent the first three (H1, H2, H3) and the last three trials (H-3, H-2, H-1). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Discussion

Infant did not detect the relation in *Object Label* trials, but they did look longer at the novel relation for *New* and *Habituation* trials. This pattern is consistent with those seen in previous infant relational learning studies, where the effect of salient objects is limited to the trials in which they appear (Ferry, et al. 2015; Anderson et al., 2018). It also demonstrates that 12-month-olds, like their younger counterparts, are capable of learning *same-different* relations. This provides further evidence that the null results in Experiment 1 were not due to an inability to do relational learning at this age and in this paradigm. Returning to the question of whether language can impact relational learning in infancy, the results of Experiment 2 suggest that labels can have detrimental effects before it can have beneficial effects. Drawing attention to object differences can hinder the ability to align across relational exemplars, and this study and work by Bergelson and Swingley (2012) suggest that infants at 12 months already have the early lexicon to map labels to objects. Thus labeling individual objects interfered with infants' ability to learn the relation on those pairs, despite generalizing to *New* pairs.

The results also revealed two findings that ran counter to our predictions. First, there was an interaction between Relation and Condition for the *Habituation* test type, which indicated that infants performed better in the *same* condition, and did not learn in the habituation to *different* condition. Although this is the first time our lab has found an instance where *same* is easier than *different*, other studies have found such patterns (Addyman & Mareschal, 2010; Hochmann et al., 2018). However, we would not typically expect *same* to be easier on such a concrete measure of learning. A methodological reason for this interaction could be that infants in the *different* habituated to five unique objects (A, B, C, E & F), while infants in the *same* condition only habituated to four (A, B, C & D). This may have placed greater working memory demand on infants in different, and also differed from Experiment 1 and Ferry et al. (2015), where the habituation object were exactly the same in both conditions. However, this change was deliberate, so that we could include a *Memory Check* trial where the familiar pair had familiar objects and a familiar relation (e.g., DD and *same*), compared to the novel pair that had novel objects and a novel relation (e.g., EF and *different*). In the *Memory Check* trials, infants completely failed to discriminate between pairs. This findings not only ran counter to our predictions, but it is puzzling that infants would succeed on the *Habituation* trial but not the *Memory Check* when the latter was designed to be an easier version of the former. It is unclear what drove this effect, and because these two trial types have never been used within the same set of subjects before. One factor may be that the sample size was smaller in this test trial type than in the *Habituation* test type (37 vs. 42), due to experimenter error.

General Discussion

We outlined two factors that could potentially interact with relational learning: 1) common labels which might support comparison and the abstraction of a shared relation, and 2) the highly concrete nature of infants' early lexicon, which could limit infants' ability to detect the abstract relation when it is the referent of a label. The results of Experiment 1 showed that relational language interfered with relational discrimination not only for test trials that featured labeled pairs, but for the other two test types as well. In contrast, infants already had the early lexicon to map individual object labels to their referents in Experiment 2. With an early vocabulary that is already dominated by concrete object referents, infants in Experiment 2 were able to take individual labels in stride. This is apparent from the fact that they failed to discriminate between pairs made of the labeled objects but looked longer at the novel relation on *Habituation* and *New* trials. In sum, early lexicon seems to be the strongest factor in how language shapes infants' relational learning.

There are some remaining questions. Infants' relative success with object labels in Experiment 2 indicates that infants in Experiment 1 were not construing the common relational label as a basic-level object label. What was the nature of the disruption that relational labels created in Experiment 1? One possibility is that infants mapped the common label onto a superordinate object category, like "toys". Because the vast majority of studies comparing individual and common labels have looked at solely at object categories (and always with individual objects instead of pairs), it is hard to know whether relational labels would need additional support even in older children or adults. As an example of this, Christie and Gentner (2010) presented preschool age children with two examples that shared a novel label. In one experiment, the two examples were presented simultaneously, and in addition to hearing each example labeled, children were also asked, "Can you see why both of these are jiggies?" In another experiment, the examples were presented sequentially, but still given the same labels. Though both groups heard a common label for these relations, it was only the children in the simultaneous comparison condition who could identify the relational match. Other studies have also noted that although relational terms are common in English (Asmuth & Gentner, 2005), there is a protracted learning period for them, during which time children mistake these terms for object categories and make more concrete mappings than necessary (Gentner, 2005; Gentner & Rattermann, 1991; Keil & Batterman, 1984). In light of this long learning period, it is perhaps not so surprising that 12-month-olds were not able to map a common relational label to its referents.

The results show that the salient object signature of relational learning is supported by individual labels before the comparison signature is supported by a common label. This work shows that comprehension of object labels is already in place, in line with research by Bergelson and Swingley (2012) and studies using novel labels to incite category formation. By 12 months, infants are beginning to acquire some abstract terms like “all gone”, yet this age group did not benefit from relational labels. It is not until four years that children have shown benefits from hearing *same* and *different* labeled (Christie & Gentner, 2014). Future studies should investigate this three-year gap, by examining the dual timelines of abstract vocabulary acquisition and benefits of common labels on relational learning. Doing so can help us determine where the tipping point is for abstract language experience, and what other factors might allow young children to begin mapping labels onto abstract relations.

This work again makes clear that language and relational learning are two separate processes, where the results of Experiment 2 are consistent with previous studies finding relational learning before vocabulary acquisition (Anderson et al., 2018). Moreover, the interactions between these two processes are slow to develop and are strongly shaped by the highly concrete nature of infants’ early lexicon. Because of this, in sharp contrast to theories that relational learning depends on language, the results of these experiments show that the detrimental effects of object labels are present far before the beneficial effects of a common relational label (with the earliest evidence of benefits at 18 months; Casasola, 2005). Ultimately, the results of our studies show that language’s tendency to amplify the signatures of relational learning – for better and for worse – is a protracted process.

Chapter 3: Broadening initial abstractions

Humans have an incredible ability to acquire new information. Not only we can generalize based on perceptual similarity, but on relational similarity as well. That is, unlike many other species, humans can recognize when relational structures are shared between items or events, even when the perceptual features differ. Generalizing relations despite perceptual differences allows us to make broader inferences. For example, learners can benefit from comparing hard-to-conceptualize phenomena to familiar events, such as science students learning about the circulatory system by comparing it to a highway. Making comparisons like these allows learners to take a familiar structure (e.g., of vehicles transporting passengers and goods) and align it with an unfamiliar target (e.g., blood vessels transporting oxygen and nutrients). It has been argued that relational ability underlies not only learning but also categorization, problem-solving and is a critical component of higher-order reasoning (Gentner, 2003).

However, relational similarity is often not obvious compared to perceptual similarity, and this can make relational learning challenging. Detecting shared relations is especially difficult when perceptual and relational similarity are in conflict. This is evident in the literature on relational learning with preschool age children. Relational learning in this age group is typically tested with relational match-to-sample task, where children are given a sample card showing a relation and then asked to match it to one of two choice cards. With feedback and support, even three- and four-year-old children can find the relational match. They perform far worse, however, when the two choices are a relational match and an object match (Gentner & Toupin, 1986; Paik & Mix, 2006; Rattermann & Gentner, 1998). Fortunately, the literature also reveals that there is a way to overcome this focus on the object match. Like other studies, Christie & Gentner (2010) found that three-year-olds tended to make object matches rather than relational matches when they were shown a single sample card. In a follow-up experiment, Christie & Gentner (2010) showed another group of children two side-by-side samples of the relation. Unlike their peers, the children who could compare multiple exemplars chose the relational match significantly above chance. These patterns are part of two signatures of relational learning that appear across age groups: first, increasing comparison improves relational learning by increasing the alignment of shared structures, and second, increasing the salience of individual objects hinders relational learning.

Based on preschooler's difficulties with relational learning, one might predict that relational ability begins in childhood. However, Ferry et al. (2015) showed that even 7- and 9-month-old infants are capable of learning *same-different* relation. After habituating infants to pairs that instantiated either *same* or *different* relations, Ferry et al. (2015) found that infants succeeded in discriminating this relation from a novel relation. Like preschoolers, infants also made a predictable failure. Prior to seeing the habituation pairs, the infants had been shown a set of individual toys. Despite the fact that infants generalized the *same* and *different* relations to new pairs at test, they failed to discriminate these relations in test pairs that contained the individually-salient objects. Thus, these infants also showed the signatures of relational learning: facilitating comparison across exemplars (the habituation pairs) promoted relational abstraction, but rendering individual objects salient (the toys shown before the experiment) disrupts this ability.

More recently, this paradigm has been extended to 3-month-olds, and the results likewise show that relational ability and the signatures of relational learning are present at this young age

(Anderson et al., 2018b). The results of this experiment also revealed a nuance of how and when comparison can help infants align the shared relation. Infants in Anderson et al. (2015) were habituated to either six unique exemplars or to two exemplars that repeated on alternating habituation trials. Counter to evidence from other domains that chances of learning increase with the number of exemplars (Gerken, 2010; Gerken & Bollt, 2008; Gweon et al., 2010; Needham et al., 2005; Quinn & Bhatt, 2005; Xu & Tenenbaum, 2007), 3-month-olds in Anderson et al. (2018) failed to detect the relations with six unique exemplars and only learned the relation with two repeating exemplars. While counter to many learning theories, this result is not the first study to indicate that greater experience with a small number of exemplars may be more beneficial when it comes to learning abstract categories such as relations and verbs (Bulf et al., 2011; Casasola, 2005b; Maguire et al., 2008). Again, we argued, this relates to the difficulty of detecting relations compared to perceptual features. In this context, the six-exemplar condition acted as a parade of new objects that attracted learners' natural attention to perceptual features. We hypothesized that over the course of the repetitions in the two-exemplar condition, the novelty of the objects diminished and infants were able to align the shared relation (Anderson et al., 2018b). Ultimately, the condition that eased alignment was more beneficial, for generalization cannot occur without it.

The finding that "less is more" for infants learning relations raises key questions about the development of relational concepts. The greatest advantages of relational learning stem from the broad generalization it offers. How does a learner arrive at a broad relational abstraction if a relation is most likely to be abstracted from close comparisons? The theory of progressive alignment and abstraction posits that, after this initial abstraction, later comparisons with new exemplars lead to further abstractions (Gentner et al., 1995; Kandaswamy et al., 2014; Kotovsky

& Gentner, 1996; Kuehne et al., 2000). According to this theory, learners are likely to make their first relational alignments through close comparisons, and these first comparisons result in narrower, more context-dependent abstractions. Yet with each subsequent comparison and abstraction, the relational concept becomes more and more abstract. This, in turn, allows learners to make alignments across increasingly broad comparisons. In line with this, Anderson et al. (2018) argued that increasing the number of exemplars to six should ultimately be beneficial, but only after infants had made the alignment.

The gradual nature of progressive alignment makes it difficult to test experimentally. However, there is evidence of progressive alignment from preschoolers. Kotovsky & Gentner (1996) tested four-year-olds on *ABA* vs. *ABB* relations in a relational match-to-sample task. On each trial, children were given a “standard” (the sample card) and then asked to match it to one of two choice cards (see Figure 1). Consistent with the idea that perceptual similarity is more salient, this age group could easily detect the relational match when the standard and choices all shared common a dimension like size (e.g., the standard and relational match were both “little-big-little”). Children struggled, however, to detect the relation when the standard showed the relation in one dimension (e.g., “little-big-little”), but the choices showed the relation in a different dimension (e.g., “light-dark-light”).

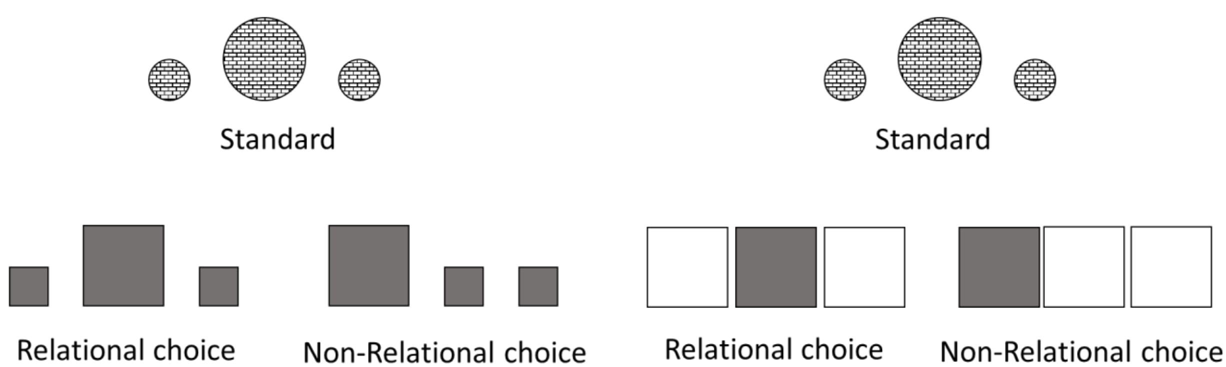

Figure 1. Stimuli from Kotovsky & Gentner (1996). Children were asked to match a triad like *ABA* (top) within the same dimension (size; left) and across dimensions (from size to lightness; right).

Progressive alignment came in to play when Kotovsky & Gentner (1996) had one group of children start by only matching the same-dimension items, capitalizing on these close comparison to facilitate the initial alignment. Then in the second block, children progressed to seeing only the cross-dimension items. Now four-year-olds were able to transfer the relations across dimensions. This series of progressive comparisons was uniquely positioned to accomplish two things: 1) encourage alignment during a close comparison and 2) push children to eventually arrive at a less context-dependent relational abstract, allowing them to generalize further.

In theory, progressive alignment and abstraction should apply to any learner abstracting a new relation. However, there is a gap of almost four years between the preschool-age participants of Kotovsky & Gentner (1996) and the three-month-old participants in Anderson et al. (2018). Unlike preschoolers, infants do not have the benefits of a more mature working memory and inhibitory control processes, nor do they have cultural supports like relational language, all of which predict improvements in relational ability (Begolli et al., 2015; Christie & Gentner, 2014; Rattermann & Gentner, 1998; Richland et al., 2017; Simms et al., 2018). Though Anderson et al. (2018) offers evidence that the alignment process benefits from close comparisons for infants as it does for preschoolers, it is unknown how gradual the process of progressive abstractions would need to be, or whether infants can make any further abstraction in the first months of life. In the following experiments, we test for whether progressive alignment can balance the need for alignment and generalization in infancy as well. In doing so, it will further demonstrate how similar the infant relational learning process is to the one seen in children and adults.

As mentioned above, Anderson et al. (2018) had found that 3-month-old infants could learn *same* and *different* from two repeating exemplars, but not from six unique exemplars. Yet they argued that six exemplars would have been helpful for generalization if infants *had* been able to make an initial relational alignment. Experiment 1 tests this claim by first showing 3-month-old infants three perceptually similar exemplars of a *same* or *different* relation to facilitate alignment, and then introducing three less similar exemplars, to provide the means of an initial alignment and subsequent broader experience. If progressive abstractions are already possible in the first months of life, then habituating infants to a series of six exemplars that are progressively aligned should increase infants' chances of learning *same-different* relations in this experiment.

Though *same* and *different* relations have been used to demonstrate infant relational learning before, using these relations in the context of two progressive alignment series presents new challenges. Kovotsky & Gentner (1996) used *ABA* and *AAB* relations, where both target relations were forms of *different*. This meant that the objects used in the standard and test items could be identical (e.g., two A objects and one B object) even though the relations that they formed differed (*AAB* vs. *ABA*). In contrast, *different* pairs have twice as many objects as *same* pairs. This has practical implications for the progressive alignment in two ways. First, it is easy to think about how a single series of objects could progress from more to less similar, and just as easy to double those objects into *same* pairs, but it is challenging to think about how two separate objects should progress side-by-side in a way where the relation different is equally alignable between pairs. Secondly, to ensure that infants who habituate to *different* will not have to learn more objects in one condition than in another, our studies typically keep the total number of objects the same (e.g., if a *same* progression was *d-d, b-b, p-p, q-q,* then the *different* progression would be something like *d-b, b-p, p-q, q-d*). Given the possibility that the *same* progressive alignment might be clearer that the *different* condition, it may be that infants who habituate to *same* in this experiment would be more likely to succeed.

Experiment 1

In Experiment 1, we ask whether 3-month-olds could learn *same-different* relations with six exemplars if they progressed from higher to lower similarity, potentially allowing for an initial alignment and then progressive abstraction. We used a habituation/ dishabituation looking time paradigm modeled on earlier infant *same-different* studies (Anderson et al., 2018b; Ferry et al., 2015). Prior to the study, the infants saw eight individual objects in the waiting room (see Figure 2A). In the study room, infants were then habituated to a series of six pairs that shared either the *same* or the *different* relation. In the *same* condition, the first three object pairs were all animals and the next three were all blocks. In the *different* condition, the first pair was two animals, the next three pairs were animal-block combinations, and the fifth pair was two blocks. These were presented over the course of 6 to 9 habituation trials, in a fixed progression (see Figure 2B). Finally, infants saw three types of test trials: *New* trials with previously unseen objects, *Salient Object* trials that contained the objects shown before habituation, and *Salient Object + Habituation* trials that contained the objects that had been shown individually and then in the context of pairs in the habituation phase. Each of these test trial types were composed of two back-to-back trials: one featuring the familiar relation (e.g., *same* if they had habituated to *same*) and one featuring the novel relation (e.g., *different* if they had habituated to *same*).

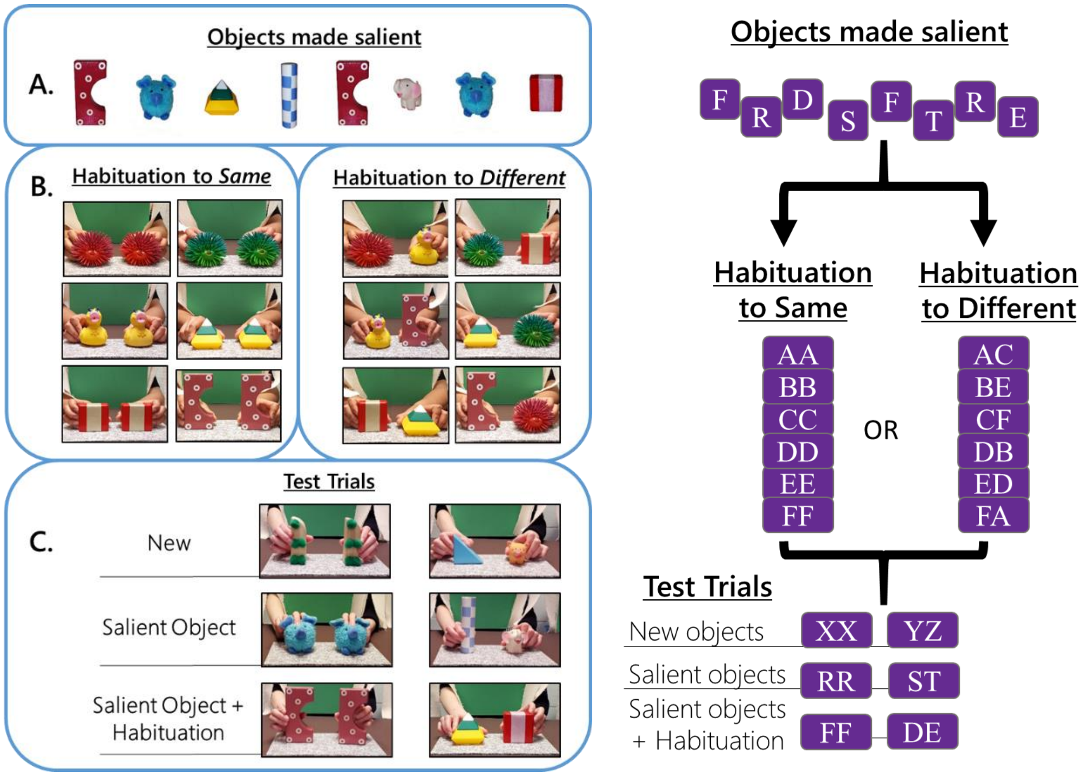

Figure 2. Pictures of the stimuli used in Experiment 1 and a schematic of the procedure. (A) Infants saw a subset of the individual toys before the experiment. (B) All infants saw a minimum of 6 habituation trials, all same or different, with one pair was shown per trial. If infants' looking did not decline within 6 trials, the habituation sequence began again with the top left pair. (C) Infants saw three types of test trials: *New*, *Salient Object*, and *Salient Object + Habituation*.

If the progressive alignment order allows 3-month-olds to abstract the relations, then infants should be able to generalize the relation to never-before-seen objects and look longer at the novel relation in the *New* trials. The second trial type is *Salient Object*. Because one of the signatures of relational learning is that drawing attention to individual objects will hurt relational learning, we predict that even if infants can generalize to *New*, infants should look equally between the relations on *Salient Object* trials where objects that were shown individually before habituation are reintroduced. The third test trial is *Salient Object + Habituation*, where infants have seen objects individually, but also as an exact pair from habituation. Though infants may be able to recognize this habituation pair, these objects have also been made individually salient, increasing the chance that infants will not focus on the relation. In fact, while 7- and 9-month-old infants from Ferry et al., (2015) successfully discriminated the novel and familiar relations in this trial type, 3-month-old infants from Anderson et al., (2018) did not. Thus it is an open question as to whether infants in this experiment will discriminate *Salient Object + Habituation* test trials.

Additionally, because the *different* progression was more challenging to construct than the *same* progression, there may be an interaction with habituation condition, such that *same* is easier to align than *different*. If this is the case, infants in the *same* condition should generalize to *New* but infants in the *different* condition may not. Infants may also show this pattern with *Salient Object + Habituation* pairs, which are challenging for this age group. Infants in both conditions should look equally at the relations in *Salient Object* trials, regardless of habituation condition.

Experiment 1 Methods

Participants

The participants were 32 healthy, full-term 3-month-old infants (12 female, *mean age* = 3 months, 13 days). This sample size was used because it was the same size as an experiment that had successfully demonstrated relational learning with 3-month-olds (Anderson et al., 2018)[6]. Half of the 3-month-old infants were assigned to the *same* condition; half to the *different* condition. Nine additional infants were excluded from this: four for fussiness (judged as fussy or crying by two independent coders for more than half the test trials), three because they

[6] Since this data was collected, a target sample size of $N = 48$ was determined based on power analyses of the 3-month-olds who demonstrated relational learning on *New* trials in Anderson et al. (2018). Therefore, we conduct Bayesian analyses in the results to determine how strong the evidence is with the current sample size.

experienced bowel movements during the test trials, and two because they took long or frequent breaks (defined as taking breaks longer than five minutes or taking three or more breaks).

Parents of infants were recruited through online ads and word of mouth. Parents who agreed to their infants' participation were provided informed consent before the experiment and given $20 as compensation. The race, ethnicity and parental education level of the sample is described in the table below.

Table 1. Participant demographic information

Ethnicity	%	Race	%	Education	Maternal %	Paternal %
Hispanic	12	African American	7	Some high school	0	2
Non-Hispanic	81	Asian/ Hawaiian/ Pacific Islander	5	High school diploma	0	3
		American Indian	2	Some college	2	9
		White	63	College degree	93	81
		Multiracial	14			
No response	7	No response	9	No response	5	5

Apparatus

Parents sat in a chair with infants on their lap facing a wooden puppet stage that displayed all stimuli. The parents were asked to refrain from interacting with the infant during the experiment and to close their eyes during the test trials. The stage measured 243.5 cm high, 128 cm wide, and 61 cm deep. The opening in the front of the stage that displayed the objects was 93 cm above the floor, 61 cm high, and 106 cm wide. The back wall had two rectangular openings with cloth fringe over the openings that allowed the experimenter to manipulate the objects. A screen that covered the infants' view of the stage was raised and lowered between trials. The MATLAB program Baby Looking Time (BLT), was used to record looking times for habituation and test trials during the experiment (Chang et al., 2018).

The stimuli were 16 three-dimensional objects (see Figure 2). Half of these objects had been used in the six-exemplar condition of Anderson et al. (2018), while the other half were replaced to increase the level of similarity between the first three pairs. Eight of the objects were shown one at a time before the experiment began. Six pairs of objects were shown in the habituation phase within *same* or *different* pairs. Finally, 12 objects appeared in pairs during the test phase. Of these, 8 objects had been seen individually before the experiment, 4 of these had also been seen in habituation trials, and 4 objects had never been seen before this point.

Procedure

The experiment consisted of three sequential parts:

(A) Objects made salient. We manipulated infants' attention on some individual objects by showing four of the test objects to the infant in the waiting room prior to the experiments, during naturalistic play interactions (see Figure 2A). Showing the objects one at a time, the experimenter held each object between the infant and themselves for 5 seconds while they jointly attended to it and made comments such as "Look!" and "See this one?". The two identical objects were never shown in immediate succession.

(B) Pair habituation trials. When the screen was raised at the start of every trial, a pair of objects rested on the cardboard tray on the stage. To engage infants' attention, in both habituation and test trials, the pairs of objects were moved during the trial. The experimenter grasped one object in each hand and raised the objects straight up (1 s), tilted them to the left (1 s), returned them to the center (1 s), tilted them to the right (1 s), returned them to the center (1s), returned them to the tray (1 s), and paused on the tray (2 s). This 8-s cycle repeated continuously until the trial ended. In the *same* condition, infants saw habituation trials in which the pairs of objects were the *same* (see Figure 2B). In the *different* condition, infants saw habituation trials in which the pairs of objects were *different*. The number of habituation trials was infant-controlled (see Coding section for the criterion), ranging from 6 to 9 trials.

(G) Test trials. For each test trial type, infants were presented with a novel relation, followed by a familiar relation (or vice versa), summing to six total test trials. In each test trial, infants viewed one pair of objects, presented in the same motion pattern as in the habituation trials, while their looking time was recorded. The three kinds of test trials were (a) objects that had not been seen before test (*New*); (b) objects that the infant had seen individually in the waiting room (*Salient Object*); and (c) objects that the infant had seen individually in the waiting room and then again in pair habituation (*Salient Object + Habituation*). There were three trial orders (abc, cab, bca) which were counterbalanced across infants.

Coding

There was a small hole in the front face of the stage containing a camera that captured a video image of the infant's face. While the experimenter conducted habituation and test trials in the room with the infants, two research assistants in a separate room viewed the video and coded infants' visual fixations online as either on target or off. Each researcher depressed a computer button when the infant attended to the events on stage and released the button when the infant looked away. Each trial ended when the infant either looked away for 2 consecutive seconds after having looked at the event for at least 2 s or looked at the event for 60 cumulative seconds without looking away for 2 consecutive seconds. The BLT program determined the end of the trial and beeped, signaling to the experimenter to lower the screen and move to the next trial. After each test trial, research assistants also checked one or more boxes to indicate the behavioral state of the infant on the preceding trial: sleepy, quiet and alert, active, fussy or crying. Coders also noted any breaks and their length. If two coders independently judged the infant’s state as fussy, crying, or falling asleep for more than half the test trials, the infant's data was excluded from the analysis. The coders were blind to the condition and the trial order. Interobserver agreement was measured for all infants and averaged 89.7%. The Intraclass Correlation Coefficient for a fixed set of raters (ICC3K) was .98 with a 95% confidence interval from .97 to .99, $F(31, 189) = 50$, $p < .001$. Our looking time data significantly deviated from a normal distribution per the Shapiro-Wilks test. Therefore, we performed parametric tests on log-transformed data, following recommendations outlined by Csibra, Hernik, Mascaro, Tatone, and Lengyel (2016).

Experiment 1 Results

Our first question was whether 3-month-olds infants looked longer at a novel relation after habituation to a progressive alignment series of six exemplars. To answer this, we conducted a repeated-measures ANOVA on the *New* test trial type, looking at the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variables of Condition (Habituation to *same* vs. *different*), Order (novel vs. familiar first) and Sex (male vs. female). Infants looked evenly across relations in the *New* trial: 28.56 s ($SD = 20.07$) at novel relations and 29.15 s ($SD = 18.51$) at familiar ones, and there was no main effect of relation, $F(1,21) < 1$, $p = .58$, (see Figure 3). This lack of discrimination was also reflected in the binomial comparison, where only 12 of the 29 infants with usable *New* trials looked longer at the novel relation compared to the familiar one, $p = .87$. There was no interaction between relation and condition, $F(1,21) < 1$, $p = .82$, nor any other main effects or interactions.

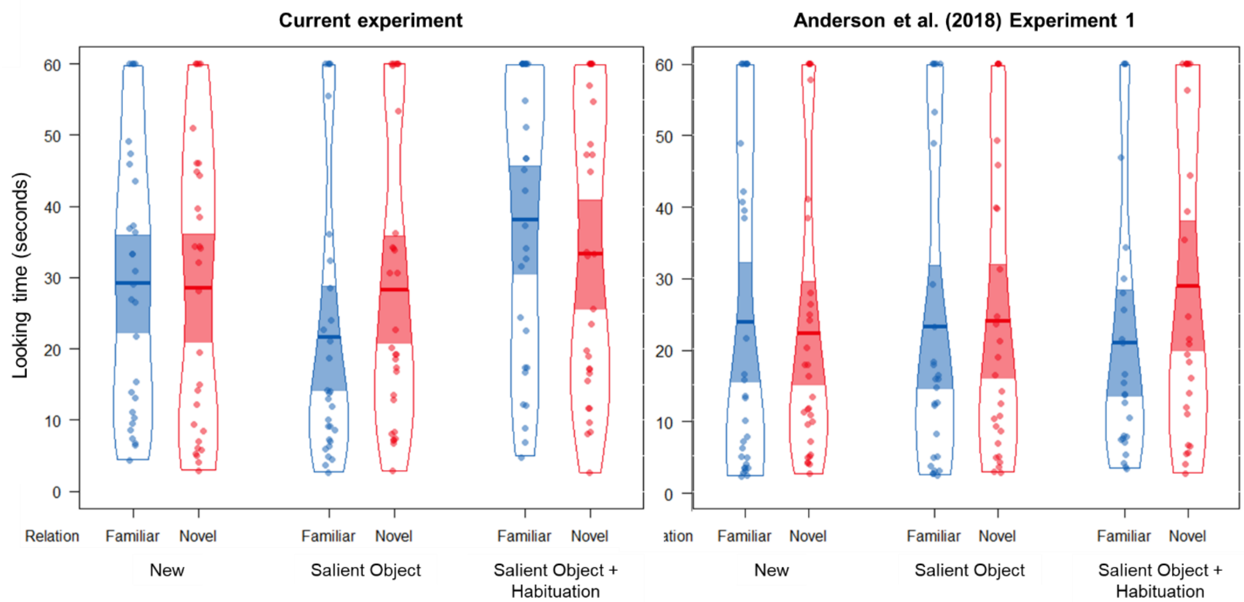

Figure 3. Looking times at novel and familiar pairs for each test type collapsed across *same* and *different* conditions, for the current experiment (left) and for Anderson et al. (2018) (right).The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Next, we predicted that infants would fail to generalize the relation to pairs containing objects that had been made salient. We conducted a repeated-measures ANOVA on the *Salient Object* test trial type, looking at the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variable of Condition (Habituation to *same* vs. *different*), Order (novel vs. familiar first) and Sex (male vs. female). Collapsing across conditions, infants looked somewhat longer at the novel relation (M = 28.31 s, SD = 19.95) than the familiar relation (M = 21.55 s, SD = 19.42), but there was no main effect of relation, $F(1,21) = 2.45$, $p = .13$. This lack of discrimination was also reflected in the binomial comparison: only 18 of the 29 infants with usable *Salient Object* trials looked longer at the novel relation compared to the familiar one, p =

.13. Again, there was no interaction between condition and relation, $F(1,21) = 2.16$, $p = .16$, nor any other main effects or interactions.

Finally, we asked whether infants would discriminate the novel from the familiar relation with test objects that had been presented individually but also presented in the context of habituation pairs. We conducted a repeated-measures ANOVA on the *Salient Object + Habituation* test trial type, looking at the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variable of Condition (Habituation to *same* vs. *different*), Order (novel first vs. familiar first) and Sex (male vs. female). Here, infants on average looked longer at the familiar relation ($M = 38.12$ s, $SD = 19.92$) than at the novel relation ($M = 33.32$ s, $SD = 20.20$), though there was no main effect of relation, $F(1,21) = 1.13$, $p = .299$. This non-significant familiarity preference was reflected in the binomial comparison: only 10 of the 29 infants with usable *Salient Object + Habituation* trials looked longer at the novel relation compared to the familiar one, $p = .97$. An error in the design of the study meant that infants in the *same* condition had harder *Salient Object + Habituation* trials than infants in *different* condition. Infants in the *different* condition looked slightly longer at the novel relation compared to the familiar for this test trial type (39 s vs. 36 s) while infants in the *same* condition showed a familiarity preference (27 s vs. 40 s), but there was no interaction between Condition and Relation, $F(1,21) = 2.36$, $p = .14$. Similarly, while 7 of 15 infants in the *different* condition looked longer at the novel pair, $p = .70$, only 3 of 14 infants looked longer at the novel *Salient Object + Habituation* pair in the *same* condition, $p = .99$. There was an interaction between Condition and Order, such that infants habituated to *same* had longer looking times in general in the familiar-relation-first order, but infants habituated to *different* had similar average looking times in both orders, $F(1,21) = 5.24$, $p = .033$, $\eta_p^2 = .20$. There were no other main effects or interactions.

Bayesian analyses

To determine how conclusive these null results are with the sample size of $n = 32$, we found the Bayes Factors for paired t-tests of the novel and familiar pair in each test trial type. In contrast to parametric test, which only compares an alternative hypothesis to a null hypothesis, Bayes Factors indicate whether there is clear evidence for the null or for the alternative or whether the evidence is too weak to conclude either way. A Bayes Factor greater than 10 indicates strong evidence of an effect, a factor greater than 3 indicates moderate evidence of an effect, while a factor in between 3 and 0.33 suggests that the evidence is inconclusive. BF_{10} indicates the Bayes Factor for the alternative hypothesis (i.e., that infants are discriminating between novel and familiar relations), while BF_{01} indicates the Bayes Factor for the null hypothesis (i.e., that infants did not discriminate). In Experiment 1, there was moderate evidence that infants did not generalize the relation on the *New* trial, $BF_{01} = 3.69$ nor on the *Salient Object + Habituation trials*, $\mathrm{BF}_{01} = 3.04$. However, there was only weak evidence that infants did not generalize on the *Salient Object* trial, $BF_{01} = 1.38$.

Comparison to Anderson et al. (2018)

As in the current experiment, Experiment 1 in Anderson et al. (2018) was one where 32 3-month-old infants habituated to six exemplars of *same* or *different*. These two experiments also shared many of the same stimuli in habituation and test trials. The critical difference between them is that in the current experiment, a fixed progression during habituation was used across all infants. To test whether this intervention improved generalization, we conducted a repeated-measures ANOVA on the *New* test trial type across these two experiments, looking at the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variables of Experiment (Anderson et al. 2018 vs. current experiment), Condition (Habituation to *same* vs. *different*), and Order (novel first vs. familiar first). Sex wasn't included here because the cell sizes were too imbalanced to meet the assumptions of the ANOVA. There was no main effect of Relation in *New* trials across these studies, $F(1,51) < 1$, $p = .73$, and Relation did not interact with Experiment, $F(1, 51) < 1$, $p = .38$. There was a main effect of Experiment, where infants in the current experiment had longer looking times overall during *New* test trials ($M = 28.85$, $SD = 19.14$) compared to infants in Anderson et al., (2018)'s Experiment 1 ($M = 22.96$, $SD = 21.14$), $F(1, 51) = 4.81$, $p = .033$, $\eta_p^2 = .09$. There was also a Condition by Order by Experiment interaction such that infants in Anderson et al. (2018) had longer looking times in the familiar-first test order when they habituated to *different* but the reverse was true when they habituated to *same*. Infants in the current experiment, however, looked longer overall in the novel-first order in both conditions, $F(1, 51) = 5.16$, $p = .027$, $\eta_p^2 = .09$. Finally there was a Relation by Condition by Order interaction where infants across experiments looked longer at *same* trials when they were first in the test pair, whether or not this was the relation they had habituated to, $F(1, 51) = 4.81$, $p = .033$, $\eta_p^2 = .09$. There were no other interactions or main effects.

Next, we conducted a repeated-measures ANOVA on the *Salient Object* test trial type across the two experiments, with the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variables of Experiment (Anderson et al. 2018 vs. current experiment), Condition (Habituation to *same* vs. *different*), and Order (novel first vs. familiar first). There was no main effect of Relation across these studies, $F(1, 48) = 3.19$, $p = .08$, and Relation did not interact with Experiment, $F(1, 48) < 1$, $p = .34$. There was a Condition by Experiment interaction, such that infants in Anderson et al. (2018) had longer overall looking times in the *same* condition but infants in the current experiment had longer looking times in the *different* condition, $F(1, 48) = 7.73$, $p = .008$, $\eta_p^2 = .14$. There were no significant main effects or other interactions.

Finally, we tested for whether the progressive alignment intervention improved performance on the *Salient Object + Habituation* test trial, by conducting a repeated-measures ANOVA on this test trial type across the two experiments, with the within-subjects variable of Relation (Novel vs. Familiar) and the between-subjects variables of Experiment (Anderson et al. 2018 vs. current experiment), Condition (Habituation to *same* vs. *different*), and Order (novel first vs. familiar first). There was no main effect of Relation across these studies, $F(1, 55) < 1$, $p = .70$, and Relation did not interact with Experiment, $F(1, 46) < 1$, $p = .96$. Here there was a Relation by Experiment by Condition interaction, such that infants habituated to *same* preferred the novel relation in Anderson et al. (2018)'s Experiment 1 while infants habituated to *same* in the current experiment preferred the familiar relation, $F(1, 46) = 4.07$, $p = .05$, $\eta_p^2 = .08$. There was also a main effect of Experiment, where infants in the current experiment had longer looking times overall during *Salient Object + Habituation* test trials ($M = 35.72$, $SD = 20.03$) compared to infants in Anderson et al., (2018)'s Experiment 1 ($M = 24.83$, $SD = 20.50$), $F(1, 46) = 10.48$, $p = .002$, $\eta_p^2 = .19$. Additionally, there was a Condition by Order by Experiment interaction, such that infants in Anderson et al. (2018) had longer overall looking times in the *same* condition if they saw the novel relation first and longer looking times in the *different* condition if they saw the familiar relation first, but the reverse was true in the current experiment, $F(1, 46) = 12.51$, $p < .001$, $\eta_p^2 = .21$. There were no other significant main effects or interactions.

Habituation patterns

Finally, we examined looking time during the habituation phase. Because it is easier to conceptualize a habituation pair progression when both the objects are the *same*, we anticipated that there might be differences in the habituation decline between conditions, commensurate to the ease of learning the *same* progression compared to the *different* one. A repeated measures ANOVA examined looking time by the within-subjects factor of Habituation Trial (first three and last three) and the between-subjects factor of Condition (*same* and *different*), and revealed that, across conditions, there was a significant decline between habituation trials, $F(5, 150) = 7.25$, $p < .001$, $\eta_p^2 = .20$ (see Figure 4). On average, infants looked 49.82 s (SD = 17.69) on their first habituation trial and 27.20 s (SD = 21.35) on their last trial. There was not a significant interaction between habituation trial and condition, $F(5, 150) < 1$, $p = .71$, nor a main effect of condition, $F(1, 30) < 1$, $p = .473$. Additionally, there were no between-condition differences in terms of total looking time during habituation, $F(1,30) < 1$, $p = .74$, in how many trials it took to reach the habituation criterion, $F(1,30) < 1$, $p = .70$, or on whether infants reached the habituation criterion, X^2 (1, N = 32) = 1.17, $p = .28$.

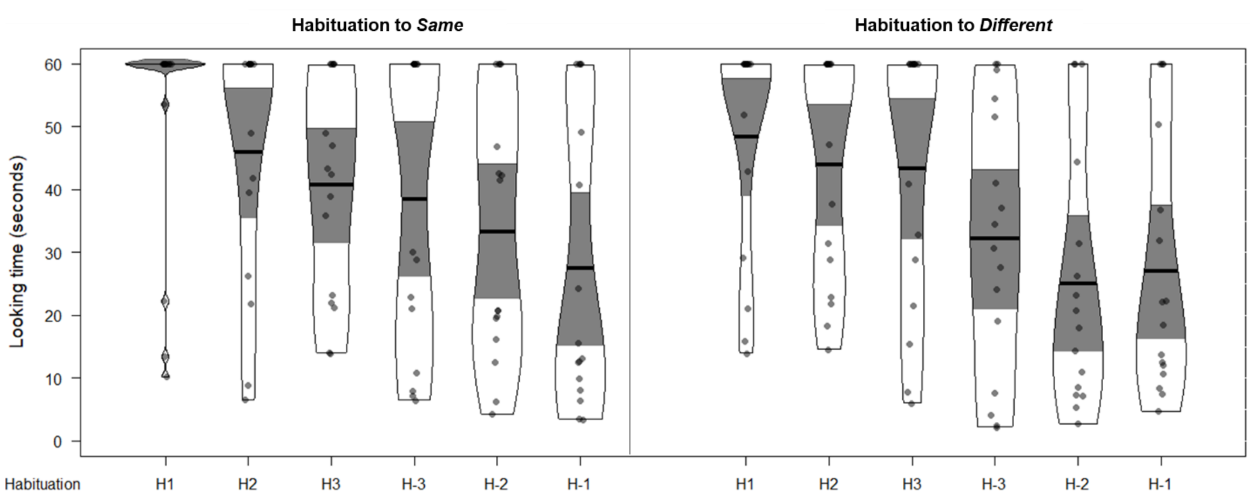

Figure 4. Looking decline over the course of 6 to 9 habituation trials, separated by habituation condition. Because the number of habituation trials was infant-controlled, the graphs represent the first three (H1, H2, H3) and the last three trials (H-3, H-2, H-1). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Experiment 1 Discussion

We did not find evidence that 3-month-olds learned *same* or *different* from six unique exemplars, even with a fixed progression designed to go from higher to lower similarity. Infants did not discriminate the novel and familiar relations on any of the three test types. We had predicted that infants would fail to discriminate in the *Salient Object* trial type. However, because infants also failed to generalize the relation to *New* pairs, it is unclear whether this result reflects a focus on the objects involved or simply a lack of relational learning. Furthermore, although the progression from high to low similarity was less straightforward in *different* condition, the results show that relational learning did not depend on the habituation condition: infants in both conditions looked equally at test.

This set of null results mirrors what was found in Anderson et al. (2018), Experiment 1, and there were no significant interactions between Relation and Experiment for any of the trial types. In the very least, this indicates that the fixed progression we used in Experiment 1 was not more helpful for making a relational abstraction than the random progressions in Anderson et al. (2018). Together, these findings suggest that any series of six unique pairs over the brief course of habituation were too challenging for 3-month-olds to align, no matter how similar the objects are to each other. However, it is important to note some limitations before drawing this larger conclusion. One is that we do not know how similar the pairs need to be for 3-month-olds to make an initial relational alignment. While the *same* habituation progression in Experiment 1 shifted from animals to blocks and tried to place transitions across similar colors, infants may need even higher levels of similarity to not be drawn by the new objects in subsequent habituation pairs. A potential piece of evidence that this age group needs a more similar set of exemplars to make the initial alignment comes from another study of *same-different* learning in 3-month-olds, which found that infants abstracted the relations from a single exemplar that repeated throughout habituation (Anderson et al., in preparation). If infants in this age group can learn the relations from identical presentations but not from the progression in Experiment 1, then the question remains as to whether they could align the relations across a more highly similar set of six exemplars.

To increase the level of similarity in Experiment 2, we used Greeble objects, which have a set of predictable features organized in highly constrained configurations (Gauthier & Tarr, 1997). Though individual Greeble objects differ from each other, they have been normed through similarity ratings so that they vary along two predictable dimensions. This means that any progression of pairs will be more similar to each other than those of Experiment were. It also means that we have more control over the level of similarity between subsequent pairs by varying one or both dimensions across these transitions. If 3-month-old infants can learn from six unique exemplars over the habituation period, then increasing the level of similarity and better manipulating the progression should help them to abstract the relations.

Raising the overall level of similarity across habituation trials could come at a cost, though. High similarity is more likely to enable learners to align the relation, but higher similarity comparisons also result in narrower generalizations (e.g., Rosch, 1978; Xu & Tenenbaum, 2007). While the progressive alignment method is designed to navigate this tradeoff, the final abstraction may not be much broader if the progressive abstractions occur across exemplars that are also fairly similar to the initial pairs. Therefore, Experiment 2 also changed the three test trial types to assess whether resulting abstractions were narrower. To do so Experiment 2 used two types of new trials: a *Near Transfer* trial featuring new Greeble objects and a *Far Transfer* trial featuring new non-Greeble objects. If infants learn the relations from the highly similar progression, then we predict that they will generalize to the *Near Transfer* trial, but they may not generalize to the *Far Transfer* trial.

Additionally, the *Salient Object + Habituation* trial – a challenge even when three-month-olds can learn the relation (Anderson et al., 2018) – is replaced by a *Memory Check*, which pits novel objects in a novel relation against familiar objects in a familiar relations. By adding a *Memory Check* trial, the experiment is better able to test whether infants learned anything from the six exemplars, even if infants fail to generalize to either new trial type.

Finally, as in Experiment 1, it is easier to create a progression of pairs when the objects are the *same* than it is to manage two simultaneous object progressions, as in the different condition. We again predict that *same* may be easier to learn from these progressions than *different*.

Experiment 2

As Experiment 1, infants were habituated to six pairs of objects that shared a common relation (e.g., all *same* or all *different*). These were presented over the course of 6 to 9 habituation trials in a fixed progression. In contrast to Experiment 1, though, all of the habituation objects were highly constrained Greeble objects (see Figure 5). Next, infants saw three types of test trials: *Near Transfer* trials with previously unseen Greeble objects, *Far*

Transfer trials with previously unseen objects that were not Greebles, and the *Memory Check* trial. The *Memory Check* trial type was added as an easier alternative to the *Salient Object + Habituation* trial in response to the widespread failures to discriminate relations in Experiment 1. Each of the four test trial types included two sequential trials: one featuring the familiar relation (e.g., *same* if they had learned *same*) and one featuring the novel relation (e.g., *different* if they had learned *same*).

If this more similar progression allows 3-month-olds to align across six exemplars, then infants should look longer at the novel relation in the *Near Transfer* trials, featuring new Greeble objects. If increasing the similarity of the habituation set has not limited their generalization, then infants should also look at the novel relation in the *Far Transfer* trials, featuring new objects that are not Greebles. Regardless of whether relational learning occurs, infants should discriminate the relation in *Memory Check* trials.

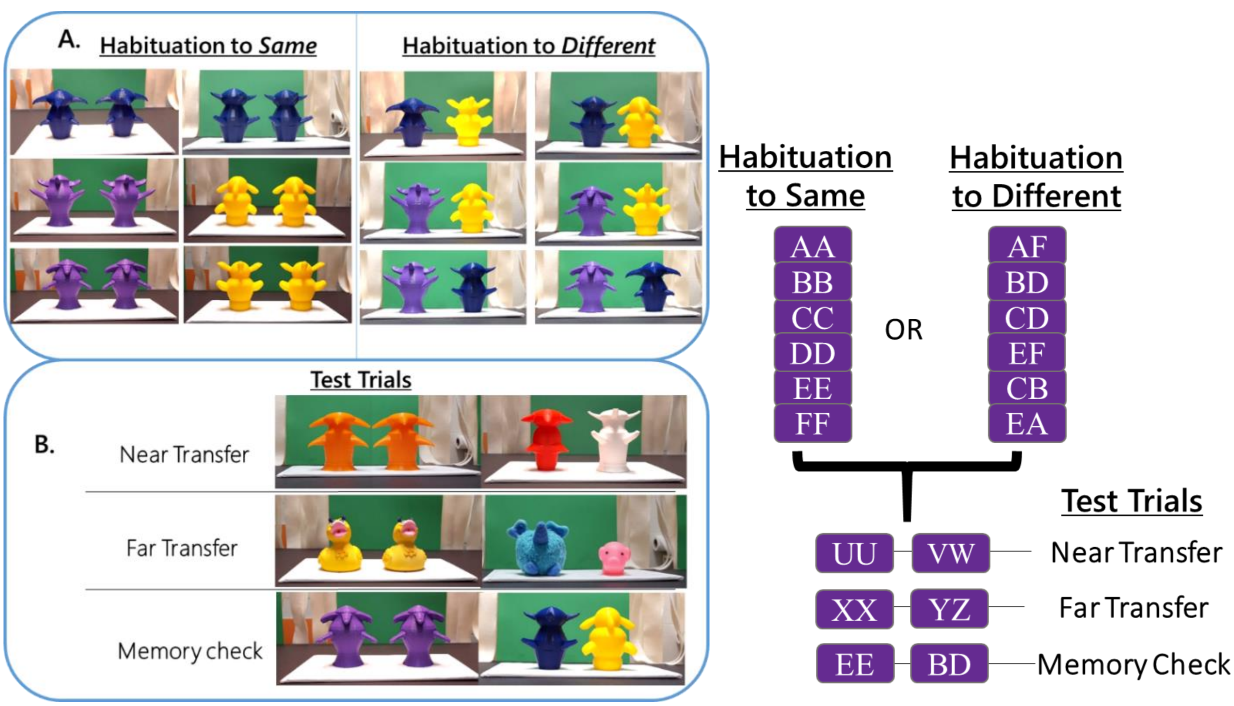

Figure 5. Stimuli and schematic of the procedure for Experiment 2. Infants are habituated to *same* or *different* Greeble objects, then tested on their recognition of novel vs. familiar relations for pairs made of habituation objects (*Memory Check*); new Greeble objects (*Near transfer*) and non-Greeble objects (*Far transfer*). The order of test trials is counter-balanced across infants, while the habituation progression remains fixed for all participants.

Another nuance is the possibility that *same* will be easier than *different*. If this is the case, we may see generalization to *Near Transfer* trials, and potentially to *Far Transfer* trials, in the *same* condition, but not in the *different* condition. However, we predict that infants should be able to discriminate novel and familiar pairs in the *Memory Check* trials in all cases.

Participants

The final sample will include 48 healthy, full-term 3-month-old infants, a sample size based on power analyses of the 3-month-olds who demonstrated relational learning on *New* trials in Anderson et al. (2018). Currently, the sample includes 35 usable infants. Another 15 infants have been excluded: 8 for fussiness, 5 for looking the maximum amount of time on all test trials, 1 for falling asleep, and 1 for a 9-minute break during test. This exclusion rate (~30% of participants) is consistent with that of other 3-month-olds studies using this paradigm: 31% Anderson et al. (2018) and 27% in the 3-month-old age group in Experiment 1. Recruitment and compensation were the same as in Experiment1. Demographics are below.

Table 2.

Participant demographics for Experiment 2

Ethnicity	%	Race	%	Parent Education (combined)	%
Hispanic	10	African American	17	Some high school	0
Non-Hispanic	88	Asian/ Hawaiian/ Pacific Islander	5	High school diploma	1
		American Indian	0	Some college	5
		White	67	College degree	93
		Multiracial	12		

No response	2	No response	0	No response	1

Apparatus

The stage is the same as in Experiment 1. The stimuli are 20 three-dimensional objects. All of the habituation objects and all but four of the test objects are 3-D printed adaptations of Gauthier & Tarr (1997)'s Greeble objects. Twelve Greeble objects are shown in the context of six *same* or *different* pairs in the habituation phase (see Figure 3). Twelve objects then appear across pairs during the test phase. Of the test objects, four are Greeble objects had been seen previously in habituation, four are Greeble objects that have not been seen before, and four are non-Greeble objects that have also not been seen before.

Design & Procedure

The design and procedure were modeled on Experiment 1 with the following differences. First, there was no "objects made salient" phase. Second, the six habituation pairs were all composed of Greeble objects, which varied on two dimensions: family (indicated by body shape and color) and gender (indicated by whether the nose, arms, and ears pointed upward or downward). The test trials differed from Experiment 1 in that there was were two trial types with new objects: *Near Transfer* and *Far Transfer*, measuring relational generalization to pairs with Greebles and without Greebles, respectively. Additionally, the *Salient Object* trial type was dropped and the *Salient Object + Habituation* trial type was replaced with *Memory Check* trial. As in Experiment 1, the order of the test trials was counterbalanced across participants.

Coding

Coding was the same as in Experiment 1. The coders were blind to the condition and the trial order. Interobserver agreement and the intra class correlation between coders will be calculated after data collection has been completed. Our looking time data skewed towards the lower bound and significantly deviated from a normal distribution per the Shapiro-Wilks test, so we performed parametric tests on log-transformed data, as in Experiment 1.

Experiment 2 Preliminary Results

Current n = 35. The additional 13 infants will be tested after COVID-19 suspensions end.

Our main question is whether three-month-olds can abstract the relation from six highly similar pairs and generalize it to new objects. When we reach our target sample ($n = 48$), we will conduct a series of repeated measures ANOVAs on the within-subjects variable of Relation (novel vs. familiar) and the between-subjects variables of Condition (habituation to *same* vs. *different*), Order (novel or familiar first), and Sex (male or female) for each of the three test trial types. To understand the directions that the data is taking with the current sample ($n = 35$), we use non-parametric binomial comparisons and derive the Bayes Factors for each test type. As discussed earlier, BF_{10} indicates the Bayes Factor for the alternative hypothesis (i.e., that infants are discriminating between novel and familiar relations), while BF_{01} indicates the Bayes Factor for the null hypothesis (i.e., that infants did not discriminate). A Bayes Factor greater than 3 indicates moderate evidence of an effect, while a factor in between 3 and 0.3 suggests that the evidence is inconclusive.

The first trial type is *Near Transfer*. The current means do not indicate any evidence of transfer to pairs of new Greebles, with infants looking 26.63 s ($SD = 21.57$) at the novel relation and 25.80 s ($SD = 21.57$) at the novel relation (see Figure 6). The binomial comparison also reflects this lack of generalization, with only 16 of 34 infants looking longer at the novel relation, $p = .70$. Additionally, a Bayesian t-test comparing the novel relation to the familiar one for *Near*

Transfer shows that there is already moderate evidence (3 < *BF* < 10) that the null hypothesis is true and that infants are not generalizing, BF_{01} = 4.27. Differences between novel and familiar looking times are more pronounced in the *same* condition than the *different* condition (see Table 3), but the binomial comparisons are similar in both conditions: 9 of 16 infants in the *same* condition looked longer at the novel relation, *p* = .40, and 10 of 19 infants did so in the *different* condition, *p* = .50. Bayesian t-tests show that there is currently weak to moderate evidence that the null hypothesis is true and that infants are not generalizing in either the *same* (BF_{01} = 3.04) or the *different* condition (BF_{01} = 3.10).

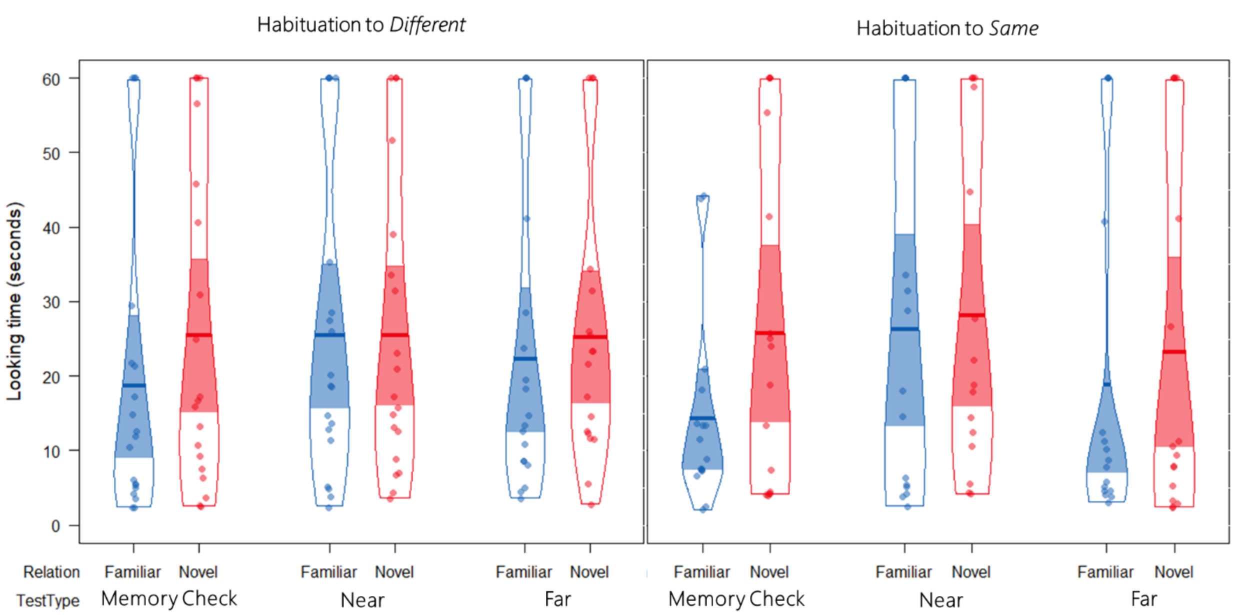

Figure 6. Preliminary looking time results from 35 infants in the current sample. Looking times are separated by test type and by *same* vs. *different* conditions. The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Table 3. Average looking times at test, separated by test type and habituation condition

		Memory Check		*Near Transfer*		*Far Transfer*	
Condition	*N*	*Novel*	*Familiar*	*Novel*	*Familiar*	*Novel*	*Familiar*
Same	16	25.76	14.29	28.11	26.24	23.18	18.88
Different	19	25.52	18.63	25.47	25.45	25.20	22.28
Combined	35	25.63	16.65	26.63	25.80	24.25	20.68

The second test trial type is *Far Transfer*. If progressive alignment has allowed infants a broader abstraction despite the high similarity of the set, then infants should look longer at the novel relation than familiar relation even if the objects in the pairs are not Greebles. Though infants looked equally at *Near Transfer* test trials, they looked slightly longer at the novel *Far Transfer* trial (M = 24.25 s; SD = 20.81), compared to the familiar Far trial (M = 20.68 s; SD = 20.69). However, a Bayesian t-test comparing the novel relation to the familiar one for this test type shows that there is already moderate evidence for the null, i.e., that infants are not generalizing to the *Far Transfer* test trials, BF_{01} = 3.24. Similarly, a binomial comparison shows that only 19 of 35 infants looked longer at the novel relation, p = .37. Infants looked nearly equally at the novel and familiar looking times in both the *same* and *different* conditions (see Table 3), and the binomial comparisons are similar in both conditions: 8 of 15 infants in the *same* condition looked longer at the novel relation, p = .05, and 8 of 19 infants did so in the *different* condition, p = .82. Similarly, Bayesian t-tests for *Far Transfer* show that there is already moderate evidence that the null hypothesis is true and that infants are not generalizing in either the *same* (BF_{01} = 3.20) nor in the *different* condition (BF_{01} = 4.11).

The final test trial type is a *Memory Check*. The current means suggest that infants may discriminate a novel pair from to a familiar pair: infants are looking 25.63 s (SD = 21.57) at the novel pair compared to 16.65 s (SD = 16.85) at the familiar pair. However, the Bayesian t-test comparing the novel relation to the familiar one for the *Memory Check* trial reveal that there is currently only weak evidence (.3 < *BF* < 3) for this effect, BF_{10} = 1.72. Furthermore, a binomial comparison shows that only 19 of the 35 infants are looking longer at the novel relation and driving this difference, *p* = .37. Again, differences between novel and familiar looking times are more pronounced in the *same* condition than the *different* condition (see Table 3), but the binomial comparisons are similar in both conditions: 9 of 16 infants in the *same* condition looked longer at the novel relation, *p* = .4, and 10 of 19 infants did so in the *different* condition, *p* = .50. Additionally, Bayesian t-tests show that there is currently only weak evidence that infants are generalizing in the *same* (BF_{10} = 0.51) and the *different* condition (BF_{10} = 1.63).

Habituation patterns

Finally, we examined looking time during the habituation phase. Because it is easier to conceptualize a habituation pair progression when both the objects are the *same*, we anticipated that there might be differences in the habituation decline between conditions, commensurate to the ease of learning the *same* progression compared to the *different* one. On average, infants looked 42.94 s (SD = 21.25) on their first habituation trial and 20.32 s (SD = 16.73) on their last trial (see Figure 7). A Bayesian ANOVA revealed that there was 3.23 times the evidence for an effect of Relation only (BF_{10} = 8.84e+7) compared to the evidence for an interaction with Relation and Condition (BF_{10} = 2.74e+7).

The averages reflected the prediction that *same* was easier, in terms of total looking time during habituation, (M_{same} = 253.29 s; $M_{different}$ = 349.52 s), but there was only weak evidence of this, BF_{10} = 1.75. There was also a difference in the average number of trials it took to reach the habituation criterion, (M_{same} = 7.56 s; $M_{different}$ = 8.60), and there was moderate evidence to support this, BF_{10} = 3.78. This *same* advantage was also reflected on whether infants reached the habituation criterion, with 11 of 16 infants habituating in the *same* condition, p = .11, but only 6 of 19 habituating in the *different* condition, p = .97.

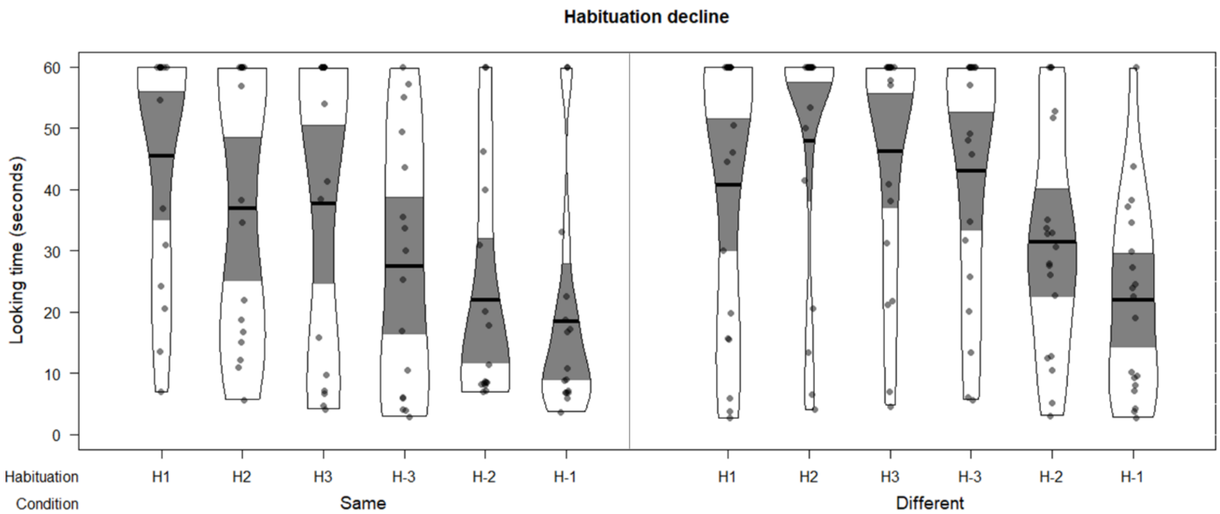

Figure 7. Looking decline over the course of 6 to 9 habituation trials, separated by habituation condition. Because the number of habituation trials was infant-controlled, the graphs represent the first three (H1, H2, H3) and the last three trials (H-3, H-2, H-1). The thick central line in each box is the mean, and the upper and lower shaded portions represent the 95% Confidence Intervals (CIs) for this mean (i.e., there is a 95% probability that the true population mean falls within this interval). Dots indicate the raw data points. The width of the bean indicates the density of the data distribution at a looking time value.

Discussion

The current evidence suggests that infants are not generalizing to neither *Near Transfer* nor *Far Transfer* trials. However, there is weak evidence that infants are discriminating relations in the *Memory Check*. If these trends are born out with the full sample, then it will indicate that increasing the similarity of the habituation set did not increase the opportunity for infants to align the relations across six exemplars. We had also predicted that the *same* progression might be easier to learn than the *different* progression. While the means reflect this possibility in all test types, the binomial comparisons and Bayesian t-test results argue that relational learning did not depend on the habituation condition. Ultimately these preliminary findings suggest that six unique exemplars are challenging for infants to relations align at three months, even when there is a progression of objects that are highly similar to each other.

Currently, there is moderate evidence for the null hypothesis in the *Far Transfer* trial type, but only weak evidence for the null in the *Near Transfer* trial type. An alternate possibility is that infants will generalize to *Near Transfer* but not to *Far Transfer* trials. If infants can demonstrate *Near Transfer*, this will show that three month-olds can benefit from multiple comparisons, but that they need more similar objects to be able to encode them within this time frame. By not demonstrating *Far Transfer,* it would show that these relational concepts begin as narrow and context-specific abstractions.

General Discussion

The results of Experiment 1 and the preliminary results of Experiment 2 suggest that it is incredibly challenging for infants to align *same* and *different* relations across six unique exemplars in the short space of habituation. Comparing these results to the similar findings from the six-exemplar condition of Anderson et al. (2018) reveals no major differences. Infants did not show significant improvement when we attempted to progressively align the pairs in habituation (Experiment 1 & 2) or even when we dramatically increased the level of similarity across all of the habituation pairs (Experiment 2).Thus these experiments show that even providing supports for alignment does not help young 3-month-old infants overcome allure of new objects.

Despite these three (presumptive) failures with 3-month-olds habituated to six exemplars, we know from previous experiments that this age group can successfully learn *same-different* relations (Anderson et al, 2018; Anderson et al., in prep). However, these successes 3-month-olds have only occurred in cases where there are a small number of exemplars (two or even just one) that are repeated across habituation trials. Thus we return to the idea that "less is more" when it comes to facilitating a first relational alignment. A remaining question is whether this means infants benefit from fewer exemplars because there is more repetition or whether they benefit from it because there is less object variability across the set.

Because infants could not align the relations in these experiment, we cannot be sure of whether progressive abstraction is a component of infant relational learning. If infants can make progressive abstractions, then these findings suggest that there are limits to this process. One factor may be that the habituation phase (which lasts ten minutes at most) is not a long enough period to permit both an initial abstraction and further comparisons. Future studies might consider longer periods over which this process could be implemented. For example, an experiment could habituate infants to fewer exemplars in a first session to permit an initial alignment and abstraction and then habituate the same infants to a greater number of exemplars during a second session.

Another factor is that, with the age group we tested, other processes might interact with infants' relational learning. In Anderson et al. (in prep) we explored how a panoply of encoding-related processes (acuity, scanning and attention, and short-term memory capacity) are still undergoing dramatic improvements in the first months, leading to recognition failures that could interact with relational learning. These recognition failures highlight the fact that the level of variability we present 3-month-olds with might not match the level of variability these infants perceive. While this artificial source of variability could increase the breadth of generalization from very close comparisons, it could also make it harder for young infants to align even six highly similar exemplars.

Questions remain as to how infants derive broad relational concepts from narrow initial abstractions. Ultimately, these results reiterate the challenges of detecting relations compared to perceptual features. In the face of what a precarious process it is to make the initial alignment, it is that much more amazing that we find evidence of relational learning in infancy and that relational ability becomes so characteristic of human thinking and reasoning.

References

Addyman, C., & Mareschal, D. (2010). The perceptual origins of the abstract same/different concept in human infants. *Animal Cognition*, *13*(6), 817–833. https://doi.org/10.1007/s10071-010-0330-0

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Interrelations Between Non-Linguistic and Linguistic Representations of Cognition and Action in Development*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Anderson, E. M., Chang, Y.-J., Hespos, S., & Gentner, D. (2018a). Comparison within pairs promotes analogical abstraction in three-month-olds. *Cognition*, *176*, 74–86.

Anderson, E. M., Chang, Y.-J., Hespos, S., & Gentner, D. (2018b). Comparison within pairs promotes analogical abstraction in three-month-olds. *Cognition*, *176*, 74–86. https://doi.org/10.1016/j.cognition.2018.03.008

Anderson, E. M., Hespos, S. J., & Gentner, D. (in preparation). *When one exemplar behaves like many*.

Arterberry, M. E., & Kellman, P. J. (2016). *Development of perception in infancy: The cradle of knowledge revisited.* Oxford University Press.

Asmuth, J. A., & Gentner, D. (2005). *Context Sensitivity of Relational Nouns*. http://www.escholarship.org/uc/item/8p9181j1

Baillargeon, R., & DeVos, J. (1991). Object Permanence in Young Infants: Further Evidence. *Child Development*, *62*(6), 1227–1246. https://doi.org/10.1111/j.1467-8624.1991.tb01602.x

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.

Begolli, K., Richland, L. E., & Jaeggi, S. M. (2015). The Role of Executive Functions for Structure-Mapping in Mathematics. *CogSci*.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9).

Bergelson, E., & Swingley, D. (2013). The Acquisition of Abstract Words by Young Infants. *Cognition*, *127*(3), 391–397.

Best, C. A., Robinson, C. W., & Sloutsky, V. M. (2010). *The effect of labels on visual attention: An eye tracking study*.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., Pêcheux, M.-G., Ruel, J., Venuti, P., & Vyt, A. (2004). Cross-Linguistic Analysis of Vocabulary in Young Children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115–1139. https://doi.org/10.1111/j.1467-8624.2004.00729.x

Bornstein, M. H., Pêcheux, M.-G., & Lécuyer, R. (1988). Visual habituation in human infants: Development and rearing circumstances. *Psychological Research*, *50*(2), 130–133. https://doi.org/10.1007/BF00309213

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual Statistical Learning in the Newborn Infant. *Cognition*, *121*(1), 127–132.

Casasola, M. (2005a). Can Language Do the Driving? The Effect of Linguistic Input on Infants' Categorization of Support Spatial Relations. *Developmental Psychology*, *41*(1), 183–192. https://doi.org/10.1037/0012-1649.41.1.183

Casasola, M. (2005b). When Less Is More: How Infants Learn to Form an Abstract Categorical Representation of Support. *Child Development*, *76*(1), 279–290.

Chang, Y.-J., Wang, S.-W., & Way, B. (2018). *BLT: Experimental program for online coding babies' looking time*.

Christie, S., & Gentner, D. (2010). Where Hypotheses Come from: Learning New Relations by Structural Alignment. *Journal of Cognition and Development*, *11*(3), 356–373.

Christie, S., & Gentner, D. (2014). Language Helps Children Succeed on a Classic Analogy Task. *Cognitive Science*, *38*(2), 383–397.

Christie, S., Gentner, D., Call, J., & Haun, D. B. M. (2016). Sensitivity to Relational Similarity and Object Similarity in Apes and Children. *Current Biology*, *26*(4), 531–535. https://doi.org/10.1016/j.cub.2015.12.054

Clark, E. (1979). Building a vocabulary: Words for objects, actions and relations. In P. Fletcher & M. Garman (Eds.), *Language acquisition: Studies in first language development*. Cambridge University Press.

Cohen, M., Horowitz, T., & Wolfe, J. (2009). Auditory recognition memory is inferior to visual recognition memory. *Journal of Vision*, *9*(8), 568–568.

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review Psychology*, *52*, 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Colombo, J., & Mitchell, D. W. (2009). Infant visual habituation. *Special Issue: Neurobiology of Habituation*, *92*(2), 225–234. https://doi.org/10.1016/j.nlm.2008.06.002

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical Treatment of Looking-Time Data. *Developmental Psychology*, *52*(4), 521–536.

Fagot, J., & Thompson, R. K. R. (2011). Generalized Relational Matching by Guinea Baboons (Papio papio) in Two-by-Two-Item Analogy Problems. *Psychological Science*, *22*(10), 1304–1309. https://doi.org/10.1177/0956797611422916

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63. https://doi.org/10.1016/0004-3702(89)90077-5

Feigenson, L., & Carey, S. (2005). On the Limits of Infants' Quantification of Small Object Arrays. *Cognition*, *97*(3), 295–313.

Ferry, A. L., Hespos, S. J., & Gentner, D. (2015). Prelinguistic Relational Concepts: Investigating Analogical Processing in Infants. *Child Development*.

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.

Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (Macaca mulatta): Judgment of relations and relations-between-relations. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(1), 55–63. https://doi.org/10.1037/0097-7403.33.1.55

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2016). Wordbank: An open repository for developmental vocabulary data. Journal of Child Language. *Journal of Child Language*, 1–18.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Vol. 2. Language, thought and culture* (pp. 301–334). Erlbaum.

Gentner, D. (2003). Why we're so smart. In *Language in mind: Advances in the study of language and thought* (pp. 195–235). MIT Press.

Gentner, D. (2005). The Development of Relational Category Knowledge. *Building Object Categories in Developmental Time.*, 245–275.

Gentner, D., Anggoro, F. K., & Klibanoff, R. S. (2011). Structure Mapping and Relational Language Support Children's Learning of Relational Categories. *Child Development, 82*(4), 1173–1188.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge University Press.

Gentner, D., & Markman, A. B. (1997). Structure Mapping in Analogy and Similarity. *American Psychologist*, *52*(1), 45–56.

Gentner, D., & Medina, J. (1998). Similarity and the Development of Rules. *Cognition*, *65*, 263–97.

Gentner, D., & Namy, L. L. (1999). Comparison in the Development of Categories. *Cognitive Development*, *14*(4), 487–513.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In *Perspectives on language and thought: Interrelations in development* (pp. 225–277). Cambridge University Press.

Gentner, D., Rattermann, M. J., Markman, A., & Kotovsky, L. (1995). Two forces in the development of relational similarity. *Developing Cognitive Competence: New Approaches to Process Modeling.*, 263–313.

Gentner, D., & Toupin, C. (1986). Systematicity and Surface Similarity in the Development of Analogy. *Cognitive Science*, *10*(3), 277–300.

Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*(2), 362–366.

Gerken, L., & Bollt, A. (2008). Three Exemplars Allow at Least Some Linguistic Generalizations: Implications for Generalization Mechanisms and Constraints. *Language Learning and Development*, *4*(3), 228–248.

Gerken, L., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants Consider Possible Bases of Generalization for a Single Input Example. *Developmental Science*, *18*(1), 80–89.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience*, *24*(3), 564–574. PubMed. https://doi.org/10.1162/jocn_a_00157

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, *105*(37), 14222. https://doi.org/10.1073/pnas.0806530105

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), 1–38.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature*, *410*(6831), 930–933. https://doi.org/10.1038/35073582

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431–436. https://doi.org/10.1111/1467-9280.00476

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization.(PSYCHOLOGICAL AND COGNITIVE SCIENCES)(Author abstract). *Proceedings of the National Academy of Sciences of the United States*, *107*(20).

Hochmann, J.-R., Carey, S., & Mehler, J. (2018). Infants learn a rule predicated on the relation same but fail to simultaneously learn a rule predicated on the relation different. *Cognition*, *177*, 49–57.

Hochmann, J.-R., Mody, S., & Carey, S. (2016). Infants' representations of same and different in match- and non-match-to-sample. *Cognitive Psychology*, *86*, 87–111.

Hochmann, J.-R., Tuerk, A. S., Sanborn, S., Zhu, R., & Carey, S. (2017). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2017.11.001

Holyoak, K. J., & Thagard, P. (1995). *Mental Leaps: Analogy in Creative Thought*. The MIT Press.

Hoyos, C., Shao, R., & Gentner, D. (2016). *The paradox of relational development: Could language learning be (temporarily) harmful?* [Proceedings of the 38th Annual Conference of the Cognitive Science Society]. Cognitive Science Society, Austin, Tx.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., & Slemmer, J. A. (2010). Abstract Rule Learning for Visual Sequences in 8- and 11-Month-Olds. *Infancy*, *14*(1), 2–18. https://doi.org/10.1080/15250000802569611

Káldy, Z., & Leslie, A. M. (2005). A memory span of one? Object identification in 6.5-month-old infants. *Cognition*, *97*(2), 153–177. https://doi.org/10.1016/j.cognition.2004.09.009

Kandaswamy, S., Forbus, K., & Gentner, D. (2014). Modeling Learning via Progressive Alignment using Interim Generalizations. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, 2471–2476. https://www.scholars.northwestern.edu/en/publications/modeling-learning-via-progressive-alignment-using-interim-general

Katz, J. S., & Wright, A. A. (2006). Same/different abstract-concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(1), 80–86. https://doi.org/10.1037/0097-7403.32.1.80

Keil, F. C., & Batterman, N. (1984). A characteristic-to-defining shift in the development of word meaning. *Journal of Verbal Learning & Verbal Behavior*, *23*(2), 221–236. https://doi.org/10.1016/S0022-5371(84)90148-8

Kibbe, M. M., & Feigenson, L. (2016). Infants use temporal regularities to chunk objects in memory. *Cognition*, *146*, 251–263.

Kibbe, M. M., & Leslie, A. M. (2013). What's the object of object working memory in infancy? Unraveling 'what' and 'how many.' *Cognitive Psychology*, *66*(4), 380–404.

Kibbe, M. M., & Leslie, A. M. (2016). The ring that does not bind: Topological class in infants' working memory for objects. *Cognitive Development*, *38*, 1–9. https://doi.org/10.1016/j.cogdev.2015.12.001

Kibbe, M. M., & Leslie, A. M. (2019). Conceptually Rich, Perceptually Sparse: Object Representations in 6-Month-Old Infants' Working Memory. *Psychological Science*, *30*(3), 362–375. Scopus. https://doi.org/10.1177/0956797618817754

Kotovsky, L., & Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child Development*, *67*(6), 2797–2822.

Kuehne, S. E., Forbus, K. D., & Gentner, D. (2000). *SEQL: Category learning as progressive abstraction using structure mapping*.

LaTourrette, A., & Waxman, S. R. (submitted). *The power of a name: Naming Systematically alters infants' representations of objects*.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*(4), 315–353.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the Relation: Fewer Exemplars Facilitate Children's Initial Verb Learning and Extension. *Developmental Science*, *11*(4), 628–634.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule Learning by Seven-Month-Old Infants. *Science*, *283*(5398), 77. https://doi.org/10.1126/science.283.5398.77

Mareschal, D., & Johnson, M. H. (2003). The "what" and "where" of object representations in infancy. *Cognition*, *88*(3), 259–276.

Markman, A. B., & Gentner, D. (1993). Structural Alignment during Similarity Comparisons. *Cognitive Psychology*, *25*(4), 431–67.

Namy, L. L., & Gentner, D. (2002). Making a Silk Purse Out of Two Sow's Ears: Young Children's Use of Comparison in Category Learning. *Journal of Experimental Psychology: General*, *131*(1), 5–15.

Needham, A., Dueker, G., & Lockhead, G. (2005). Infants' formation and use of categories to segregate objects. *Cognition*, *94*(3), 215–240. https://doi.org/10.1016/j.cognition.2004.02.002

Norcia, A. M., & Tyler, C. W. (1985). Spatial frequency sweep VEP: Visual acuity during the first year of life. *Vision Research*, *25*(10), 1399–1408. https://doi.org/10.1016/0042-6989(85)90217-2

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology*, *4*. https://doaj.org/article/d4d6c27355aa48799a0ff41ceea6272d

Oakes, L. M., & Rakison, D. H. (2020). *Developmental cascades: Building the infant mind*. /z-wcorg/.

Paik, J. H., & Mix, K. S. (2006). Preschoolers' Use of Surface Similarity in Object Comparisons: Taking Context into Account. *Journal of Experimental Child Psychology*, *95*(3), 194–214.

Pickron, C. B., Iyer, A., Fava, E., & Scott, L. S. (2018). Learning to individuate: The specificity of labels differentially impacts infant visual attention. *Child Development*, *89*(3), 698–710. https://doi.org/10.1111/cdev.13004

Premack, D. (1983). Animal Cognition. *Annual Review of Psychology*, *34*(1), 351–362. https://doi.org/10.1146/annurev.ps.34.020183.002031

Quinn, P. C., & Bhatt, R. S. (2005). Learning Perceptual Organization in Infancy. *Psychological Science*, *16*(7), 511–515. https://doi.org/10.1111/j.0956-7976.2005.01567.x

Rattermann, M. J., & Gentner, D. (1998). The effect of language on similarity: The use of relational labels improves young children's performance in a mapping task. In *Advances in analogy research: Integration of theory & data from the cognitive, computational, and neural sciences.* (pp. 274–282). New Bulgarian University.

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2017). Supporting Mathematical Discussions: The Roles of Comparison and Cognitive Load. *Educational Psychology Review.*, *29*(1), 41–53.

Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, *35*(1), 28–43. https://doi.org/10.1016/j.cedpsych.2009.09.001

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's Development of Analogical Reasoning: Insights from Scene Analogy Problems. *Journal of Experimental Child Psychology*, *94*(3), 249–273.

Robinson, J. S. (1955). The sameness-difference discrimination problem in chimpanzee. *Journal of Comparative and Physiological Psychology*, *48*(3), 195–197. https://doi.org/10.1037/h0042463

Rosch, E. (1978). Principles of categorization. In *Cognition and categorization*. L. Erlbaum Associates ; distributed by Halsted Press.

Ross-Sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The Development of Visual Short-Term Memory Capacity in Infants. *Child Development*, *74*(6), 1807–1822.

Simms, N. K., Frausel, R. R., & Richland, L. E. (2018). Working memory predicts children's analogical reasoning. *Journal of Experimental Child Psychology*, *166*, 160–177. https://doi.org/10.1016/j.jecp.2017.08.005

Simon, T. J., Hespos, S. J., & Rochat, P. (1995). Do infants understand simple arithmetic? A replication of Wynn (1992). *Cognitive Development*, *10*(2), 253–269. https://doi.org/10.1016/0885-2014(95)90011-X

Son, J. Y., Smith, L. B., Goldstone, R. L., & Eleslie, M. (2012). The Importance of Being Interpreted: Grounded Words and Children's Relational Reasoning. *Frontiers in Psychology*, *3*. https://doaj.org/article/2c2530ad8e4f44358161fc503f4d14cd

Stavans, M., Lin, Y., Wu, D., & Baillargeon, R. (2019). Catastrophic Individuation Failures in Infancy: A New Model and Predictions. *Psychological Review*, *126*(2), 196–225.

Wasserman, E. A., & Young, M. E. (2010). Same–different discrimination: The keel and backbone of thought and reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, *36*(1), 3–22. https://doi.org/10.1037/a0016327

Waxman, S. R., & Booth, A. (2003). The origins and evolution of links between word learning and conceptual organization: New evidence from 11-month-olds. *Developmental Science*, *6*(2), 128–135.

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

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*(2), 125–166.

Xu, F., & Tenenbaum, J. B. (2007). Word Learning as Bayesian Inference. *Psychological Review*, *114*(2), 245–272.