The Less is More Paradox in Relational Learning

Conference Paper · July 2023

4 authors:

Erin M Anderson
Indiana University Bloomington

Apoorva Shivaram
Northwestern University

Susan J Hespos
Northwestern University

Dedre Gentner
Northwestern University

The Less is More Paradox in Relational Learning

Erin M. Anderson (ema1@iu.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Apoorva Shivaram (ashivaram@u.northwestern.edu)
Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

Susan Hespos (s.hespos@westernsydney.edu.au)
The MARCS Institute for Brain, Behaviour, and Development, 160 Hawkesbury Rd,
Westmead NSW 2145 Australia

Dedre Gentner (gentner@u.northwestern.edu)
Department of Psychology, 2029 Sheridan Road
Evanston, IL 60208 USA

Abstract

The ability to generalize previous knowledge to new contexts is a key aspect of human cognition and relational learning. A well-known learning maxim is that *breadth of training predicts breadth of transfer.* When examples vary in their surface features, this provides evidence that only the common relational structure is relevant. However, there is some evidence suggesting that the above maxim may not apply well in early relational learning. Here, we present a further test whether the maxim holds for young infants. We find that 3-month-old infants perform better with a narrow, perceptually similar training set than with a broad, perceptually variable set. We argue that lower-level perceptual similarities can prompt comparison processes that facilitate relational abstraction. These findings cohere with research arguing relational learning depends on relational alignment.

Keywords: relational learning; similarity; generalization; infant learning; comparison

Introduction

Few topics in cognitive science have been more intensely researched than learning and transfer. Within this broad arena, relational learning and generalization has emerged as a key challenge for researchers in human cognition, development, and computer science. Relational learning is the ability to recognize the structures between elements despite perceptual differences, allowing the learner to make new inferences as they transfer this structure to novel contexts (e.g., comparing passengers and goods being transported on highways to oxygen and nutrients being transported in blood vessels). This process is crucial to higher-order cognition, with clear roles in learning language, math, and in problem-solving (Bettoni et al., 2020; Gentner, 2003; Gentner & Namy, 2006; Richland & Begoli, 2016).

What makes this process challenging is that the relations between elements are often less obvious than the perceptual features surrounding them (for evidence of these challenges, see Gentner & Namy, 1999; Gentner & Toupin, 1986; Paik & Mix, 2006). One approach to overcome the pull of perceptual features is to train learners on highly dissimilar examples. According to this approach, when a relational structure is seen across a variety of contexts, there will be increasing evidence that the perceptual elements local to each example are not relevant to the relational structure. This increased abstraction in turn should allow greater generalization. In other words, the *breadth of training should predict the breadth of transfer*. Evidence supporting this approach has been seen in adults (see Raviv et al., 2022 for a review); and even in young children (Walker et al., 2018). This is notable because young children are particularly susceptible to focusing on perceptual similarity (Gentner et al., 2011; Gentner & Toupin, 1986; Richland et al., 2006).

However, there is some evidence showing the opposite effect: these studies report that learners only succeed in learning relational concepts when examples are less variable. For example, one study trained 3-month-olds on *same* and *different* relations, with either six examples or two examples that repeated (Anderson et al., 2018). By virtue of having more examples, the six-example set had more variability, but only infants who saw the two-example set succeeded in generalizing at test. Other studies of infants and toddlers have found similar results, not only with fewer, repeated examples, but also when examples are more similar to each other (Bulf et al., 2011; Casasola, 2005; Childers et al., 2016; Maguire et al., 2008; Oshima-Takane et al., 2011; Scott & Fisher, 2012; but see Gómez, 2002 for the reverse).

Though seemingly counterintuitive, these findings point to a key tenet of relational learning: learners must be able to align the relational structure across examples, and this might not occur when examples are more variable from each other. For adults or older children, explicit instruction can bridge the gap between highly dissimilar examples. For example, labeling relations facilitates relational learning for children as young as 3 years (Christie & Gentner, 2014; Du et al., 2018;

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

Gentner & Rattermann, 1981). Labeling relations may not be as effective for infants, though (Anderson et al., 2022). For these learners, increased similarity might be a more reliable route to comparison. Gentner and Hoyos (2017) identified two ways in which high similarity is particularly beneficial to beginning learners: first, it prompts the initiation of a comparison process; and equally important, it supports *structural alignment*, where shared relational structures become more apparent when the examples are put into alignment with each other, without which no relational abstraction would occur (Forbus et al., 2017; Gentner & Markman, 1997). Because of this, close comparisons – where the perceptual similarities between examples support the relational similarity – can be extremely useful for initiating relational learning (Gentner et al., 2011; Loewenstein & Gentner, 2001; Kotovsky & Gentner, 1996).

In the current studies, we demonstrate that increasing the likelihood of comparison – via increasing the visual similarity of trained relations – has a clear impact on relational generalization in infants. Specifically, we show that 3-month-olds can generalize from limited experience with many examples (6), but only if visual similarity between pairs is increased to invite comparison between them.

Experiment 1 Methods

To test relational generalization in 3-month-olds, we used a habituation/dishabituation looking time paradigm modeled on earlier infant studies with *same-different* relations (Anderson et al., 2018; Ferry et al., 2015). As in these studies, there were two between-subjects habituation conditions: half of the infants tested were habituated to a series of object pairs where the relation between objects was always *same* and half saw pairs that were always *different*. These progressions took place over the course of 6 to 9 habituation trials, providing experience with the relation and the opportunity for infant looking to decline. To increase the likelihood of spontaneously comparing the pairs, the six examples in this study were presented in a fixed progression designed to have some similarities between consecutive pairs (see Figure 1). For example, 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.

After habituation trials, infants saw test trials, featuring the pairs of objects in 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*). The full procedure included *New* test trials featuring previously unseen objects, as well as test types featuring objects seen before habituation. For the purpose of this study, we focus on *New* trials, targeting the question of generalization. If infants are able to align across the set and form a relational abstraction, then they should perceive the *same*-relation test as a continued example of the *same* relational pattern, and the novel test pair as a violation of the pattern. Thus, they should look longer at the novel relation than at the familiar relation in the *New* trials.

Participants

The participants were 32 healthy, full-term infants (12 female, mean age = 3 months, 13 days, range = 2 months to 4 months 13 days). Based on a post-hoc power analysis of Anderson et al. (2018)'s main effect of relation (novel vs. familiar, $\eta_p^2 = .17$)[1], this was a sufficient sample size to reject the null hypothesis ($\alpha = .95$). 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 sample: four for fussiness (judged as fussy or crying by two independent coders for more than half the test trials), three because they 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). This exclusion rate (27% of participants) is consistent with the 31% reported in Anderson et al. (2018).

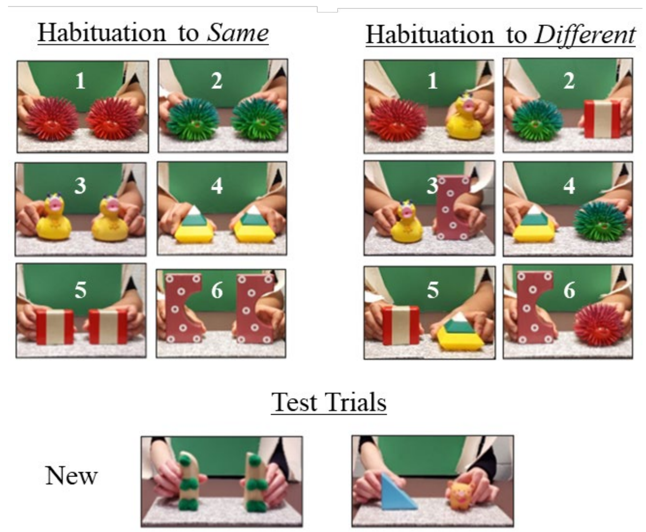


Figure 1: Images of the stimuli used in Experiment 1 habituation (training) and test trials. Infants were habituated to either the *same* or *different* sequence, but all participants saw the same test trials.

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 self-reported race of sample was 63% white, 14% multiracial, 7% African-American, 5% Asian, 2% American Indian, and 9% unreported. The ethnicity of the sample was 81% non-

[1] The post-hoc analysis used GPOWER (Faul et al., 2009) where the statistical test was set to ANOVA: Repeated measures, within factors, the effect size was set to 0.45, calculated from $\eta_p^2 = .17$, number of groups was 2, number of measurements was 2 and correlation among repeated measures was set to 0.34, calculated from the Anderson et al. (2018) supplemental dataset.

Hispanic, and 93% of primary caretakers in the sample had a college degree or higher.

Apparatus & Procedure

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 three-dimensional objects (see Figure 1). 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. The number of habituation trials was infant-controlled (see Coding section for the criterion), ranging from 6 to 9 trials. Test trials were presented in the same motion pattern as in the habituation trials, and the infants' looking time on each trial was recorded.

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 pressed 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. The habituation criterion was a looking time decline of 50% or more from the first three habituation trials to the last three trials. Because of this, infants saw a minimum of 6 habituation trials. If infants completed 9 habituation trials without reaching criterion, they continued to test. After each 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 unaware of the experimental 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. Our looking time data significantly deviated from a normal distribution per the Shapiro-Wilks test. Therefore, we performed analyses on log-transformed data, following recommendations outlined by Csibra et al. (2016). For ease of interpretation, however, all summary statistics, coefficients, and plot axes are reported or shown in non-transformed scales.

Experiment 1 Results

The main question was whether infants would generalize the *same* or *different* relation from training (habituation) to test, looking longer at the novel relation even when it was instantiated with new objects. To analyze this, we used a mixed effects model that included fixed effects for condition (habituation to *same* or *different*), test relation (novel or familiar), and the interaction term, as well as including a random intercept for each subject. Infants showed similar looking times for both the familiar relation ($M = 29.15$ s, $SD = 18.51$) and the novel relation ($M = 28.56$ s, $SD = 20.07$); see Figure 2A. Our analysis confirmed that there was no effect of relation, $F(1, 27) = .66$, $p = 0.42$, with a $\beta = -1.12$ s ($SE = 1.26$) estimated difference in looking time for novel compared to familiar relations when controlling for variance from test type, condition, and subject intercepts.

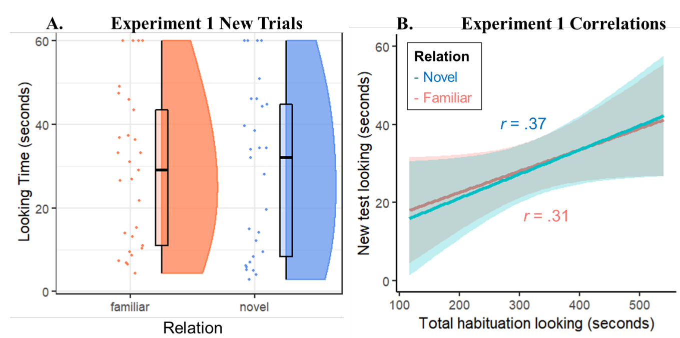


Figure 2: *A*. Raincloud plot showing looking time to familiar vs. novel relations in Experiment 1. Dots indicate individual looking times on each trial, the width of the colored curve corresponds with the density of the data at a given looking time, and the boxplot summarizes the upper and lower quartiles (top and bottom boxes) and the median of the data (thick middle line). *B*. Regression lines representing the linear relationship between total looking time during habituation (x-axis) and looking time at novel and familiar (blue & coral) relations in the *New* test trials (y-axis). The thick center of each line is the estimated test

looking time per habituation time, and the shaded area is the standard error.

Infants in the *different* condition had longer looking times at test, M = 31.42 s (SD = 21.37) than those in the *same* condition, M = 26.46 s (SD = 16.81), but the model showed no effect of condition, $F(1, 27) = .09$, $p = 0.77$, $\beta = -1.10$ s (SE = 1.41) estimated difference for the *same* compared to the *different* condition. Additionally, there was no interaction between condition and relation, $F(1, 27) < .001$, $p > .99$, $\beta <$ 1 s (SE = 1.35) additional difference in looking at the novel relation in the *same* condition.

Because infants differed in how much total exposure (in seconds) they received to the relation during the habituation period, we asked if there was a relationship between this and their generalization at test. If more accrued experience with the relational pairs during habituation predicts better generalization at test, then we would expect to see a positive relationship between looking time during habituation and looking time at the novel test relation. In contrast, because the increased experience should lead to recognition of the familiar relation and thus reduced looking to it at test, we might expect a negative relationship between habituation looking time and looking time at the familiar test relation. Instead, both novel and familiar looking times were weakly correlated with total habituation looking time (r = .37 for novel and r = .31 for familiar), suggesting that infants were longer- or shorter-lookers in general, both in training and test (see Figure 2B).

In sum, infants did not generalize from the six training examples in Experiment 1.

Experiment 1 Discussion

The results from this study show the same pattern as Experiment 1 in Anderson et al. (2018): 3-month-old infants failed to generalize from training on 6 examples over the course of 6 to 9 training trials. This occurred despite our attempts to facilitate comparison across training pairs by making consecutive pairs of stimuli more visually similar to each other. Instead, infants in both training conditions (*same* and *different*) looked equally between these test relations made of novel objects. In particular, the heightened looking at the familiar relation suggests that infants did not recognize the relation from training (aka habituation).

Further increasing the level of visual similarity between training examples should facilitate relational learning in 3-month-olds by creating close comparisons where the perceptual similarities between examples support the relational similarity. Therefore, in Experiment 2, we investigated whether increasing the visual similarity of the entire training set would facilitate relational abstraction. To do so, we used Greeble objects, which have a set of predictable features 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 all training pairs will be more similar to each other than those in Experiment 1.

Raising the overall level of similarity in the training set could come at a cost, though. Higher similarity is more likely to enable learners to compare and align the relation, but higher similarity comparisons also result in narrower generalization (e.g., Rosch, 1978; Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007). Therefore, it is possible that even if infants abstract the relation, they will not generalize to *New* items at test. Of course, another possibility is that infants' failure to discriminate novel from familiar relations in Experiment 1 resulted from inattention or lack of memory for the habituation materials. Therefore Experiment 2 included *Memory Check* test trials which featured an exact pair seen during the habituation period. This served as a manipulation check of whether infants were discriminating at all.

Experiment 2

The procedure of Experiment 2 was identical to Experiment 1 except that the habituation objects were highly constrained Greeble objects (see Figure 3). If the increased level of similarity between pairs allows 3-month-olds to align across six examples and generalize, then infants should look longer at the novel relation in the *New* trials that feature non-Greeble objects, as well as the *Memory Check* trials with Greebles. If infants fail to generalize, they should still look longer at the novel relation in the *Memory Check* trials, because they have habituated to the exact pair. that instantiates the familiar relation.

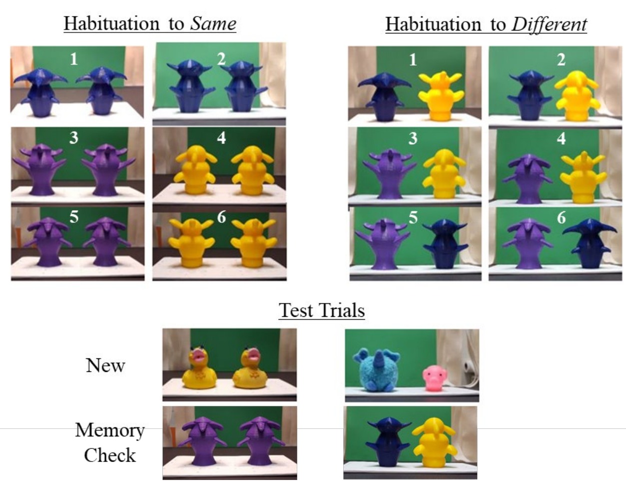


Figure 3: Images of the highly similar Greeble stimuli used in Experiment 2 habituation (training) trials and both non-Greeble and Greeble stimuli used in *New* and *Memory Check* test trials. Infants were habituated to either the *same* or *different* sequence, but all participants saw the same test trials.

Participants

The sample included 48 healthy, full-term infants (27 female, mean = 3 months, 22 days, range = 2 months 15 days to 4 months, 23 days). The sample size was increased compared to Experiment 1 to adjust for the design, because we were now not only looking for a main effect of relation, but potentially an interaction if infants generalized on *Memory Check* but not *New* trials. Another 15 infants were 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 (24% exclusion, similar to Experiment 1 and Anderson et al. (2018)). Recruitment and compensation were the same as in Experiment 1. The self-reported race of the sample was similar to that of Experiment 1: 67% white, 12% multiracial, 17% African-American, and 5% Asian. The ethnicity of the sample was 88% non-Hispanic and 93% of the primary caretakers had a college degree or higher.

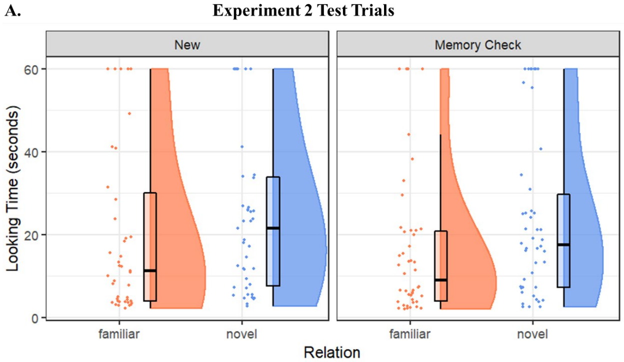
Apparatus

All of the habituation objects were 3-D printed adaptations of Gauthier & Tarr (1997)'s Greeble objects. The Greeble objects 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).

Experiment 2 Results & Discussion

The main question was whether infants would generalize the *same* or *different* relation in test trials, looking longer at the novel relation. We used a mixed effects model that included fixed effects for condition (training in *same* or *different*), and test relation (novel or familiar). The model also included a term for test type (*Memory Check* vs. *New*), interaction terms, and a random intercept for each subject. Across *Memory Check* and *New* trials, infants looked longer at the novel relation (M = 24.04 s, SD = 19.66) compared to the familiar relation (M = 18.07 s, SD = 18.97) (see Figure 4A). Our analysis likewise indicated a main effect of relation, $F(1, 41) = 12.03$, $p < .001$, showing a β = 1.55 s (SE = 1.29) estimated difference in looking time for novel compared to familiar relations (controlling for variance from test type, condition and subject intercepts).

A.



B.

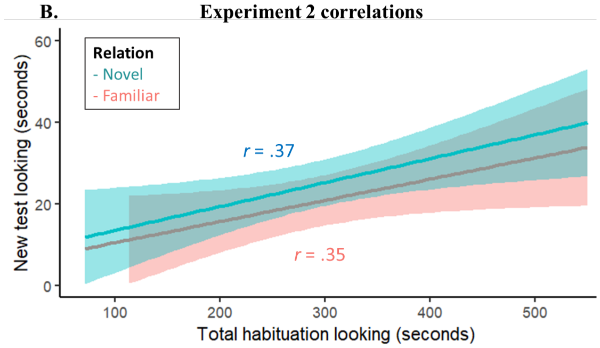


Figure 4: *A*. Raincloud plots showing looking time at novel and familiar relations in Experiment 2 test trials (New and Memory Check). *B*. Regression lines representing the linear relationship between total looking time during habituation (x-axis) and looking time at novel and familiar (blue & coral) relations in the *New* test trials (y-axis). See Figure 2 caption for a full explanation of the plot elements.

Longer looking at the novel relation occurred on both the *New* trials (*M-novel* = 24.66 s, SD = 19.85; *M-familiar* = 20.37 s, SD = 20.87), and the *Memory Check* trials (*M-novel* = 23.46 s, SD = 19.86; *M-familiar* = 15.91 s, SD = 16.96). In line with this, there was not a significant interaction between relation and test type, $F(1, 41) < 1$, $p = .69$.

There was no main effect of test type, F (1, 41) = 3.07, p = .09, B = 1.38 (SE = 1.11), $F(1, 41) = 2.39$, $p = .12$, nor of condition, with β = -1.35 (SE = 1.35) estimated difference in looking for *same* compared to the *different* condition.

Additionally, there was no interaction between condition and relation, F (1, 41) < 1, $p = .94$, β = 1.07 (SE = 1.41) for novel relations in the *same* condition, and no interaction between condition and test type, F (1, 41) < 1, $p = .43$, β = -.1.14 (SE = 1.41) for *New* trials in the *same* condition. There was no interaction between the three variables, F (1, 41) < 1, $p = .83$, β = -1.12 (SE = 1.66) for the novel relation in the *New* test in the *same* condition.

As in Experiment 1, there were positive weak correlations between the total habituation looking time and looking at test relations, both for the novel relation ($r = .37$) and the familiar relation ($r = .35$; see Figure 4B). That is, in general, infants who looked longer during training also looked longer at test.

Cross-Experiment Comparison

Turning to the key question, we next compare Experiments 1 & 2 to determine whether they differed in degree of generalization on the *New* test trials.

Generalization For the cross-experiment analysis, the mixed effects model included a fixed effects term for Experiment (1 or 2) as well as those for condition and relation, and interaction terms between these three factors. The model also included random intercepts between subjects. In general, there were longer test looking times in Experiment 1 (M = 28.85 s, SD = 19.14) than in Experiment 2 (M = 22.52 s, SD = 20.27), and the analysis showed a trending main effect of experiment, $F(1, 68)$, = 3.56, p = 0.06, β = -1.59 s (SE = 1.4) for Experiment 2 compared to Experiment 1. There was a significant interaction of relation and experiment, $F(1, 68) = 4.73$, $p = 0.03$, with a $\beta = 1.69$ s ($SE = 1.4$) additional difference in looking at the novel relation in Experiment 2. This is in line with the previously reported differences in looking time to novel vs. familiar test relations in the high-similarity experiment (E2; see Figure 4), but not the lower-similarity experiment (E1; see Figure 2). There was no main effect of condition across experiments, $F(1, 68)$, = 1.73, p = 0.19, β = -1.53 s for *same* compared to *different*, nor of relation, $F(1, 68)$, = 1.22, $p = 0.27$, $\beta = 1.49$ ($SE = 1.25$) for novel compared to familiar.

Habituation Finally, we asked whether the high similarity training set (Experiment 2) would correspond with a faster learning curve, as measured by the decline in looking times over habituation trials. If so, this could indicate that the relational pairs had been easier to learn when they were more visually similar. An experiment by habituation trial effect was not significant, $F(5, 380) = 0.54$, $p = .75$ as both experiments showed similar declines; $\beta = -1.15$ s (SE = 1.04); see Figure 5. Across both experiments, looking times were longer on the first three habituation trials ($M = 45$ s, $SD = 19.30$) than on the last three trials ($M = 27.86$ s, $SD = 20.41$). In line with this, a mixed effects model showed a main effect of habituation trial, $F(5, 380) = 21.96$, $p < .001$, with an estimated $\beta = -1.15$ s (SE = 1.04) decrease in looking per subsequent habituation trial.

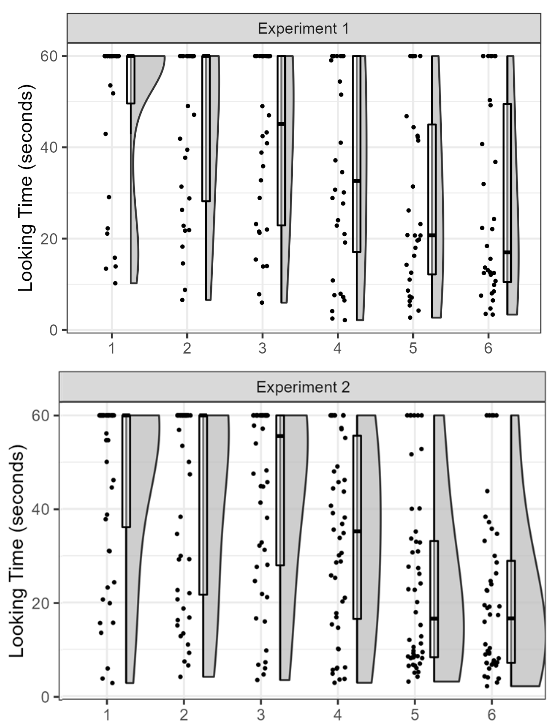


Figure 5: Raincloud plots showing the habituation looking time declines by trial number in Experiments 1 and 2. (See Figure 2 caption for a full explanation of plot elements.)

General Discussion

Infants showed generalization of *same* and *different* relations after seeing the high-similarity training set in Experiment 2, but not the more variable set in Experiment 1. This finding supports our predictions that the perceptual similarity of a set of examples can facilitate detection of their relational similarity. The differences in within-set similarity between experiments and resulting differences in generalization at test suggest that 1) high perceptual similarity may be more likely to initiate comparison and 2) highly similar instances of a relation may be easier to align and abstract. These findings provide converging evidence that 3-month-old infants are capable of relational learning, complementing the initial demonstration in Anderson et al. (2018). In addition, these data suggest that perceptual similarity helps foster comparison and generalization in this context. This leads us to amend the maxim that breadth of training predicts breadth of transfer. We suggest that for relational learning, breadth of *alignable* examples is what matters.

Successfully learning far generalizations from close comparisons has deep implications for development in a variety of domains. Experiment 2 may constitute an ideal learning environment, where learners saw highly similar presentations over a short period of time, but the result of better generalization from a narrower training set gives us insight into how infants might be able to learn in more complex learning conditions of everyday life. In home and daycare, infants are also likely to see certain sets of similar objects compose the same relations again and again (e.g., sippy cups *on* the table). While research must confirm how broad or limited this everyday "training data" is, finding of generalization from narrow training offers a proof of concept for how abstract learning could occur in the first months.

These results make a clear argument for how a training set that has both perceptual and relational similarity can help with relational learning, but it remains an open question as to why relational generalization was not limited. It is possible that the *New* test objects, which were also all animate beings, were more similar to the Greebles than in our conceptualization. However, Experiment 1 also included a range of animals in training but saw no generalization in test. There are also factors unique to 3-month-olds that may in fact support more abstract learning in the visual modality, including poorer acuity that creates a more schema-like experience (e.g., Vogelsang et al., 2018). Undoubtedly developmental factors such as these interact with comparison and alignment to shape the outcomes of relational learning and should be targeted by future work.

In sum, this work has implications for understanding learning and transfer in and out of the lab. By studying young infants, we get a sense of what prompts relational generalization when learners have no explicit instruction. Our study underscores two aspects that seem critical: first, training data where low-order perceptual similarity coincides with higher-order relational similarity, and second, a mechanism that can compare and align these.

Acknowledgments

This research was supported by the National Science Foundation grants [BCS-1423917, BCS-1729720] awarded to Susan J. Hespos and Dedre Gentner, by NSF grant [SBE-0541957] awarded to the Spatial Intelligence and Learning Center (SILC), by Office of Naval Research grant [N00014-16-1-2613] awarded to Dedre Gentner, and by a US Department of Education, Institute of Education Sciences Training Grant (Multidisciplinary Program in Education Science) [R305B140042] awarded to Erin Anderson. We are indebted to parents who agreed to have their infants participate in the research. We thank members of the Infant Cognition Lab for their help, especially Shelley Powers. All reported data are archived in the following OSF repository: https://osf.io/m6rju/?view_only=c4eb133915b1431882a477c0d8bcf994

References

Anderson, E. M., Chang, Y-J, Hespos, S.J., & Gentner, D. (2018). Comparison Within Pairs Promotes Analogical Abstraction in Three-Month-Olds. *Cognition, 176*, 74 – 86. DOI: 10.1016/j.cognition.2018.03.008

Anderson, E. M., Chang, Y-J, Hespos, S.J., & Gentner, D. (2022). No evidence for language benefits in infant relational learning. *Infant Behavior and Development, 66*, 101666. DOI: 10.1016/j.infbeh.2021.101666

Bettoni, R., Riva, V., Cantiani, C., Molteni, M., Macchi Cassia, V., & Bulf, H. (2020). Infants' Learning of Rule-Based Visual Sequences Predicts Language Outcome at 2 Years. *Frontiers in psychology*, *11*, 281. https://doi.org/10.3389/fpsyg.2020.00281

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. Cognition, 121(1), 127–132. https://doi.org/10.1016/j.cognition.2011.06.010

Casasola, M. (2005). When Less Is More: How Infants Learn to Form an Abstract Categorical Representation of Support. *Child Development, 76*(1), 279–290. https://doi.org/10.1111/j.1467-8624.2005.00844.x

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

Childers, J. B., Parrish, R., Olson, C. V., Burch, C., Fung, G., & McIntyre, K. (2016). Early Verb Learning: How Do Children Learn How to Compare Events?. *Journal of cognition and development : official journal of the Cognitive Development Society*, *17*(1), 41–66. https://doi.org/10.1080/15248372.2015.1042580

Christie, S., & Gentner, D. (2010). Where hypotheses come from: learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373.

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521–536.

Du, C., Miyazaki, Y., Cook, M., Papadopoulos, J., & Hao, Y. (2018). Relational Language Improves Preschool Children's Performance of Analogical Reasoning. *International journal of psychological studies, 10*, 91.

Ferry, A., Hespos, S.J., & Gentner, D. (2015). Prelinguistic Relational Concepts: Investigating Analogical Processing in Infants. *Child development, 86 5*, 1386-405.

Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, *41*, 1152-1201.

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought*. MIT Press.

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. https://doi.org/10.1111/j.1467-8624.2011.01599.x

Gentner, D., & Hoyos, C. (2017). Analogy and abstraction. *Topics in Cognitive Science, 9*(3), 672–693. https://doi.org/10.1111/tops.12278

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*(4), 487–513. https://doi.org/10.1016/S0885-2014(99)00016-7

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science, 15*(6), 297–301. https://doi.org/10.1111/j.1467-8721.2006.00456.x

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

Gentner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive science*, *10*(3), 277-300.

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

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*(6), 2797–2822.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development*, *2*(2), 189-219.

Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., & Poulin-Dubois, D. (2011). Early verb learning in 20-month-old Japanese-speaking children. *Journal of child language*, *38*(3), 455–484. https://doi.org/10.1017/S0305000910000127

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. https://doi.org/10.1016/j.jecp.2006.06.002
Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in cognitive sciences*, *26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007
Richland, L. E., & Begolli, K. N. (2016). Analogy and Higher Order Thinking: Learning Mathematics as an Example. Policy Insights from the Behavioral and Brain Sciences, 3(2), 160–168. https://doi.org/10.1177/2372732216629795
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. https://doi.org/10.1016/j.jecp.2006.02.002
Rosch, E. (1978). Principles of categorization. In *Cognition and categorization*. L. Erlbaum Associates; distributed by Halsted Press.
Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180. https://doi.org/10.1016/j.cognition.2011.10.010
Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *The Behavioral and brain sciences*, *24*(4), 629–791. https://doi.org/10.1017/s0140525x01000061
Vogelsang, L., Gilad-Gutnick, S., Ehrenberg, E., Yonas, A., Diamond, S., Held, R., & Sinha, P. (2018). Potential downside of high initial visual acuity. *PNAS*, *115*(44), 11333–11338. https://doi.org/10.1073/pnas.1800901115
Walker, C. M., Hubachek, S. Q., & Vendetti, M. S. (2018). Achieving abstraction: Generating far analogies promotes relational reasoning in children. *Developmental Psychology, 54*(10), 1833–1841. https://doi.org/10.1037/dev0000581
Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272.