When do Children Pass the Relational-Match-To-Sample Task?

Apoorva Shivaram (ashivaram@u.northwestern.edu)
Department of Psychology, Northwestern University, USA

Ruxue Shao (ruxueshao@sina.com)
Department of Psychology, Northwestern University, USA

Nina Simms (simms@northwestern.edu)
Department of Psychology, Northwestern University, USA

Susan J. Hespos (S.Hespos@westernsydney.edu.au)
Department of Psychology, Northwestern University, USA
MARCS Institute for Brain, Behaviour, & Development, Western Sydney University, Australia

Dedre Gentner (gentner@northwestern.edu)
Department of Psychology, Northwestern University, USA

Abstract

Relational ability—the ability to compare situations or ideas and discover common relations – is a key process in higher-order cognition that underlies transfer in learning and creative problem solving. For this reason, it has generated intense interest both among developmentalist and in cross-species comparative studies. The gold standard for evaluating relational ability is the Relational-Match-to-Sample (RMTS) task (Premack, 1983). Current work in cognitive development has produced inconsistent results as to when children are able to pass the RMTS, with Christie and Gentner (2014) finding earlier success than Hochmann et al. (2017) and Kroupin and Carey (2022). In this research, we attempt to resolve this issue. We first describe two studies that bear out and extend Christie and Gentner's (2014) findings. We then discuss factors that might explain the discrepancy between the findings.

Keywords: relational reasoning; spontaneous reasoning; development; same and different relations; alignment; complexity

Introduction

Relational ability—the ability to compare distinct situations or ideas and discover common relations—is a key component of higher-order cognition (Gentner, 2003, 2010; Penn et al., 2008). As relational ability has been argued to be central in learning science and mathematics (Goldwater & Schalk, 2016; Richland & Simms, 2015), it is critical to understand the development of this ability.

The gold standard for assessing children's relational ability is the Relational Match-to-Sample (RMTS) task, devised by Premack (1983). The RMTS is a triad task in which the participant is given a pair that instantiates either a *same* relation or a *different* relation (the standard) and must choose the alternative that instantiates the relation that is represented in the standard. For example, in a *same* triad, the participant is given AA [standard] and is asked to choose which of two alternatives is a better match for the standard: XX [relational match—correct] or YZ [non-relational match] (see Figure 1 for an example). In a *different* triad, if AB is the standard, the participant must choose YZ over XX. Success on this task requires the participants to encode the relation within each pair of objects and choose the alternative that shares the relation with the standard.

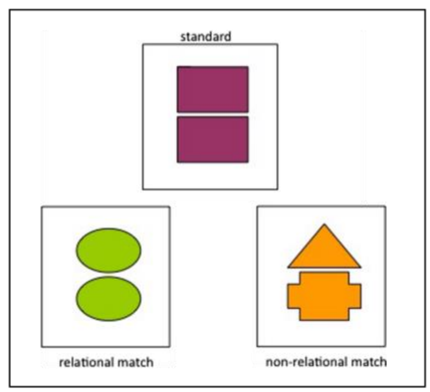

Figure 1: A sample triad of the *same*-only version of the RMTS task as used in Christie & Gentner (2014).

This task is ideal for developmental and cross-species work. It does not require specialized content knowledge – only an understanding of *same* and *different* relations. These relations are perceptually available and are fundamental relations both for concept formation and for everyday reasoning. Finally, *same-different* relations have been studied extensively across species, offering the possibility of phylogenetic as well as ontogenetic comparisons (e.g., Wasserman & Young, 2010).

Passing the RMTS task provides clear evidence of the ability to form and match relational representations. Apart from humans, only a few other species have demonstrated success on the RMTS task but only after extensive training prior to the task (for example, chimpanzees (Thompson et al., 1997), crows (Smirnova et al., 2015), orange-winged

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

amazons (Obozova et al., 2015), and baboons (Fagot & Thompson, 2011; Flemming et al., 2013)).

Because of its importance, there has been considerable interest in the key question of when young humans acquire the ability to succeed on this task spontaneously—that is, without training or feedback. To address this question, Christie and Gentner (2014, Experiment 1) gave 2.5-, 3.5-, and 4.5-year-old children the RMTS task as exemplified in Figure 1. Children saw a standard that depicted the relation of *same* (e.g., two squares). Then two alternatives—one depicting *same* and one depicting *different*—were placed below the standard, and the child was asked "Which of these two pictures is more like this one [standard]?" Critically, there were no training trials prior to the RMTS task, nor was any feedback given during or after the task. The findings revealed that only 4.5-year-olds performed above chance on this task; neither 2.5- nor 3.5-year-olds succeeded. Because children had received neither training nor feedback, Christie and Gentner concluded that by 4.5 years of age, children can spontaneously pass the RMTS.

However, in another investigation of children's performance on the RMTS task, Hochmann et al. (2017, Experiment S1 in Appendix A) found different results. They used the full RMTS task, with both *same* and *different* standards. Unique objects were used on every trial (Figure 2). The participants were 3- to 6-year-old children. At the outset, children were introduced to two puppets. Children were told that each puppet 'likes a certain type of card' and that they would get a sticker if they gave the puppet a card that it liked. Children completed eight test trials. On each trial, the experimenter first placed the standard card in front of the appropriate puppet. Then the two alternatives were placed in front of the standard, and the child was asked which one the puppet would like. They counterbalanced which puppet liked *same* cards, and which liked *different* cards across participants. Children were given feedback on each test trial—a sticker if they chose the correct relational match and verbal feedback if they chose incorrectly. Children's performance exceeded chance at 5 years and older, but not at 3 or 4 years.

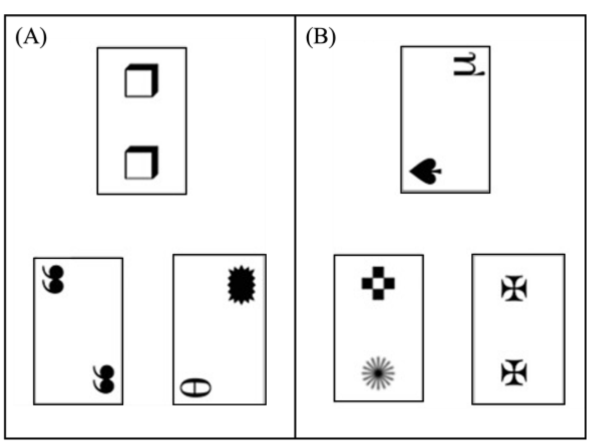

Figure 2: A sample triad of the (A) *same* and (B) *different* relations of the *mixed* RMTS task as used in Hochmann et al. (2017).

In a further study, children were given eight trials of training with feedback prior to the RMTS task, but still failed to succeed until 5 years of age (Hochmann et al., 2017; Experiment 2). Once again, the full RMTS task was used (i.e., with both *same* and *different* standards). Again, unique objects were used on every trial. The participants were 4- and 5-year-old children (mean ages were 4.51 and 5.53 years, respectively). On each training trial, the experimenter first placed the two alternatives before the child (one *same* and one *different*) and demonstrated that they 'do not go together because the cards are not alike'. Then children were presented with the standard and asked to choose 'which of these cards [pointing to the two alternatives] goes with this card [the standard]'. Children were given corrective feedback during the eight training trials. In the test, children were given eight test trials without feedback. Children's performance on the test trials exceeded chance at 5.5 years, but not at 4.5 years of age.

In Hochmann et al.'s (2017) Experiment S1, children were given corrective feedback on each trial. In Experiment 2, children received eight training trials with feedback prior to the test. Yet, these children required a full year more to master the task than did children in Christie and Gentner's (2014) study, in which neither training nor feedback was given. This is a striking discrepancy.

In discussing the difference between their findings and those of Christie and Gentner (2014), Hochmann et al. (2017, page 22) noted that Christie and Gentner (2014) had used the *same*-only version of the RMTS task, rather than the full classic version with both *same* and *different* relations as standards. As Hochmann et al. (2017) pointed out, "Christie and Gentner (2014) tested children on a XX-AA problem that could be solved by learning a single rule "choose same;" i.e., this could be a discrimination learning task…". It is quite possible that detecting *same* is less challenging than detecting *different*. It has been argued that *same* is an elemental relation, while *different* could be represented as 'not *same*' (Clark & Chase, 1972; Hochmann et al., 2016; Smith et al., 2008). For both these reasons, the success of 4.5-year-olds in Christie and Gentner's (2014) *same*-only version of the RMTS task does not justify the conclusion that this age group can pass the full RMTS task.

To address this limitation, in this research, we examined 4.5-year-olds' performance on the full RMTS task, as well as on the *same*-only and *different*-only versions. As in the original Christie and Gentner (2014) task, there were no training trials and no feedback during the test trials.

Experiment 1

The aim of this experiment is to examine whether 4.5-year-old children can spontaneously perceive relational similarities across exemplars, as evidenced by passing the full RMTS task. We tested three conditions: *same*-only, *different*-only, and *mixed* (half *same* and half *different*). In each condition, there were eight test trials, as in the studies by Christie and Gentner (2014) and Hochmann et al. (2017). The materials and procedures were identical to those used by

Christie and Gentner (2014), except for additions needed to test both *same* and *different* conditions.

Method

Participants Participants were 79 4.5-year-old children approximately evenly split between the three conditions: *same*-only (n = 26), *diff*-only (n = 26), and *mixed* (n = 27). The mean ages for each condition were as follows: *same*-only (M = 4 years 5 months; range = 3 years 11 months to 4 years 11 months; 13 female), *diff*-only (M = 4 years 5 months; range = 4 years 1 months to 4 years 10 months; 12 female), and *mixed* (M = 4 years 6 months; range = 4 years 2 months to 4 years 10 months; 14 female). Fourteen additional children were excluded from statistical analyses due to failing to pass all catch trials at the end of the RMTS task (10) (described below), experimenter error (2), refusal/non-compliance (1), or parental interference (1). Before beginning data collection, this target sample size of at least 24 children in each condition was calculated based on the goal of obtaining a power of .80 with an alpha level of $p <$ .05 and an effect size of d = .60 (Faul et al., 2007). The target sample size was determined based on the effect size (d) derived from the performance of 4.5-year-olds who achieved success on Christie and Gentner's (2014) RMTS task at above chance levels.

Children were recruited using a voluntary participant pool or through local preschools and received a small gift for their participation. The ethnic makeup of the sample was approximately 53% non-Hispanic, 11% Hispanic, and 36% No response. Of this sample, 52% of the sample was White, 9% were multiracial, 5% were Asian, 3% were Black/African American, 1% were American Indian or Alaskan Natives, and the remaining 30% had no responses.

Materials The materials were as in Christie and Gentner's (2014) study, except for the addition of additional cards required to test *different* as well as *same*. There were eight triads in each condition. Each of the three cards in a triad had two colored geometric shapes placed vertically. The individual shapes were familiar figures such as triangles and circles; each appearance of a given shape had a unique color. Each triad had a standard card depicting a *same* or *different* relation and two alternatives—one depicting *same* and the other depicting *different*, as shown in Figure 3. To avoid object matches, neither alternative included any shape that was also present in the standard. There were three conditions. In the *same*-only condition, all eight standards depicted the *same* relation; in the *different*-only condition, all eight standards depicted the *different* relation; and in the *mixed* condition, four of the standards depicted *same* and four depicted *different* (Figure 3). Left/right placement of the alternatives was counterbalanced. Three catch trials were given after the eight test trials to verify that children understood the task. In these triads, the standard was a single object (e.g., blue fish) and the alternatives were a card with an object that was highly similar to the standard (e.g., red fish) and another card with an object that was highly dissimilar to the standard (e.g., yellow cup).

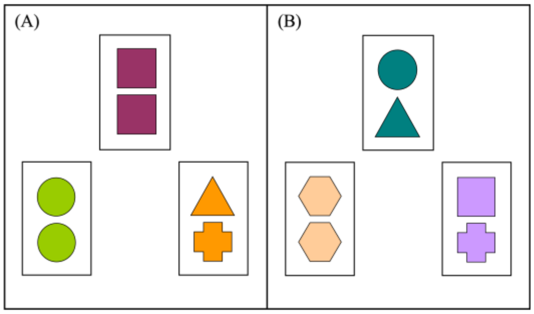

Figure 3: Sample triads for the RMTS task depicting (A) *same* and (B) *different* relations. For the *mixed* version of the RMTS task, half the trials depicted the *same* relation and the other half depicted *different*.

Procedure After obtaining parental consent and the child's verbal assent, the experimenter invited the child to a quiet room. The child was seated across from the experimenter. On each trial, the experimenter placed the standard card on the table and said "Look at this one! Do you see this one?" Then the experimenter placed the two alternatives below the standard simultaneously and said "Look at these two! Do you see these two? Which of these two is more like this one?" The child indicated their response by pointing to one of the two alternatives and the experimenter recorded it. No feedback was given apart from general positive encouragement throughout the study.

At the end of the eight test trials and the three catch trials, two additional questions were included but will not be examined in this paper: First, the experimenter asked children to revisit the last two test triads and explain their choice[1]. Second, to assess children's understanding of the words 'same' and 'different' affected their relational performance, children were asked to hand over a card that depicted either *same* or *different*, depending on condition.

Analysis Plan All statistical analyses for this experiment were performed using R and R Studio (R Core Team, 2022) using the 'tidyverse' (Wickham et al., 2019), 'stats' (R Core Team, 2022), 'rstatix' (Kassambara, 2021) packages. Depending on the research question, three types of analyses were conducted. First, as the dependent variable (percentage of relational match chosen) was not normally distributed, Wilcoxon's one-sample t-tests and its corresponding effect sizes (r) were calculated to test whether children's performance was significantly different from chance. Second, to examine whether children's relational performance was comparable across the three conditions, an Analysis of Variance (ANOVA) on the dependent variable of percentage of relational match chosen was conducted with

[1] This method follows Hochmann et al.'s (2017) technique, intended to trace the role of *same-different* language in this task.

between-subjects factors of condition (*same*-only, *different*-only, and *mixed*), age (in months), and sex (male or female). Finally, to examine whether there was a change in performance across the course of the experiment, a mixed ANOVA was conducted on children's relational performance with between-subjects factors of age (in months) and sex (male or female), and a within-subject factor of trial (first three or last three).

Results

The main question was whether 4.5-year-old children will succeed at relational matching on the full RMTS task. The answer is yes. Children chose the relational match at above-chance levels in each of the three conditions: *same*-only (M = 66.83%, SD = 28.71, Z = 222, p = .03 with Bonferroni correction, r = .48), *different*-only (M = 65.87%, SD = 19.86, Z = 209, p = .003, r = .64), and *mixed* (M = 67.59%, SD = 24.82, Z = 218.5, p = .008, r = .57) (see Figure 4).

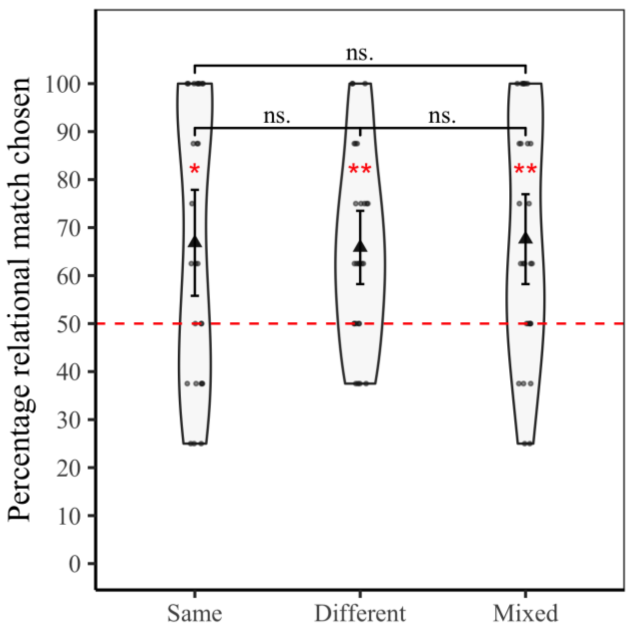

Figure 4: Percentage of times children chose the relational match in Experiment 1. For this figure and the ones that follow, the black triangles represent the mean, and the whiskers are standard error bars. The grey dots represent the individual data points, and the violin plot indicates the density of the distribution at different points of the dependent variable. * $p < .05$, ** $p < .01$.

We also found no evidence that children performed worse on the mixed *same-different* RMTS than on the *same*-only (or *different*-only) RMTS. An ANOVA with between-subjects factors of condition (*same*-only, *different*-only, and *mixed*), age (in months), and sex (male or female) revealed no main effects of condition (F(2, 73) = .10, p = .91, NS.), age (F(1, 73) = .61, p = .44, NS.), or sex (F(1, 73) = .38, p = .54, NS.). Further, the performance of the *mixed* group on the *same* trials (M = 67.59%, SD = 32.39) and the *different* trials (M = 67.59%, SD = 30.87) did not differ from each other, t(26) = 0, p = 1, nor did they differ from the performance of the *same*-only group, t(44.59) = -.24, p = .81.

Finally, we asked whether children's performance was consistent across trials. The answer is yes: Children's average relational performance on the first three trials did not differ significantly from that on the last three in any condition (*same*-only: F(1, 48) = .28, p = .60, NS.; *different*-only: F(1, 46) = 2.98, p = .09, NS.; *mixed*: F(1, 50) = .43, p = .51, NS.).

Discussion

The results from Experiment 1 show that 4.5-year-old children can spontaneously perceive relational similarities across exemplars. Children received neither training nor feedback, yet they succeeded on all three versions of the RMTS task: *same*-only, *different*-only, and *mixed*, and they performed comparably between the three conditions of the RMTS task. These findings suggest that children are indeed able to spontaneously succeed on the full RMTS task at 4.5 years of age, providing support for the results of Christie and Gentner (2014). However, before drawing conclusions, we need to consider another relevant study, that of Kroupin and Carey (2022).

Kroupin and Carey (2022, Experiment 1) assessed 4.5-year-olds' performance on the RMTS task with the shapes aligned on the horizontal axis (Figure 5). The experimenter placed the alternatives first, then the standard. Then, children chose which alternative "goes with" the standard. In the bare RMTS task, with no feedback nor prior training, children performed significantly below chance levels at 4.5 years, only succeeding when provided with training trials.

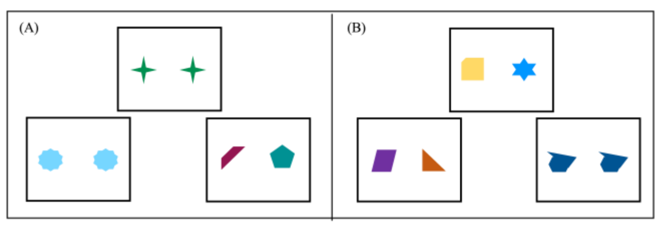

Figure 5: A sample triad of the (A) *same* and (B) *different* relations of the *mixed* RMTS task as used in Kroupin and Carey (2022).

Again, we see a marked difference in performance. In Christie and Gentner's (2014) study, 4.5-year-old children demonstrated spontaneous success, while in Kroupin and Carey's (2022) study, 4.5-year-old children were unable to spontaneously succeed on the RMTS task.

In Experiment 1, we demonstrated that 4.5-year-old children can succeed on the bare RMTS task with vertically placed stimuli. In Experiment 2, we examine whether they would also succeed on the RMTS task when given horizontally placed stimuli. This study will allow a closer comparison with Kroupin and Carey's (2022) results.

Experiment 2

Experiment 2 was identical to Experiment 1 except for three important differences: First, we modified the materials such that the shapes within each card were aligned on a horizontal axis rather than a vertical axis as in Experiment 1. Second, as there were no differences across the three conditions in Experiment 1, we ran only the *mixed* version of the RMTS task (i.e., with both *same* and *different* relations). Third, we used 16 trials in the current experiment compared to the 8 trials used in previous studies. Like Experiment 1, we predict that 4.5-year-old children will be able to spontaneously perceive relational similarities across exemplars with the new configuration of stimuli.

Method

Participants Participants were 22 4.5-year-old children (M = 4 years 8 months; range = 4 years 3 months to 5 years 0 months; 12 female). Six additional children were excluded from analysis for failing one or more catch trials. The target sample size is 24 children, based on the power analysis as described in Experiment 1.

The ethnic makeup of the sample was approximately 86% non-Hispanic and 14% Hispanic/Latinx. Of this sample, 65% of the sample was White, 8% were Black/African American, 8% were Asian/Hawaiian/Pacific Islander, 16% were multiracial, and the remaining 3% chose not to answer.

Materials The materials were identical to those of Experiment 1 with two differences: First, the stimuli were placed on a horizontal axis (Figure 6). Second, we increased the number of trials from 8 to 16 to better permit testing whether performance would change across trials. As before, each triad consisted of a standard card and two alternatives. Within the first eight trials, half had a standard that depicted the *same* relation, and the other half had a standard depicting the *different* relation; and likewise for the second eight trials.

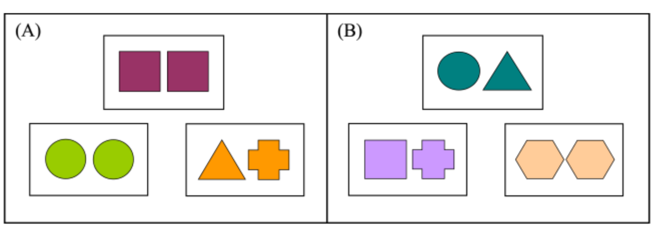

Figure 6: Sample triads of the RMTS task with stimuli placed in a horizontal orientation as used in Experiment 2. Half the trials depicted the (A) *same* relation and the other half depicted (B) *different*.

Procedure The procedure was identical to that of Experiment 1. The only exception was that families were given $20 (instead of a book) as compensation for their participation.

Analysis Plan The statistical plan was identical to Experiment 1. In addition, to determine whether children performed differently across Experiments 1 and 2, we conducted an independent samples t-test between children's performance on the first eight trials of the *mixed* RMTS task.

Results

The main question was whether 4.5-year-old children are capable of relational matching on the *mixed* RMTS using a new configuration of the stimuli. The data suggest that the answer is yes: Across all 16 triads, the 4.5-year-old children chose the relational match at above-chance levels (M = 78.13%, SD = 22.96, Z = 201.5, $p < .001$, r = .77) (see Figure 7). Performance was above-chance on the first eight trials (M = 75.57%, SD = 22.98, Z = 182, $p < .001$, r = .88) and on the last eight trials (M = 80.68%, SD = 25.51, Z = 199, $p < .001$, r = .78). There was no significant difference in performance on the first eight vs. the last eight trials (Z = 18.5, p = .20, NS.). An ANOVA on the dependent variable of relational performance across all 16 trials with between-subjects factors of age (in months) and sex (male or female) revealed no main effects of age ($F(1, 19)$ = .60, p = .45, NS.) nor sex ($F(1, 19)$ = 1.15, p = .30, NS.).

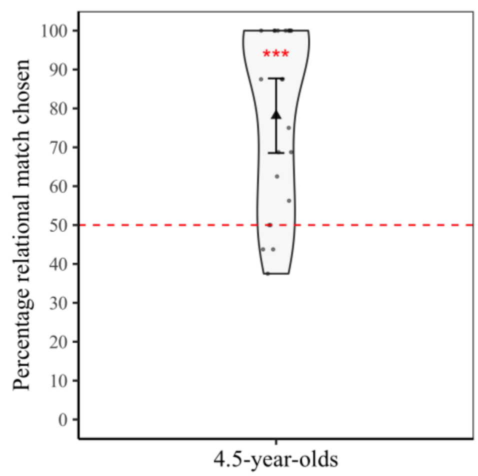

Figure 7: Percentage of the relational match chosen across all 16 trials in Experiment 2. See Figure 4 for a full explanation of the plot elements. *** $p < .001$.

Our second question was whether children would perform better on *same* triads (i.e., triads whose standard depicted the *same* relation) than on *different* triads. The answer is no: we found no significant difference in performance between *same* triads (M = 77.8%, SD = 41.7) and *different* triads (M = 78.4%, SD = 41.3), $t(21)$ = -.15, p = .88, NS.

We also asked whether children's performance would be stable across trials. As in the previous study, the answer is yes: Across all 16 trials, children's average relational performance was similar on the first three trials when compared to the last three ($M_{1\text{-}3}$ = 71.21%, $SD_{1\text{-}3}$ = 33.01, $M_{14\text{-}16}$ = 78.79%, $SD_{14\text{-}16}$ = 31.78), $F(1, 40)$ = 0.57, p = .45, NS. To compare performance on the current study to that of our prior studies, we analyzed the data from the first eight trials only. Here too, performance was comparable on the first three and the last three (M_{1-3} = 71.21%, SD_{1-3} = 33.01, M_{6-8} = 78.79%, SD_{6-8} = 30.07), F(1, 40) = .64, p = .43, NS.

Finally, we compared Experiments 1 and 2 to determine whether there was a difference in 4.5-year-old children's relational matching performance on the *mixed* RMTS task with a change in the configuration of stimuli. To compare performance across the two experiments, we examined only the first eight trials of Experiment 2. We found no significant differences in performance between Experiment 1 (M = 67.59%, SD = 24.82) and Experiment 2 (M = 75.57%, SD = 22.98., t(46.2) = -1.17, p = .25.

Discussion

The results of Experiment 2 replicate and extend those of Experiment 1 and provide additional evidence for spontaneous relational matching among 4.5-year-old children. Even with a new configuration of the stimuli compared to Experiment 1, children from this age group were able to successfully perceive relational similarities across exemplars. We found no difference in performance between the mixed RMTS and the *same*-only version of the RMTS. We also found no difference in performance between *same* trials and *different* trials. Children's relational performance was high and remained consistent across trials. Critically, this successful spontaneous relational performance was observed without the use of any training trials or corrective feedback during the task.

General Discussion

Across two experiments, 4.5-year-old children demonstrated successful relational performance on the full RMTS task. The findings from these two experiments provide robust evidence that this ability is present among children at 4.5 years of age without any training trials or feedback. The finding that 4.5-year-old children can succeed on the base RMTS task is not an artifact of running same-only trials. When given mixed *same-different* trials, with no training nor feedback, 4.5-year-olds can succeed on the RMTS task with both vertically and horizontally arranged stimuli.

Along with Christie and Gentner's (2014) findings, these findings provide clear evidence that 4.5-year-old children are able to spontaneously match *same* and *different* relations, with neither training nor feedback. This is a full year earlier than that found by Hochmann et al. (2017), even though their studies included either training prior to the test trials or feedback during the test. Likewise, Kroupin and Carey (2022) found that 4.5-year-olds were unable to pass the bare RMTS task.

An important question, then is *why* the findings on this task differ so widely between these studies. There is no obvious difference in the population. However, there are some differences in methods. In our studies, the cards all had vertical configurations of shapes (Experiment 1 and Christie & Gentner, 2014) or else all horizontal configurations (Experiment 2). This could have facilitated spatial alignment and detection of common or distinctive relations within the triads. In contrast, in Hochmann et al.'s (2017) experiments the spatial configuration of shapes within the cards was varied between vertical, right-diagonal, or left-diagonal (Figure 2). Within each triad, the configuration within the standard differed from that in the two alternatives. This could have made it difficult for children to align the cards within a triad and detect which alternative shared a relation with the standard (Matlen et al., 2020; Zheng et al., 2022). However, it is important to note that Kroupin and Carey (2022) used stimuli that were aligned on the horizontal axis and yet, failed to find spontaneous success at 4.5 years. Therefore, the differences in findings cannot be solely attributed to differences in spatial alignment.

Another difference is in the kinds of shapes used to make up the pairs. In Christie and Gentner's (2014) study and in our current studies, the pairs were made up of simple shapes such as triangle, circles, and squares, which are familiar to children by the age of 4 (Clements et al., 1999). The shapes were repeated across trials, though always with different colors. In contrast, Hochmann et al. (2017) and Kroupin and Carey (2022) used novel shapes throughout; that is, none of the component shapes was used more than once. While this seems a reasonable principle, it meant that many of the individual shapes used in the stimuli were rather complex and relatively unfamiliar to 4-year-olds. There is considerable evidence that attention to individual objects can interfere with children's relational matching (Christie & Gentner, 2007; Gentner & Toupin, 1986; Kroupin & Carey, 2022). Some prior studies have found poorer relational learning when the relation is instantiated across varied and distinctive objects than when it is instantiated with fewer and more familiar objects (Anderson et al., 2018; Casasola, 2005; Casasola & Park, 2013; Maguire et al., 2008). Thus, it is possible that the use of all-unique shapes on each trial might have distracted 4-year-olds from focusing on the relational patterns. Although further research will be needed to pinpoint the factors that impede or support children in this task, the key point is that 4.5-year-olds are capable of spontaneous success on the RMTS task.

Developmental research is still unpacking the cognitive, cultural, and linguistic factors that influence relational ability from its early manifestation in infancy (Anderson et al., 2018, 2022; Ferry et al., 2015) through childhood (Christie & Gentner, 2010, 2014; Gentner & Namy, 1999; Hochmann et al., 2017; Kroupin & Carey, 2022; Walker et al., 2016; Walker & Gopnik, 2014) and into adulthood (Doumas & Hummel, 2013; Gentner, 2010; Gick & Holyoak, 1983; Goldwater & Gentner, 2015; Vendetti et al., 2014).

In sum, the present findings demonstrate that 4.5-year-old children can spontaneously pass the full RMTS task. Relational reasoning is a key process in higher-order cognition, learning, and creative problem solving (Gentner et al., 2001; Holyoak & Thagard, 1997; Richland & Simms, 2015). Understanding the factors that promote relational insight will enable us to create learning environments that foster children's learning and discovery.

Acknowledgments

This research was supported by National Science Foundation grants [BCS-1423917, BCS-1729720BCS-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. The opinions expressed are those of the authors and do not necessarily represent views of the National Science Foundation or Office of Naval Research. We are grateful to the caregivers who agreed to have their children participate in this research. We thank the members of the Infant Cognition Lab and the Cognition and Language Lab for their assistance and insights.

References

Anderson, E. M., Chang, Y.-J., Hespos, S., & Gentner, D. (2018). 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., Chang, Y.-J., Hespos, S., & Gentner, D. (2022). No evidence for language benefits in infant relational learning. *Infant Behavior and Development*, *66*, 101666. https://doi.org/10.1016/j.infbeh.2021.101666

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

Casasola, M., & Park, Y. (2013). Developmental changes in infant spatial categorization: When more is best and when less is enough. *Child Development*, *84*(3), 1004–1019. https://doi.org/10.1111/cdev.12010

Christie, S., & Gentner, D. (2007). *Relational similarity in identity relation: The role of language*. 601–666.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), Article 3. https://doi.org/10.1080/15248371003700015

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

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

Clements, D. H., Swaminathan, S., Hannibal, M. A. Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education*, *30*(2), 192. https://doi.org/10.2307/749610

Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLoS ONE*, *8*(6), e63889. https://doi.org/10.1371/journal.pone.0063889

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), Article 10. https://doi.org/10.1177/0956797611422916

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferry, A. L., Hespos, S. J., & Gentner, D. (2015). Prelinguistic relational concepts: Investigating analogical processing in infants. *Child Development*, *86*(5), Article 5. https://doi.org/10.1111/cdev.12381

Flemming, T. M., Thompson, R. K. R., & Fagot, J. (2013). Baboons, like humans, solve analogy by categorical abstraction of relations. *Animal Cognition*, *16*(3), Article 3. https://doi.org/10.1007/s10071-013-0596-0

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. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), Article 5. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The analogical mind: Perspectives from cognitive science* (pp. xii, 541). The MIT Press. https://doi.org/10.7551/mitpress/1251.001.0001

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., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, *10*(3), Article 3. https://doi.org/10.1207/s15516709cog1003_2

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*(1), Article 1. https://doi.org/10.1016/0010-0285(83)90002-6

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*(7), Article 7. https://doi.org/10.1037/bul0000043

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. https://doi.org/10.1016/j.cogpsych.2016.01.005

Hochmann, J.-R., Tuerk, A. S., Sanborn, S., Zhu, R., Long, R., Dempster, M., & Carey, S. (2017). Children's representation of abstract relations in relational/array match-to-sample tasks. *Cognitive Psychology*, *99*, 17–43. https://doi.org/10.1016/j.cogpsych.2017.11.001

Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist*, *52*(1), 35–44. https://doi.org/10.1037/0003-066X.52.1.35

Kassambara, A. (2021). *rstatix: Pipe-friendly framework for basic statistical tests* (0.7.0). https://CRAN.R-project.org/package=rstatix

Kroupin, I. G., & Carey, S. E. (2022). You cannot find what you are not looking for: Population differences in relational reasoning are sometimes differences in inductive biases alone. *Cognition*, *222*, 105007. https://doi.org/10.1016/j.cognition.2021.105007

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), Article 4. https://doi.org/10.1111/j.1467-7687.2008.00707.x

Matlen, B. J., Gentner, D., & Franconeri, S. L. (2020). Spatial alignment facilitates visual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *46*, 443–457. https://doi.org/10.1037/xhp0000726

Obozova, T., Smirnova, A., Zorina, Z., & Wasserman, E. (2015). Analogical reasoning in amazons. *Animal Cognition*, *18*(6), 1363–1371. https://doi.org/10.1007/s10071-015-0882-0

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, *31*(2), Article 2. https://doi.org/10.1017/S0140525X08003543

Premack, D. (1983). Animal cognition. *Annual Review of Psychology*, *34*, 351–362.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *WIREs Cognitive Science*, *6*(2), Article 2. https://doi.org/10.1002/wcs.1336

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, *25*(2), Article 2. https://doi.org/10.1016/j.cub.2014.11.063

Smith, J. D., Redford, J. S., Haas, S. M., Coutinho, M. V. C., & Couchman, J. J. (2008). The comparative psychology of same-different judgments by humans (Homo sapiens) and monkeys (Macaca mulatta). *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(3), Article 3. https://doi.org/10.1037/0097-7403.34.3.361

Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (Pan troglodytes) judge relations between relations in a conceptual Matching-to-Sample task. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 20.

Vendetti, M. S., Wu, A., & Holyoak, K. J. (2014). Far-out thinking: Generating solutions to distant analogies promotes relational thinking. *Psychological Science*, *25*(4), Article 4. https://doi.org/10.1177/0956797613518079

Walker, C. M., Bridgers, S., & Gopnik, A. (2016). The early emergence and puzzling decline of relational reasoning: Effects of knowledge and search on inferring abstract concepts. *Cognition*, *156*, 30–40. https://doi.org/10.1016/j.cognition.2016.07.008

Walker, C. M., & Gopnik, A. (2014). Toddlers infer higher-order relational principles in causal learning. *Psychological Science*, *25*(1), Article 1. https://doi.org/10.1177/0956797613502983

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), Article 1. https://doi.org/10.1037/a0016327

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Zheng, Y., Matlen, B., & Gentner, D. (2022). Spatial alignment facilitates visual comparison in children. *Cognitive Science*, *46*(8), e13182. https://doi.org/10.1111/cogs.13182