Cognition

The Role of Language in Forming the Concepts Same and Different

--Manuscript Draft--

Manuscript Number:	COGNIT-D-25-00291R2
Article Type:	Full Length Article
Section/Category:	Ori Friedman- Social cognitive development,conceptual development, judgment, experimental philosophy
Keywords:	Relational Reasoning; analogy; early childhood; Cognitive Development; abstract learning
Corresponding Author:	Apoorva Shivaram Northwestern University UNITED STATES
First Author:	Apoorva Shivaram, PhD
Order of Authors:	Apoorva Shivaram, PhD
	Ruxue Shao, PhD
	Susan J. Hespos, PhD
	Dedre Gentner, PhD
Abstract:	Relational ability – the ability to compare situations or ideas and discover common relations – is key to our capacity for higher-order cognition. The Relational-Match-to-Sample task (RMTS) is the gold standard to evaluate relational ability, yet recent reports conflict on when and how this ability emerges. Here, we focus on a version of the task in which children receive no training or feedback and ask whether possessing the relational concepts same and different is the key to success. We present two experiments that found 4-year-old children succeeded on three RMTS conditions aligned on the vertical axis (N = 79; 39 female and 40 male; 52% White; Wilcoxon's r >= .48) and one on the horizontal axis (N = 31; 15 female and 16 male; 73% White; r = .62). Across both experiments, children who justified their choices using relational language performed better on the RMTS task. We conclude that possession of the same/different concept is the key to success on the RMTS and production of relational language to justify performance is a better predictor than age for indicating possession of the concept.

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Title: The Role of Language in Forming the Concepts *Same* and *Different*

Authors: Apoorva Shivaram[1]*, Ruxue Shao[2], Susan J. Hespos[3], Dedre Gentner[1]

Affiliations:
[1] Department of Psychology, Northwestern University, Evanston, IL 60208, USA
[2] School of Foreign Languages, Wuhan Textile University, Wuhan, Hubei Province 430200, China
[3] The MARCS Institute for Brain, Behaviour & Development, Western Sydney University, Penrith, NSW 2751, Australia

*Correspondence to: ashivaram@u.northwestern.edu; Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208

Declaration of Interest:
This research was supported by National Science Foundation Grants [BCS-1423917, BCS-1729720] awarded to Susan J. Hespos and Dedre Gentner, by NSF SLC 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 the views of the National Science Foundation or Office of Naval Research.

Acknowledgments:
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 at Northwestern University for their insights and assistance with data collection. We particularly thank Nina Simms for critical feedback on earlier versions of this paper.

Conflict of Interest Statement:
The authors have no conflicts of interest to disclose.

Data Availability Statement:
All materials, data, and code have been made publicly available on Open Science Framework and can be accessed through this OSF repository: https://osf.io/rqwjk/. This study was not preregistered.

Dear Prof. Friedman,

Thank you for the opportunity to do the minor revisions. We have copied the Reviewers responses below and detailed our response to them in blue below each comment.

Reviewer #1: First, I want to thank the authors for substantially engaging with my review. The manuscript is much improved, both in its theoretical framing and in the clarity of the writing. The main theoretical contribution of this article is significantly clarified too: possession of abstract same/different concepts (as evidence by spontaneous relational language) is required for success on unscaffolded version of RMTS. I have two minor suggested revisions:

1) Please report effect sizes for all statistical analyses. In particular, your t-tests are missing effect sizes.

Thank you for catching this omission. We have added standardized beta coefficients (β) with 95% CIs to all regression analyses in the main manuscript and supplementary analyses. We note that one-sample tests reported in the manuscript are Wilcoxon signed-rank tests rather than *t*-tests, as the dependent variable was not normally distributed. Effect sizes for these tests were already reported in the prior version of the manuscript in the form of Wilcoxon's *r*.

2) I was disappointed to see no discussion of the entropy-based accounts of same/different perception (e.g., Glorioso et al, 2021; Wasserman et al, 2021). I am quite sympathetic to the theoretical narrative developed in the current manuscript. I don't find the entropy-based accounts particularly explanatory either; however, they present one possible alternative explanation as to how babies succeed at these other same/different tasks while toddlers without relational language fail at the tradition RMTS task. I think you should at least cursorily mention this work and then tell us why you don't buy the account. A clear spot to do this would be in the discussion section in the paragraph starting: "At this juncture, it may seem puzzling that 3-year-olds fail the RMTS without special support, yet 7- and 9-month-old infants, and even 3-month-olds, can form abstract concepts of same and different (Anderson et al., 2018; Ferry et al., 2015; Hespos et al., 2021"

We now include the following paragraph in the discussion.

Hochmann et al. (2017) found that 3- and 4-year-old children, like pigeons, fail 2-item array-match-to-sample (RMTS) tasks but succeed with 16-item arrays. Success by preschool children and pigeons on the 16-item display is attributed to encoding entropy (defined as the internal variability of an array). In contrast, older children and adults can detect the relation of same/different between pairs of objects. Although entropy-based accounts have been proposed (e.g., Glorioso et al., 2021), in the pure version of the RMTS task used in this study, there is strong evidence that success cannot be achieved using entropy perception, as the entropy of a *same* pair (0) and that of a *different* pair (1) are not sufficiently different to be detected given our method (i.e., presenting eight RMTS trials without practice or feedback). There is debate over whether a human or pigeon could be trained to detect this difference, but there is general agreement that any such training would require many hundreds or thousands of trials. There is an excellent discussion of this issue in Hochmann et al. (2017) (p. 19 and following).

Reviewer #2: I appreciate the authors response to my comments, and they have greatly improved the paper. I still have some doubts about the size of the contribution (I don't think we should continue to incentivize using same/different as a stand in for all relational development), but I think the paper is sufficiently improved for publication. Hopefully I will get to review papers with different relational measures/tasks for young kids soon.

Research Highlights

- Four-year-old children succeeded on the RMTS task without training or feedback
- They performed equally well on *same*-only, *different*-only, and full RMTS trials
- Children using relational language to justify choices perform better on the RMTS task
- Command of the relational concepts *same* and *different* is linked to success on RMTS

The Role of Language in Forming the Concepts *Same* and *Different*

Manuscript submitted on ~~February 15~~April 25, 2026

Abstract

Relational ability – the ability to compare situations or ideas and discover common relations – is key to our capacity for higher-order cognition. The Relational-Match-to-Sample task (RMTS) is the gold standard to evaluate relational ability, yet recent reports conflict on when and how this ability emerges. Here, we focus on a version of the task in which children receive no training or feedback and ask whether possessing the relational concepts *same* and *different* is the key to success. We present two experiments that found 4-year-old children succeeded on three RMTS conditions aligned on the vertical axis (N = 79; 39 female and 40 male; 52% White; Wilcoxon's r >= .48) and one on the horizontal axis (N = 31; 15 female and 16 male; 73% White; r = .62). Across both experiments, children who justified their choices using relational language performed better on the RMTS task. We conclude that possession of the *same/different* concept is the key to success on the RMTS and production of relational language to justify performance is a better predictor than age for indicating possession of the concept.

Keywords: relational reasoning; analogy; early childhood; cognitive development; abstract learning

1. Introduction

A major component of human intelligence is our ability to perceive relations between objects, events, or ideas and transfer those relations to new situations. We use this ability every day. For example, you navigate through a new city by mentally aligning the image on your phone's map with real-life landmarks and streets. Relational ability is the vital foundation for effective learning, particularly in mathematics and science (e.g., Green et al., 2017; Jee et al., 2013; Mix et al., 2024). Thus, clarifying how relational abilities emerge and develop is key to understanding why some children learn more readily than others and how best to support this ability in all children.

To study relational ability at its core, we focus on *same* and *different* relations—the fundamental building blocks of relational matching. These relations are fundamental both for concept formation and everyday reasoning, are perceptually available to young children and nonhuman animals, and have been studied extensively across many species, offering the possibility of phylogenetic as well as ontogenetic comparisons (e.g., Christie & Gentner, 2014; Hochmann et al., 2017; Wasserman & Young, 2010). The gold standard for evaluating relational ability is the Relational-Match-To-Sample (RMTS) task (given AA, choose XX over YZ; given BC, choose YZ over XX) developed by Premack (1983). This task has been widely used to assess relational ability in both developmental (Carstensen et al., 2019; Christie & Gentner, 2014; Hochmann et al., 2017; Kroupin & Carey, 2022; Walker et al., 2016; Walker & Gopnik, 2014) and cross-species (Fagot & Thompson, 2011; Flemming et al., 2007; Premack, 1983; Thompson et al., 1997; Wasserman & Young, 2010) investigations.

In Premack's (1983) seminal paper, he proposed the RMTS as a true test of relational ability, distinguishing it from two other tasks that tap into *sameness* and *difference.* The first and easiest task is the Match-To-Sample task (MTS; Given A, choose A over B). This task simply requires detecting identical matches between objects; it can be passed by many species, including pigeons, macaques, honeybees, and baboons (Fagot & Thompson, 2011; Flemming et al., 2007; Thompson et al., 1997; Wasserman & Young, 2010; Zentall et al., 2018) as well as 10-month-old humans (Hochmann et al., 2016). The second task is the *same/different* task. In this task, the participant is first trained, with feedback, to classify a set of stimuli into *same* and *different* pairs (i.e., to make one response for AA, BB, CC etc. and a different response for AB, CD, EF etc.). If after mastering the training set, the participant can correctly classify novel pairs (e.g., XX, YZ etc.), they are considered to have passed the *same/different* task. Both the MTS and the *same/different* task involve training. However, the *same/different* task is more difficult than the MTS task, with fewer species achieving successful learning. For instance, pigeons can pass the MTS task but fail the 2-item *same/different* task (Wasserman & Young, 2010). An important distinction between the MTS and the *same/different* task in that success on the MTS task can be achieved by detecting object matches (X with X) whereas success on the *same/different* task requires the ability represent the relations *same* (XX) and *different* (YZ). Premack (1983) suggested that differences in the ability to form and use relational representations explain why fewer species succeed on the *same/different* task than on the MTS task.

The third test, the RMTS task was designed by Premack as a true test of the ability to recognize and match *same* and *different* relations. It is given after the subject has mastered the *same-different* task; the idea is that if a subject can fluently encode a new pair of items as *same* or *different*, then they will pass the RMTS task. It is ideal for developmental and cross-species comparisons because it does not require specialized content knowledge – only an ability to encode and match *same* and *different* relations. Premack hypothesized that possessing symbols for *same* and *different* was critical in passing this task – an idea that is consistent with many recent findings on the role of relational language in relational cognition (e.g., Gentner, 2016; Loewenstein & Gentner, 2005; Yuan et al., 2024). We return to this idea below.

Because of its importance as an indicator of relational ability, the RMTS has been used in a number of recent developmental studies. This research has considered a number of factors that contribute to children's performance on this task. Among these are pretraining children to label figures as *same* or *different* (Christie & Gentner, 2014), pretraining on a series of matching tasks (Kroupin & Carey, 2022, Experiment 2), embedding the RMTS within a causal *same-different* task (Carstensen et al., 2019; Walker et al., 2016; Walker & Gopnik, 2014), and training and/or feedback during the RMTS task (Hochmann et al., 2017). There is also evidence that cultural and linguistic factors influence performance (Carstensen et al., 2019; Murphy et al., 2021).

These studies of the supported RMTS task have provided valuable information about the paths through which children gain relational insight. However, our interest is in when children can pass the pure, unsupported RMTS. This is important, because true relational competence requires the ability to spontaneously notice and respond to relations in the world. We suggest that for this to happen, the child needs to have stored representations of *same* and *different* that can be accessed and used across contexts. A partial answer to the question of how children fare on the unsupported RMTS was provided by Christie and Gentner (2014, Experiment 1) who administered a *same*-only version of the RMTS task to 2.5-, 3.5-, and 4.5-year-old children. On all eight trials, children saw a standard that depicted the relation of *same* (e.g., two squares), followed by two alternatives—one depicting *same* and one depicting *different.* Children were asked to choose the alternative that was more like the standard. There were no training trials prior to the task, nor was any feedback provided during the task. The findings revealed that 4.5-year-olds succeeded and that neither 2.5- nor 3.5-year-olds did so—suggesting that children can pass the *same*-only RMTS by 4.5 years of age. However, these results do not tell us how children would fare on the full RMTS, which requires both *same* and *different* standards.

Two studies that used both *same* and *different* standards have somewhat different results. Kroupin and Carey (2022, Experiment 1) gave children eight trials of the pure RMTS task (that is, with neither pre-training nor feedback) with half the trials having a *same* standard and half a *different* standard. They found that 4.5-year-olds performed below chance. Kroupin and Carey (2022) did not include older children, but in a related study, Hochmann and colleagues (2017, Experiment S1, Appendix A) tested 3-, 4-, 5-, and 6-year-olds on a task with four *same* and four *different* trials, with corrective feedback on each test trial. The 5- and 6-year-olds succeeded, but children who were four and younger failed. In another experiment, Hochmann et al. (2017, Experiment 2) gave 4- and 5-year-old children the full RMTS task (four *same* trials and four *different* trials) with eight trials of pretraining before the test. The mean age of success was 5.5 years; again, 4-year-olds performed at chance. This discrepancy is striking given that the study by Hochmann et al. (2017) provided support—either pre-training or feedback—for the RMTS task.

One possible reason for this discrepancy is that Christie and Gentner (2014) only included *same* trials in their study, whereas the other studies used both *same* and *different* trials. As Hochmann et al. (2017, page 22) pointed out, success on the *same*-only task "could be solved by learning a single rule 'choose *same'*…". Therefore, in Experiment 1 we extended the methods in Christie and Gentner's (2014) study by testing three between-subject conditions: eight *same* trials, eight *different* trials, and four *same*- plus four-*different* trials.

A second motivation for this study concerns the role of *same-different* language. We ask whether possession of the relational concepts of *same* and *different* is a predictor of success in the RMTS task. To assess understanding of *same* and *different*, we adopted a technique used by Hochmann et al. (2017). On the last two trials of the RMTS, we asked children to explain their choice. Children who used the concepts *same* and/or *different* in their justifications were considered to have demonstrated productive command of the corresponding concepts. These manipulations will clarify whether 4-year-olds have the capacity to detect and match both *same* and *different* relations, and whether this capacity is associated with productive command of the corresponding linguistic terms or age.

2. Experiment 1

The goal of this experiment is to examine (1) whether children can succeed on the RMTS task with standards depicting *same* and *different* relations and no training or feedback; and (2) whether justifying their answers with 'same' and 'different' predicts success on the RMTS. We tested three conditions: *same*-only, *different*-only, and *full* (half *same* and half *different* trials). Experiments 1 and 2's design and analyses were not preregistered. All materials, data, and analysis code for both experiments have been made publicly available at Open Science Framework and can be accessed through this link.

2.1 Methods

2.1.1 *Participants*

Before beginning data collection, the 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 effect size was estimated based on the effect size (*d*) derived from the performance of 4-year-olds who achieved success at above chance levels on Christie and Gentner's (2014) RMTS task. This study was approved by the Institutional Review Board at a major US University (protocol number withheld for review).

Participants were 79 4-year-old children, approximately evenly split between the three conditions (*M* = 4 years 5.5 months, range = 3 years 11 months to 4 years 11 months). For a detailed breakdown of participant demographics, see Supplementary Materials, Section 1.1.1. An additional 14 children were excluded from statistical analyses due to failure to pass all three catch trials (described in Section 2.1.2 below) (n = 10), experimenter error (n = 2), refusal/non-compliance (n = 1), or parental interference (n = 1).

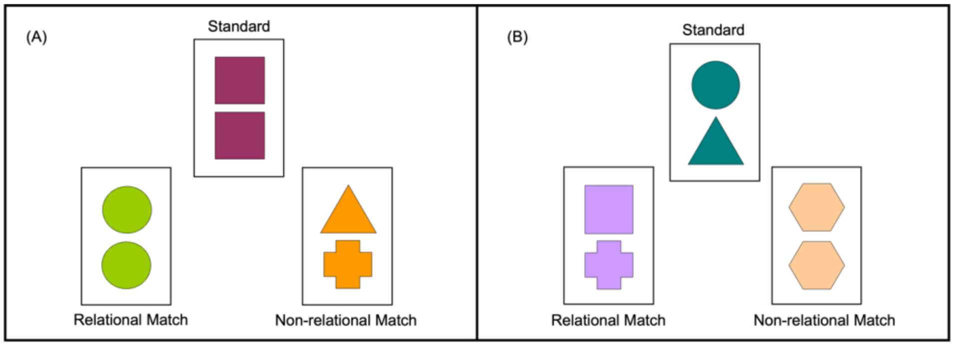
Children for this and the following experiment were recruited using a voluntary participant pool or through local preschools in a major metropolitan city in the United States. All participants received a small gift for their participation. Across both experiments, the ethnic makeup was approximately 61% non-Latine, 13% Latine, and 26% provided no response. Approximately 56% were White, 12% were multiracial, 5% were Asian, 3% were Black/African American, 1% were Indigenous American, and the remaining 23% had no response. For the two experiments, approximately 73% of primary caregivers had a college degree or higher, 2% had some college experience, 1% had a high school diploma, and the remaining 24% had no response. Among the secondary caregivers, 65% had a college degree or higher, 5% had some college experience, 2% had some high school education, and 28% had no response.

2.1.2 *Materials*

The materials were identical to those in Christie and Gentner's (2014) study, except for the additional cards required to have *different* standards. There were three conditions, each with eight trials: *same*-only, *different*-only, and *full* RMTS (four *same* and four *different*) (Figure 1). See the Supplementary Materials, Section 1.1.1 for more details.

Figure 1

Sample Triads for the RMTS Task Used in Experiment 1 Depicting (A) Same and (B) Different Relations. For the Full Version of the RMTS Task, Half the Standards Depicted the Same Relation and the Other Half Depicted Different.



In addition to the eight RMTS trials, three catch/MTS trials were given to verify that children understood the task: e.g., a blue fish as the standard with a similar red fish and a yellow cup as alternatives. After the catch trials, we repeated the final two RMTS triad trials where the child was asked to justify why they made the choice they did.

2.1.3 *Procedure*

After obtaining parental consent and the child's verbal assent, the experimenter invited the child to a quiet room. The child was seated at a table 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 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. No feedback was given apart from general positive encouragement. After each trial, the experimenter noted the child's response on a sheet of paper. When the triad trials were complete, the three catch trials were administered using the same instructions as the eight RMTS trials.

Next, children were asked to justify their choice on the last two triad trials. For these two justification trials, the trial was repeated and once the child indicated their response, the experimenter removed the non-chosen card and asked, "Why is this picture more like this one?" while gesturing to the two remaining cards.

2.1.4 *Coding*

For each participant, the primary coder scored all eight trials of the RMTS task, the three catch trials and the two justification trials. In addition, 40% of the justification trials were double coded for whether the child used the term *same* and/or *different* (e.g., "Because they are same shapes" or "Because it has 2 – because it's both different shapes"). The two coders were unaware of the child's RMTS response. Reliability was 98%, resulting in a Cohen's kappa of .94. Disagreements between coders were resolved through discussion.

2.2 Results

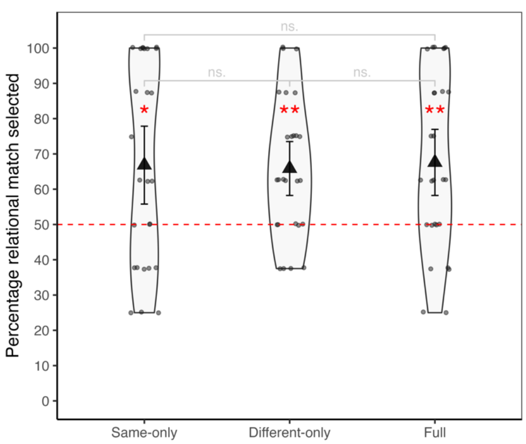
All statistical analyses in the current study were performed using R and R Studio (Posit team, 2025) using the 'tidyverse' (Wickham et al., 2019), 'stats' (R Core Team, 2025), 'rstatix' (Kassambara, 2023), and 'lme4' (Bates et al., 2015) packages. A complete list of packages used for data cleaning and visualization can be found in the R script in this OSF repository. This repository also contains an HTML report of analyses conducted for this study with complete regression tables.

2.2.1 *Performance on the RMTS task*

The first question asked whether 4-year-olds succeeded on the RMTS task. To answer this question, we conducted Wilcoxon one-sample tests and calculated the corresponding effect size (Wilcoxon's r), as the dependent variable (percentage of relational match selected) showed evidence of non-normality based on Shapiro–Wilk tests (*same*-only: $W = .86$, $p = .002$; *different*-only: $W = .93$, $p = .07$; *full*: $W = .91$, $p = .03$). The answer is yes: Children's performance was significantly above chance in all three conditions of the RMTS task: *same*-only ($M = 66.83\%$, $SD = 28.71$, $Z = 222$, $p = .03$ with Bonferroni correction, Wilcoxon's $r = .48$), *different*-only ($M = 65.87\%$, $SD = 19.86$, $Z = 209$, $p = .003$ with Bonferroni correction, $r = .64$), and *full* ($M = 67.59\%$, $SD = 24.82$, $Z = 218.5$, $p = .008$ with Bonferroni correction, $r = .57$; see Figure 2).

Figure 2

Percentage of Relational Match Selected in Experiment 1.



Note. 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 various points of the dependent variable. * $p < .05$, ** $p < .01$, *** $p < .001$. The red dashed line indicates chance performance.

To examine whether children perform comparably across the three conditions, we conducted a multiple regression predicting the percentage of relational matches selected from condition (*same*-only, *different*-only, or *full*), age (in months), gender (female or male), and order of presentation (five orders). The overall regression model was not statistically significant, $R^2 = .04$, $R^2_{adj} = -.07$, $F(8, 69) = .33$, $p = .95$, indicating that the predictors collectively did not account for variation in performance. Holding all other predictors constant, neither the *different*-only nor *full* conditions differed from the *same*-only condition (*different*-only: $b = –2.26$, 95% CI [–16.60, 12.07], $t(69) =$ ~~–0.~~–.31, $p = .754$~~;~~, $\beta = –.09$, 95% CI [–.68, .50]; *full*: $b =$ ~~0~~.14, 95% CI [–14.30, 14.59], $t(69) =$ ~~0~~.02, $p = .984$~~).~~, $\beta = .01$, 95% CI [–.59, .60]). A derived pairwise comparison between the *different*-only and full conditions, obtained by re-leveling the factor, was also nonsignificant, $b = 2.41$, 95% CI [–12.19, 17.00], $t(69) =$ ~~0~~.33, $p = .743$~~.~~, $\beta = .10$, 95% CI [–.50, .70]. None of the remaining predictors were significant (age: $b =$ ~~0~~.45, $p = .68$~~;~~, $\beta = .05$, 95% CI [–.21, .31]; gender: $b = 2.20$, $p = .71$~~;~~, $\beta = .09$, 95% CI [–.39, .57]; order: all dummy-coded contrasts *p*s > .41, |β|s < .28). These results indicate that children performed comparably across all three RMTS conditions and that age was not a significant predictor of relational performance.

2.2.2 *Effects of Relational Language*

Our second goal was to test whether production of the terms ‘same’ and ‘different’ to justify their choice would predict success on the RMTS task. We conducted a multiple regression with percentage of relational match selected as the outcome variable and the number of relational language justifications provided (0, 1, or 2 of 2 trials), condition (*same*-only, *different*-only, or *full*), age (in months), and gender (female or male) as predictors. Using the relational terms on both justification trials, relative to justifying their choices with no relational terms was marginal, but not a significant predictor of children's relational performance on the RMTS task when holding all other predictors constant, *b* = 14.33, 95% CI ~~=~~ [-1.69, 30.35], *t*(68) = ~~.35~~1.79, *p* = .079~~.~~, *β* = .58, 95% CI [–.07, 1.23]. Using *same* and *different* relational terms on one justification trial, relative to none, was also not a significant predictor of children's relational performance on the RMTS task, *b* = .65, 95% CI ~~=~~ [-14.98, 16~~,~~.28], *t*(68) = .08, *p* = .934~~.~~, *β* = .03, 95% CI [–.61, .66]. The descriptive statistics lend further details. Children who did not produce relational language showed lower accuracy and considerable variability (*M* = 63.8%, *SD* = 22.5%), including some who performed well on the task. Children who produced relational language on both trials had higher accuracy, but also showed considerable variability (*M* = 79.2%, *SD* = 26.6%), including some who failed RMTS trials.

2.3 Discussion

These data provide clear evidence that children succeed on the RMTS task across all three condition types (*same* only, *different* only, and mixed *same* and *different* standards). Further, performance was similar across the three conditions; the mean proportion correct ranged from 65.9% to 67.6%. However, we did not find the predicted relation between spontaneous production of the *same* and *different* terms and performance on the RMS task. One possible factor here is that unlike Hochmann et al. (2017), this experiment had only eight trials. It is possible that with more trials, children might have found it easier to produce the relevant relational terms. This idea was tested in Experiment 2.

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

How do we reconcile the finding that children pass the RMTS at 4 years of age with the negative results from Hochmann et al. (2017) and Kroupin and Carey (2022)? There are differences in the materials and procedures that could have affected the results. Here we highlight two of these differences. Additional differences and a table comparing experimental details across several RMTS tasks is provided in the supplementary materials (see Section 1.2.1). One difference that stands out is in spatial alignability. There is evidence that spatial alignment fosters relational alignment, both in adults (Matlen et al., 2020; Simms et al., 2023) and in children (Zheng et al., 2022). In our Experiment 1, all the pairs had the same (vertical) spatial configuration, whereas in Hochmann et al.'s (2017) studies, the pairs within a triad varied in their spatial configuration, making spatial alignment more difficult (Figure 3). In Kroupin and Carey's (2022) study, the figures were spatially aligned, but the configuration within a pair was horizontal, rather than vertical as in the current studies (Figure 4). To test whether the horizontal configuration diminishes performance, in Experiment 2 we presented the objects with a horizontal configuration.

Figure 3

A Sample Triad of the (A) Same and (B) Different Relations of the Full RMTS Task as Used in Hochmann et al. (2017).

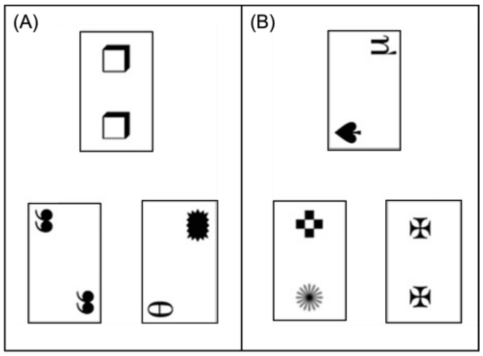
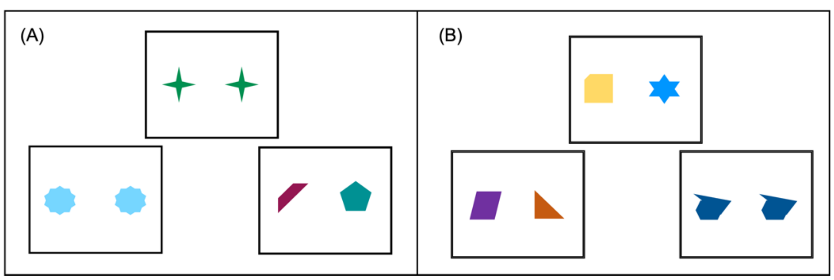


Figure 4

A Sample Triad of the (A) Same and (B) Different Relations of the RMTS Task as Used in Kroupin and Carey (2022).



A second difference is the kinds of shapes that make up the pairs. In the current studies, the pairs were made up of simple shapes such as triangles, circles, and squares, which are familiar to children (Clements et al., 1999). The shapes were repeated across trials, though always with different colors and in different positions. In contrast, Hochmann et al. (2017) and Kroupin and Carey (2022) used unique shapes throughout. Consequently, many of the individual shapes used in the stimuli were complex and unfamiliar to 4-year-olds. Prior studies have found that children perform worse on relational tasks when the objects instantiating the relations are complex rather than simple (Gentner & Rattermann, 1991; Hespos & Spelke, 2004; Kuwabara & Smith, 2012; Son et al., 2008) and novel rather than familiar (Anderson et al., 2018; Casasola, 2005; Casasola & Park, 2013; Maguire et al., 2008; Son et al., 2008).

In summary, there are many factors that could have affected performance. However, in light of the variability across labs, we wanted to ensure that our results would generalize across a different spatial configuration. Therefore, in Experiment 2, we used horizontal rather than vertical pairs. In addition, we increased the number of trials from 8 to 16, enabling us to detect possible learning across trials.

3. Experiment 2

The aim of Experiment 2 was to investigate whether preschoolers can succeed on the RMTS task given horizontally arranged stimuli and more trials. Experiment 2 was identical to the Experiment 1 condition that had both *same* and *different* standards. There were two changes: First, we modified the materials such that the shapes within each card were aligned on a horizontal axis rather than a vertical axis. Second, we ran 16 trials.

3.1 Methods

3.1.1 *Participants*

Participants were 31 4-year-old children (15 female, 16 male; M = 4;7, range = 4;3 to 5;0). An additional 10 children were excluded from analysis for failing one or more catch trials. To increase the power when compared to Experiment 1, we conducted an a priori power analysis with the goal of obtaining a power of .90 with an alpha level of $p < .05$ and an effect size of d =

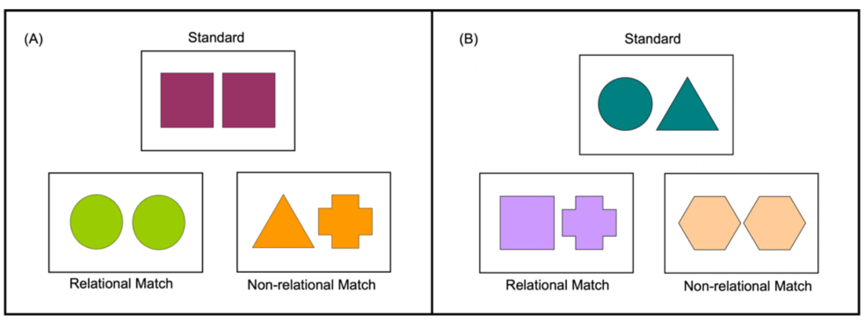
.60 (Faul et al., 2007). The resulting target sample size is approximately 32 children. This study was approved by the Institutional Review Board at a major US University (protocol number withheld for review). Families were given $20 as compensation for their participation and a child-sized T-shirt.

3.1.2 *Materials*

The materials were identical to those of Experiment 1 with two modifications. First, the stimuli were placed on a horizontal axis (Figure 5). Second, there were 16 trials.

Figure 5

Sample Triads from Experiment 2 with Stimuli Placed in a Horizontal Orientation. (A) Half the Standards Depicted the Same Relation (B) and the Rest were Different.



3.1.3 *Procedure*

The procedure was identical to that of Experiment 1. The only exception was that due to a mistake at the beginning of data collection, one participant was not administered any justification trials and five participants were administered only one of the two justification trials.

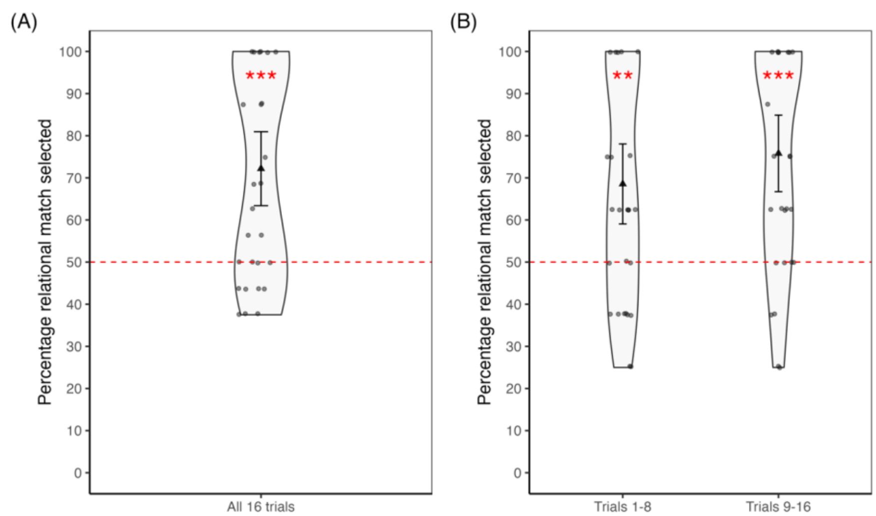
3.2 Results

3.2.1 *Performance on the RMTS task*

Similar to Experiment 1, we conducted a Wilcoxon's one-sample ~~*t*~~-test and calculated its corresponding effect size (Wilcoxon's *r*) as the dependent variable (percentage of relational match selected) was not normally distributed. Again, 4-year-old children succeed on the RMTS task. Children performed at significantly above-chance levels (M = 72.18%, SD = 24.94, Z = 338.5, $p < .001$, Wilcoxon's r = .62~~,~~ see Figure 6A). Examining each set of eight trials in turn, children performed at significantly above chance on the first set of eight trials (M = 68.55%, SD = 26.98, Z = 337, p = .004, Wilcoxon's r = .88) as well as on the second set of eight trials (M = 75.81%, SD = 25.81, Z = 350, $p < .001$, Wilcoxon's r = .88) (Figure 6B).

Figure 6

Percentage of Relational Match Selected in Experiment 2 (A) Across all 16 Trials and (B) Across Trials 1-8 and 9-16.



We also asked whether children's performance would be stable across trials of the RMTS task. We conducted a linear mixed-effects regression model with average percentage of relational match selected as the dependent variable, fixed effects of trial (first three or last three), age, and gender with random effects by participant. 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, *b* = 5.38, 95% CI~~=~~ [-2.87, 13.63], *t*(56) = 1.31, *p* = .197, *β* = .17, 95% CI [-.09, .43] (see Table S4 in Supplementary Materials, Section 1.2). For comparability with prior studies which used 8 trials, we also analyzed the data from the first eight trials only. Here too, performance was comparable on the first three and the last three trials, *b* = 3.23, 95% CI~~=~~ [-7.66, 14.11], *t*(56) = .59, *p* = .555, *β* = .10, 95% CI [-.24, .44] (see Supplementary Materials, Section 1.2, Table S4).

Analyses were also performed on children's performance on *same* vs. *different* trials and comparisons across Experiments 1 and 2. There were no significant differences in relational performance in any of these analyses (see Section 1.2 in Supplementary Materials).

3.2.2 *Effects of Relational Language*

Children who used the relational terms *same* and *different* in justifying their choices also had high accuracy on the RMTS task. We conducted a linear regression with percentage of relational match selected as the outcome variable and the number of *same-different* justifications (0, 1, or 2), age (in months), and gender (female or male) as predictors. Using the relational terms *same* and *different* on both justification trials, relative to justifying their choices with no relational terms, was a significant predictor of children's relational performance on the RMTS task, b = 44.60, 95% CI ~~=~~ [23.89, 65.31], $t(26) = 4.43$, $p < .001$~~.~~, β = 1.79, 95% CI [.96, 2.62]. Similarly, using *same* and *different* relational terms on one justification trial, relative to none, was also a significant predictor of children's relational performance on the RMTS task, b = 29.66, 95% CI ~~=~~ [12.06, 47.25], $t(26) = 3.47$, $p = .002$~~.~~, β = 1.19, 95% CI [.48, 1.89]. The descriptive statistics lend further details. Children who did not produce relational language revealed lower accuracy and considerable variability (M = 58.0%, SD = 20.4%) including some who performed well on the task. In contrast, children who produced relational language on both trials had a much higher accuracy and very little variability (M = 97.9%, SD = 5.1%).

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

3.3 Discussion

We found that 4-year-old children succeeded at the RMTS task with horizontal configurations of stimuli, just as they did with vertical configurations in Experiment 1. As predicted, children who produced *same* and *different* terms to justify their choices demonstrated significantly better relational performance than those who did not. These findings are consistent with the idea that command of the relevant relational concepts is associated with better performance on the RMTS, and that spontaneous production of the corresponding relational terms may be a better predictor of performance on the RMTS task than age. Consistent with this suggestion, children's performance in our studies was high and remained stable throughout trials—suggesting that they likely possessed the *same/different* concepts prior to participating in the experiment.

Experiment 2 found that spontaneous relational language production predicted success on the task. However, this is qualified by further inspection that revealed that producing these words was merely correlated with performance, it was not necessary. For example, there were children who produced relational language but failed the RMTS trials and the reverse; there were children who did not produce relational language but succeeded on the RMTS trials.

4. General Discussion

The results of these two experiments provide clear evidence that 4-year-old children can succeed on the pure RMTS task—that is, the RMTS without training or feedback. These experiments also show that the production of *same/different* language during the justification trials predicts success on the task. We suggest that command of the relational concepts *same* and *different* plays a key role in success on the RMTS.

Prior work in this area has also found support for this idea. Hochmann et al. (2017) asked children to justify their response on the last two trials (a technique we adopted in the present studies). Despite some differences in methods and in the age of success, they also found that the use of *same/different* language predicted performance on the RMTS task, and concluded that the ability to recruit *same* and *different* concepts is an important factor in success on the RMTS. Christie and Gentner (2014) took a direct approach by teaching younger children the terms *same* and *different*. They gave 2- and 3-year-olds—who normally fail the *same*-only RMTS—a pretraining task in which they labeled pairs of items as “same” or “different” and received feedback. With this training, 3-year-old children went on to pass the *same*-only RMTS. Other developmental studies have also found effects of relational language on cognition (Gentner et al., 2013; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005; Pyers et al., 2010; Yuan et al., 2024).

However, we are not arguing that age is irrelevant. Indeed, it is possible that there are limitations on when the concepts *same* and *different* can be understood. For example, in Christie and Gentner’s (2014) training study, there was an age effect in grasping the *same-different* language. Whereas 3-year-olds succeeded in the naming task after an average of 11 pairs, nearly all 2-year-olds failed the naming task, even after 24 pairs (and also failed the RMTS task). Thus, there could be maturational limits on when such concepts are possible. However, recent findings call the idea of age limits into question. For example, Walker and colleagues have found that 18-month-old infants can learn *same-different* relations when given a causal learning task (Carstensen et al., 2019; Walker et al., 2016; Walker & Gopnik, 2014). Indeed, even infants under a year old have shown the ability to form *same* and *different* concepts (Anderson et al., 2018; Ferry et al., 2015).

At this juncture, it may seem puzzling that 3-year-olds fail the RMTS without special support, yet 7- and 9-month-old infants, and even 3-month-olds, can form abstract concepts of *same* and *different* (Anderson et al., 2018; Ferry et al., 2015; Hespos et al., 2021). The explanation lies in the degree of support given. The method in the infant studies is habituation to just one relation (either *same* or *different*), instantiated over a small set of exemplars. The infants require perfect spatial alignment, multiple repeated trials, and short retention for successful learning. Under these conditions, 7-, 9- and even 3-month-olds can form an abstraction of *same*, as evidenced by significantly looking longer at a novel *different* pair compared to a novel *same* pair (or the reverse, if habituated to *different* pairs). These findings suggest that the mechanism for relational alignment is available early and is capable of abstracting relations through a series of readily comparable examples. However, we believe these abstractions are quite fleeting in the infants; we have no evidence that they retain the *same* or *different* concept after the study.

~~In contrast, to~~Hochmann et al. (2017) found that 3- and 4-year-old children, like pigeons, fail 2-item array-match-to-sample (RMTS) tasks but succeed with 16-item arrays. Success by preschool children and pigeons on the 16-item display is attributed to encoding entropy (defined as the internal variability of an array). In contrast, older children and adults can detect the relation of same/different between pairs of objects. Although entropy-based accounts have been proposed (e.g., Glorioso et al., 2021), in the pure version of the RMTS task used in this study, there is strong evidence that success cannot be achieved using entropy perception, as the entropy of a *same* pair (0) and that of a *different* pair (1) are not sufficiently different to be detected given our method (i.e., presenting eight RMTS trials without practice or feedback). There is debate over whether a human or pigeon could be trained to detect this difference, but there is general agreement that any such training would require many hundreds or thousands of trials. There is an excellent discussion of this issue in Hochmann et al. (2017) (p. 19 and following).

To succeed on the pure RMTS requires that the child have well-established representations of *same* and *different* that can be spontaneously recruited to encode novel stimuli (Gentner et al., 2021). Consistent with this claim, we saw no learning across trials in the present studies, even with 16 items – suggesting that children who succeeded had the relational concept of *same* and *different* prior to the experiment. In other words, the RMTS task tests whether the relational concept has become part of one's cognitive arsenal, sufficiently entrenched as to be spontaneously accessible.

Our findings are parallel to those of Premack (1983). In his studies, three chimpanzees taught to use a physical symbol system (which included symbols for *same* and *different*) passed the RMTS, while two animals not trained on symbols failed (see also Thompson et al., 1997). Premack's studies led him to argue that language (or some other symbol system) is critical in permitting a chimpanzee to engage in "…abstract judgment, as in analogical reasoning, matching proportions of physically unlike exemplars, and completing incomplete (external) representations of action." (Premack, 1983, page 125). We suggest that this comment is equally true for our species.

This paper is unusual in its focus on the simple unsupported RMTS. Despite the interest in relational learning, there are very few RMTS studies that have no training. As reviewed here, among primates, only humans and chimpanzees with extensive symbol training have shown success in the pure RMTS task (Experiments 1 & 2; Christie & Gentner, 2010, 2014; Premack, 1983; Thompson & Oden, 2000). Experiment 1 and 2 show that production of relational language is the best proxy for possessing this concept.

In the larger context, it is possible to argue that living in a society that embraces literacy and numeracy could constitute relational training. With this view, it becomes possible to argue that, like Premack's chimpanzees, the children in our studies have already been given training in relational symbol system prior to coming to our labs. This continuity view is consistent with the idea that relational learning in infancy through adulthood operates by a single underlying structure-mapping mechanism and the varied results represent a continuum in the type of training necessary to form the abstract relation. This view suggests that the change in cognition across development is one of elaboration and refinement.

4.1 Summary

In sum, the present findings demonstrate that preschoolers can succeed on the RMTS task without training. Whether the pairs were presented vertically or horizontally, 4-year-olds chose the relational alternative at above-chance rates. A key conclusion from these studies is the importance of having accessible *same/different* concepts in supporting this insight. The results from the present studies converge with those of Hochmann and colleagues (2017) in underlining the powerful role of relational terms. William James (1890) stated that the "sense of sameness is the very keel and backbone of our thinking" (p. 459). The present findings help clarify when and how that comes about.

5. 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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *arXiv:1406.5823 [Stat]*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carstensen, A., Zhang, J., Heyman, G. D., Fu, G., Lee, K., & Walker, C. M. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences*, *116*(28), 13891–13896. https://doi.org/10.1073/pnas.1818365116

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. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

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

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

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

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), 1386–1405. https://doi.org/10.1111/cdev.12381

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), Article 1. https://doi.org/10.1037/0097-7403.33.1.55

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*(8), 650–657. https://doi.org/10.1037/amp0000082

Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*(3), 318–330. https://doi.org/10.1016/j.cognition.2013.01.003

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on Language and Thought* (1st ed., pp. 225–277). Cambridge University Press. https://www.cambridge.org/core/product/identifier/CBO9780511983689A017/type/book_part

Gentner, D., Shao, R., Simms, N., & Hespos, S. (2021). Learning same and different relations: Cross-species comparisons. *Current Opinion in Behavioral Sciences*, *37*, 84–89. https://doi.org/10.1016/j.cobeha.2020.11.013

Glorioso, G. C., Kuznar, S. L., Pavlic, M., & Povinelli, D. J. (2021). Still no solution to non-verbal measures of analogical reasoning: Reply to Walker and Gopnik (2017). *Cognition*, *214*, 104288. https://doi.org/10.1016/j.cognition.2020.104288

Green, C. T., Bunge, S. A., Briones Chiongbian, V., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143. https://doi.org/10.1016/j.jecp.2016.12.005

Hespos, S., Gentner, D., Anderson, E., & Shivaram, A. (2021). The origins of same/different discrimination in human infants. *Current Opinion in Behavioral Sciences, Same-Different Conceptualization*, *37*, 69–74. https://doi.org/10.1016/j.cobeha.2020.10.013

Hespos, S., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*(6998), Article 6998. https://doi.org/10.1038/nature02634

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

James, W. (1890). *The principles of psychology* (Vol. 2). Henry Holt and Company. https://doi.org/10.1037/11059-000

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187. https://doi.org/10.1007/s10339-013-0551-7

Kassambara, A. (2023). *rstatix: Pipe-friendly framework for basic statistical tests* (Version 0.7.2) [Computer software]. 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

Kuwabara, M., & Smith, L. B. (2012). Cross-cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology*, *113*(1), Article 1. https://doi.org/10.1016/j.jecp.2012.04.009

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. https://doi.org/10.1016/j.cogpsych.2004.09.004

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. 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*(5), 443–457. https://doi.org/10.1037/xhp0000726

Mix, K. S., Bower, C. A., Hancock, G. R., Crespo, S., & Smith, L. B. (2024). Place value learning in kindergarten. *Journal of Educational Psychology*, *116*(7), 1129–1152. https://doi.org/10.1037/edu0000866

Murphy, A. N., Zheng, Y., Shivaram, A., Vollman, E., & Richland, L. E. (2021). Bias and sensitivity to task constraints in spontaneous relational attention. *Journal of Experimental Child Psychology*, *202*, 104981. https://doi.org/10.1016/j.jecp.2020.104981

Posit team. (2025). *RStudio: Integrated development environment for R* [Computer software]. Posit Software, PBC. http://www.posit.co/

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

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*(27), 12116–12120. https://doi.org/10.1073/pnas.0914044107

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Simms, N. K., Matlen, B. J., Jee, B. D., & Gentner, D. (2023). Spatial alignment supports comparison of life science images. *Journal of Experimental Psychology: Applied*, *29*(4), 747–760. (2023-60337-001). https://doi.org/10.1037/xap0000471

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*(3), 626–638. https://doi.org/10.1016/j.cognition.2008.05.002

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, *24*(3), 363–396. https://doi.org/10.1016/S0364-0213(00)00029-X

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. https://doi.org/10.1037/0097-7403.23.1.31

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), 161–169. 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), 3–22. 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

Yuan, L., Novack, M., Uttal, D., & Franconeri, S. (2024). Language systematizes attention: How relational language enhances relational representation by guiding attention. *Cognition*, *243*, 105671. https://doi.org/10.1016/j.cognition.2023.105671

Zentall, T. R., Andrews, D. M., & Case, J. P. (2018). Sameness may be a natural concept that does not require learning. *Psychological Science*, *29*(7), 1185–1189. https://doi.org/10.1177/0956797618758669

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

Formatted: Line spacing: Double

Click here to download Research data
https://osf.io/rqwjk/?view_only=7056324f8a954251b492ddb11a444688

Revised manuscript (unmarked, editable source file) Click here to view linked References

The Role of Language in Forming the Concepts *Same* and *Different*

Manuscript submitted on April 25, 2026

Abstract

Relational ability – the ability to compare situations or ideas and discover common relations – is key to our capacity for higher-order cognition. The Relational-Match-to-Sample task (RMTS) is the gold standard to evaluate relational ability, yet recent reports conflict on when and how this ability emerges. Here, we focus on a version of the task in which children receive no training or feedback and ask whether possessing the relational concepts *same* and *different* is the key to success. We present two experiments that found 4-year-old children succeeded on three RMTS conditions aligned on the vertical axis (N = 79; 39 female and 40 male; 52% White; Wilcoxon's r >= .48) and one on the horizontal axis (N = 31; 15 female and 16 male; 73% White; r = .62). Across both experiments, children who justified their choices using relational language performed better on the RMTS task. We conclude that possession of the *same/different* concept is the key to success on the RMTS and production of relational language to justify performance is a better predictor than age for indicating possession of the concept.

Keywords: relational reasoning; analogy; early childhood; cognitive development; abstract learning

1. Introduction

A major component of human intelligence is our ability to perceive relations between objects, events, or ideas and transfer those relations to new situations. We use this ability every day. For example, you navigate through a new city by mentally aligning the image on your phone's map with real-life landmarks and streets. Relational ability is the vital foundation for effective learning, particularly in mathematics and science (e.g., Green et al., 2017; Jee et al., 2013; Mix et al., 2024). Thus, clarifying how relational abilities emerge and develop is key to understanding why some children learn more readily than others and how best to support this ability in all children.

To study relational ability at its core, we focus on *same* and *different* relations—the fundamental building blocks of relational matching. These relations are fundamental both for concept formation and everyday reasoning, are perceptually available to young children and nonhuman animals, and have been studied extensively across many species, offering the possibility of phylogenetic as well as ontogenetic comparisons (e.g., Christie & Gentner, 2014; Hochmann et al., 2017; Wasserman & Young, 2010). The gold standard for evaluating relational ability is the Relational-Match-To-Sample (RMTS) task (given AA, choose XX over YZ; given BC, choose YZ over XX) developed by Premack (1983). This task has been widely used to assess relational ability in both developmental (Carstensen et al., 2019; Christie & Gentner, 2014; Hochmann et al., 2017; Kroupin & Carey, 2022; Walker et al., 2016; Walker & Gopnik, 2014) and cross-species (Fagot & Thompson, 2011; Flemming et al., 2007; Premack, 1983; Thompson et al., 1997; Wasserman & Young, 2010) investigations.

In Premack's (1983) seminal paper, he proposed the RMTS as a true test of relational ability, distinguishing it from two other tasks that tap into *sameness* and *difference.* The first and easiest task is the Match-To-Sample task (MTS; Given A, choose A over B). This task simply requires detecting identical matches between objects; it can be passed by many species, including pigeons, macaques, honeybees, and baboons (Fagot & Thompson, 2011; Flemming et al., 2007; Thompson et al., 1997; Wasserman & Young, 2010; Zentall et al., 2018) as well as 10-month-old humans (Hochmann et al., 2016). The second task is the *same/different* task. In this task, the participant is first trained, with feedback, to classify a set of stimuli into *same* and *different* pairs (i.e., to make one response for AA, BB, CC etc. and a different response for AB, CD, EF etc.). If after mastering the training set, the participant can correctly classify novel pairs (e.g., XX, YZ etc.), they are considered to have passed the *same/different* task. Both the MTS and the *same/different* task involve training. However, the *same/different* task is more difficult than the MTS task, with fewer species achieving successful learning. For instance, pigeons can pass the MTS task but fail the 2-item *same/different* task (Wasserman & Young, 2010). An important distinction between the MTS and the *same/different* task in that success on the MTS task can be achieved by detecting object matches (X with X) whereas success on the *same/different* task requires the ability represent the relations *same* (XX) and *different* (YZ). Premack (1983) suggested that differences in the ability to form and use relational representations explain why fewer species succeed on the *same/different* task than on the MTS task.

The third test, the RMTS task was designed by Premack as a true test of the ability to recognize and match *same* and *different* relations. It is given after the subject has mastered the *same-different* task; the idea is that if a subject can fluently encode a new pair of items as *same* or *different*, then they will pass the RMTS task. It is ideal for developmental and cross-species comparisons because it does not require specialized content knowledge – only an ability to encode and match *same* and *different* relations. Premack hypothesized that possessing symbols for *same* and *different* was critical in passing this task – an idea that is consistent with many recent findings on the role of relational language in relational cognition (e.g., Gentner, 2016; Loewenstein & Gentner, 2005; Yuan et al., 2024). We return to this idea below.

Because of its importance as an indicator of relational ability, the RMTS has been used in a number of recent developmental studies. This research has considered a number of factors that contribute to children's performance on this task. Among these are pretraining children to label figures as *same* or *different* (Christie & Gentner, 2014), pretraining on a series of matching tasks (Kroupin & Carey, 2022, Experiment 2), embedding the RMTS within a causal *same-different* task (Carstensen et al., 2019; Walker et al., 2016; Walker & Gopnik, 2014), and training and/or feedback during the RMTS task (Hochmann et al., 2017). There is also evidence that cultural and linguistic factors influence performance (Carstensen et al., 2019; Murphy et al., 2021).

These studies of the supported RMTS task have provided valuable information about the paths through which children gain relational insight. However, our interest is in when children can pass the pure, unsupported RMTS. This is important, because true relational competence requires the ability to spontaneously notice and respond to relations in the world. We suggest that for this to happen, the child needs to have stored representations of *same* and *different* that can be accessed and used across contexts. A partial answer to the question of how children fare on the unsupported RMTS was provided by Christie and Gentner (2014, Experiment 1) who administered a *same*-only version of the RMTS task to 2.5-, 3.5-, and 4.5-year-old children. On all eight trials, children saw a standard that depicted the relation of *same* (e.g., two squares), followed by two alternatives—one depicting *same* and one depicting *different.* Children were asked to choose the alternative that was more like the standard. There were no training trials prior to the task, nor was any feedback provided during the task. The findings revealed that 4.5-year-olds succeeded and that neither 2.5- nor 3.5-year-olds did so—suggesting that children can pass the *same*-only RMTS by 4.5 years of age. However, these results do not tell us how children would fare on the full RMTS, which requires both *same* and *different* standards.

Two studies that used both *same* and *different* standards have somewhat different results. Kroupin and Carey (2022, Experiment 1) gave children eight trials of the pure RMTS task (that is, with neither pre-training nor feedback) with half the trials having a *same* standard and half a *different* standard. They found that 4.5-year-olds performed below chance. Kroupin and Carey (2022) did not include older children, but in a related study, Hochmann and colleagues (2017, Experiment S1, Appendix A) tested 3-, 4-, 5-, and 6-year-olds on a task with four *same* and four *different* trials, with corrective feedback on each test trial. The 5- and 6-year-olds succeeded, but children who were four and younger failed. In another experiment, Hochmann et al. (2017, Experiment 2) gave 4- and 5-year-old children the full RMTS task (four *same* trials and four *different* trials) with eight trials of pretraining before the test. The mean age of success was 5.5 years; again, 4-year-olds performed at chance. This discrepancy is striking given that the study by Hochmann et al. (2017) provided support—either pre-training or feedback—for the RMTS task.

One possible reason for this discrepancy is that Christie and Gentner (2014) only included *same* trials in their study, whereas the other studies used both *same* and *different* trials. As Hochmann et al. (2017, page 22) pointed out, success on the *same*-only task “could be solved by learning a single rule ‘choose *same’*…”. Therefore, in Experiment 1 we extended the methods in Christie and Gentner’s (2014) study by testing three between-subject conditions: eight *same* trials, eight *different* trials, and four *same*- plus four-*different* trials.

A second motivation for this study concerns the role of *same-different* language. We ask whether possession of the relational concepts of *same* and *different* is a predictor of success in the RMTS task. To assess understanding of *same* and *different*, we adopted a technique used by Hochmann et al. (2017). On the last two trials of the RMTS, we asked children to explain their choice. Children who used the concepts *same* and/or *different* in their justifications were considered to have demonstrated productive command of the corresponding concepts. These manipulations will clarify whether 4-year-olds have the capacity to detect and match both *same* and *different* relations, and whether this capacity is associated with productive command of the corresponding linguistic terms or age.

2. Experiment 1

The goal of this experiment is to examine (1) whether children can succeed on the RMTS task with standards depicting *same* and *different* relations and no training or feedback; and (2) whether justifying their answers with ‘same’ and ‘different’ predicts success on the RMTS. We tested three conditions: *same*-only, *different*-only, and *full* (half *same* and half *different* trials). Experiments 1 and 2’s design and analyses were not preregistered. All materials, data, and analysis code for both experiments have been made publicly available at Open Science Framework and can be accessed through this link.

2.1 Methods

2.1.1 *Participants*

Before beginning data collection, the 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 effect size was estimated based on the effect size (*d*) derived from the performance of 4-year-olds who achieved success at above chance levels on Christie and Gentner's (2014) RMTS task. This study was approved by the Institutional Review Board at a major US University (protocol number withheld for review).

Participants were 79 4-year-old children, approximately evenly split between the three conditions (M = 4 years 5.5 months, range = 3 years 11 months to 4 years 11 months). For a detailed breakdown of participant demographics, see Supplementary Materials, Section 1.1.1. An additional 14 children were excluded from statistical analyses due to failure to pass all three catch trials (described in Section 2.1.2 below) (n = 10), experimenter error (n = 2), refusal/non-compliance (n = 1), or parental interference (n = 1).

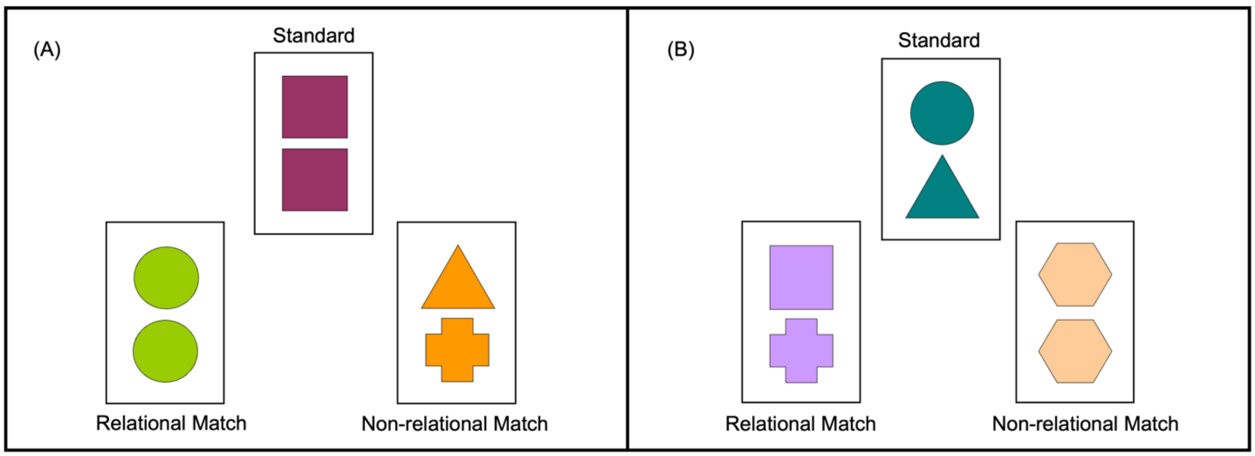
Children for this and the following experiment were recruited using a voluntary participant pool or through local preschools in a major metropolitan city in the United States. All participants received a small gift for their participation. Across both experiments, the ethnic makeup was approximately 61% non-Latine, 13% Latine, and 26% provided no response. Approximately 56% were White, 12% were multiracial, 5% were Asian, 3% were Black/African American, 1% were Indigenous American, and the remaining 23% had no response. For the two experiments, approximately 73% of primary caregivers had a college degree or higher, 2% had some college experience, 1% had a high school diploma, and the remaining 24% had no response. Among the secondary caregivers, 65% had a college degree or higher, 5% had some college experience, 2% had some high school education, and 28% had no response.

2.1.2 *Materials*

The materials were identical to those in Christie and Gentner's (2014) study, except for the additional cards required to have *different* standards. There were three conditions, each with eight trials: *same*-only, *different*-only, and *full* RMTS (four *same* and four *different*) (Figure 1). See the Supplementary Materials, Section 1.1.1 for more details.

Figure 1

Sample Triads for the RMTS Task Used in Experiment 1 Depicting (A) Same and (B) Different Relations. For the Full Version of the RMTS Task, Half the Standards Depicted the Same Relation and the Other Half Depicted Different.



In addition to the eight RMTS trials, three catch/MTS trials were given to verify that children understood the task: e.g., a blue fish as the standard with a similar red fish and a yellow cup as alternatives. After the catch trials, we repeated the final two RMTS triad trials where the child was asked to justify why they made the choice they did.

2.1.3 *Procedure*

After obtaining parental consent and the child's verbal assent, the experimenter invited the child to a quiet room. The child was seated at a table 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 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. No feedback was given apart from general positive encouragement. After each trial, the experimenter noted the child's response on a sheet of paper. When the triad trials were complete, the three catch trials were administered using the same instructions as the eight RMTS trials.

Next, children were asked to justify their choice on the last two triad trials. For these two justification trials, the trial was repeated and once the child indicated their response, the experimenter removed the non-chosen card and asked, "Why is this picture more like this one?" while gesturing to the two remaining cards.

2.1.4 *Coding*

For each participant, the primary coder scored all eight trials of the RMTS task, the three catch trials and the two justification trials. In addition, 40% of the justification trials were double coded for whether the child used the term *same* and/or *different* (e.g., "Because they are same shapes" or "Because it has 2 – because it's both different shapes"). The two coders were unaware of the child's RMTS response. Reliability was 98%, resulting in a Cohen's kappa of .94. Disagreements between coders were resolved through discussion.

2.2 Results

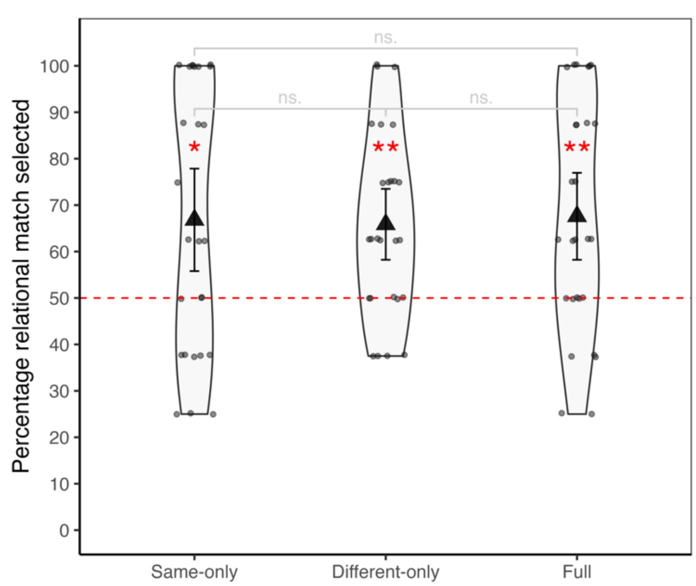
All statistical analyses in the current study were performed using R and R Studio (Posit team, 2025) using the 'tidyverse' (Wickham et al., 2019), 'stats' (R Core Team, 2025), 'rstatix' (Kassambara, 2023), and 'lme4' (Bates et al., 2015) packages. A complete list of packages used for data cleaning and visualization can be found in the R script in this OSF repository. This repository also contains an HTML report of analyses conducted for this study with complete regression tables.

2.2.1 *Performance on the RMTS task*

The first question asked whether 4-year-olds succeeded on the RMTS task. To answer this question, we conducted Wilcoxon one-sample tests and calculated the corresponding effect size (Wilcoxon's r), as the dependent variable (percentage of relational match selected) showed evidence of non-normality based on Shapiro–Wilk tests (*same*-only: $W = .86$, $p = .002$; *different*-only: $W = .93$, $p = .07$; *full*: $W = .91$, $p = .03$). The answer is yes: Children's performance was significantly above chance in all three conditions of the RMTS task: *same*-only ($M = 66.83\%$, $SD = 28.71$, $Z = 222$, $p = .03$ with Bonferroni correction, Wilcoxon's $r = .48$), *different*-only ($M = 65.87\%$, $SD = 19.86$, $Z = 209$, $p = .003$ with Bonferroni correction, $r = .64$), and *full* ($M = 67.59\%$, $SD = 24.82$, $Z = 218.5$, $p = .008$ with Bonferroni correction, $r = .57$; see Figure 2).

Figure 2

Percentage of Relational Match Selected in Experiment 1.



Note. 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 various points of the dependent variable. * $p < .05$, ** $p < .01$, *** $p < .001$. The red dashed line indicates chance performance.

To examine whether children perform comparably across the three conditions, we conducted a multiple regression predicting the percentage of relational matches selected from condition (*same*-only, *different*-only, or *full*), age (in months), gender (female or male), and order of presentation (five orders). The overall regression model was not statistically significant, R^2 = .04, R^2_{adj} = -.07, $F(8, 69) = .33$, $p = .95$, indicating that the predictors collectively did not account for variation in performance. Holding all other predictors constant, neither the *different*-only nor *full* conditions differed from the *same*-only condition (*different*-only: b = –2.26, 95% CI [–16.60, 12.07], $t(69)$ = –.31, p = .754, β = –.09, 95% CI [–.68, .50]; *full*: b = .14, 95% CI [–14.30, 14.59], $t(69)$ = .02, p = .984, β = .01, 95% CI [–.59, .60]). A derived pairwise comparison between the *different*-only and full conditions, obtained by re-leveling the factor, was also nonsignificant, b = 2.41, 95% CI [–12.19, 17.00], $t(69)$ = .33, p = .743, β = .10, 95% CI [–.50, .70]. None of the remaining predictors were significant (age: b = .45, p = .68, β = .05, 95% CI [–.21, .31]; gender: b = 2.20, p = .71, β = .09, 95% CI [–.39, .57]; order: all dummy-coded contrasts ps > .41, $|\beta|$s < .28). These results indicate that children performed comparably across all three RMTS conditions and that age was not a significant predictor of relational performance.

2.2.2 *Effects of Relational Language*

Our second goal was to test whether production of the terms ‘same’ and ‘different’ to justify their choice would predict success on the RMTS task. We conducted a multiple regression with percentage of relational match selected as the outcome variable and the number of relational language justifications provided (0, 1, or 2 of 2 trials), condition (*same*-only, *different*-only, or *full*), age (in months), and gender (female or male) as predictors. Using the relational terms on both justification trials, relative to justifying their choices with no relational terms was marginal, but not a significant predictor of children's relational performance on the RMTS task when holding all other predictors constant, b = 14.33, 95% CI [-1.69, 30.35], $t(68)$ = 1.79, p = .079, β = .58, 95% CI [–.07, 1.23]. Using *same* and *different* relational terms on one justification trial, relative to none, was also not a significant predictor of children's relational performance on the RMTS task, b = .65, 95% CI [-14.98, 16.28], $t(68)$ = .08, p = .934, β = .03, 95% CI [–.61, .66]. The descriptive statistics lend further details. Children who did not produce relational language showed lower accuracy and considerable variability (M = 63.8%, SD = 22.5%), including some who performed well on the task. Children who produced relational language on both trials had higher accuracy, but also showed considerable variability (M = 79.2%, SD = 26.6%), including some who failed RMTS trials.

2.3 Discussion

These data provide clear evidence that children succeed on the RMTS task across all three condition types (*same* only, *different* only, and mixed *same* and *different* standards). Further, performance was similar across the three conditions; the mean proportion correct ranged from 65.9% to 67.6%. However, we did not find the predicted relation between spontaneous production of the *same* and *different* terms and performance on the RMS task. One possible factor here is that unlike Hochmann et al. (2017), this experiment had only eight trials. It is possible that with more trials, children might have found it easier to produce the relevant relational terms. This idea was tested in Experiment 2.

How do we reconcile the finding that children pass the RMTS at 4 years of age with the negative results from Hochmann et al. (2017) and Kroupin and Carey (2022)? There are differences in the materials and procedures that could have affected the results. Here we highlight two of these differences. Additional differences and a table comparing experimental details across several RMTS tasks is provided in the supplementary materials (see Section 1.2.1). One difference that stands out is in spatial alignability. There is evidence that spatial alignment fosters relational alignment, both in adults (Matlen et al., 2020; Simms et al., 2023) and in children (Zheng et al., 2022). In our Experiment 1, all the pairs had the same (vertical) spatial configuration, whereas in Hochmann et al.'s (2017) studies, the pairs within a triad varied in their spatial configuration, making spatial alignment more difficult (Figure 3). In Kroupin and Carey's (2022) study, the figures were spatially aligned, but the configuration within a pair was horizontal, rather than vertical as in the current studies (Figure 4). To test whether the horizontal configuration diminishes performance, in Experiment 2 we presented the objects with a horizontal configuration.

Figure 3

A Sample Triad of the (A) Same and (B) Different Relations of the Full RMTS Task as Used in Hochmann et al. (2017).

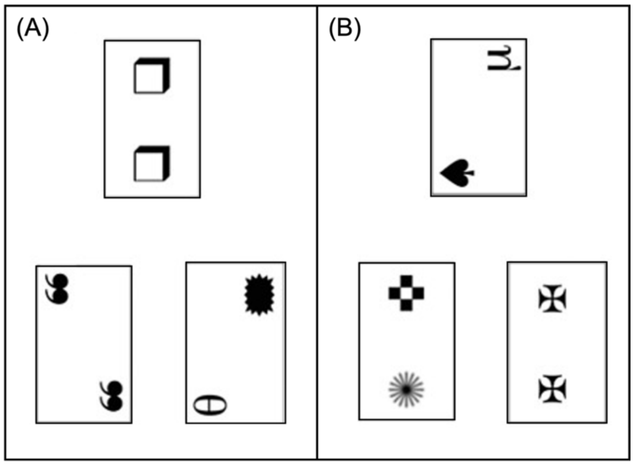
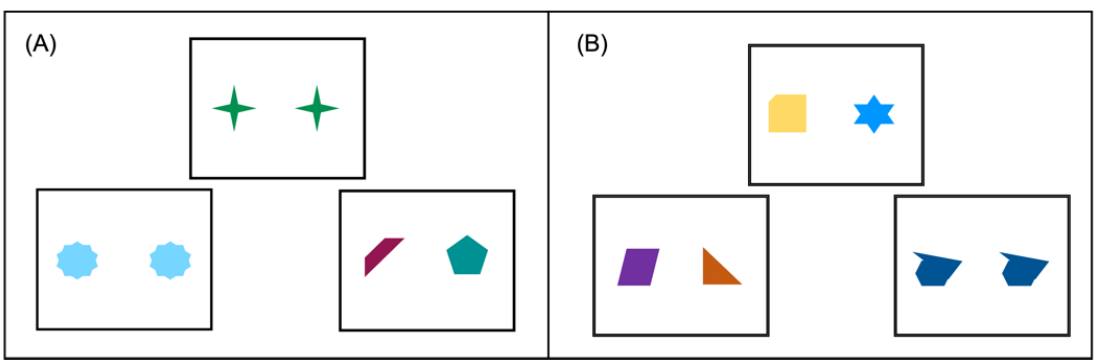


Figure 4

A Sample Triad of the (A) Same and (B) Different Relations of the RMTS Task as Used in Kroupin and Carey (2022).



A second difference is the kinds of shapes that make up the pairs. In the current studies, the pairs were made up of simple shapes such as triangles, circles, and squares, which are familiar to children (Clements et al., 1999). The shapes were repeated across trials, though always with different colors and in different positions. In contrast, Hochmann et al. (2017) and Kroupin and Carey (2022) used unique shapes throughout. Consequently, many of the individual shapes used in the stimuli were complex and unfamiliar to 4-year-olds. Prior studies have found that children perform worse on relational tasks when the objects instantiating the relations are complex rather than simple (Gentner & Rattermann, 1991; Hespos & Spelke, 2004; Kuwabara & Smith, 2012; Son et al., 2008) and novel rather than familiar (Anderson et al., 2018; Casasola, 2005; Casasola & Park, 2013; Maguire et al., 2008; Son et al., 2008).

In summary, there are many factors that could have affected performance. However, in light of the variability across labs, we wanted to ensure that our results would generalize across a different spatial configuration. Therefore, in Experiment 2, we used horizontal rather than vertical pairs. In addition, we increased the number of trials from 8 to 16, enabling us to detect possible learning across trials.

3. Experiment 2

The aim of Experiment 2 was to investigate whether preschoolers can succeed on the RMTS task given horizontally arranged stimuli and more trials. Experiment 2 was identical to the Experiment 1 condition that had both *same* and *different* standards. There were two changes: First, we modified the materials such that the shapes within each card were aligned on a horizontal axis rather than a vertical axis. Second, we ran 16 trials.

3.1 Methods

3.1.1 *Participants*

Participants were 31 4-year-old children (15 female, 16 male; M = 4;7, range = 4;3 to 5;0). An additional 10 children were excluded from analysis for failing one or more catch trials. To increase the power when compared to Experiment 1, we conducted an a priori power analysis with the goal of obtaining a power of .90 with an alpha level of $p < .05$ and an effect size of d =

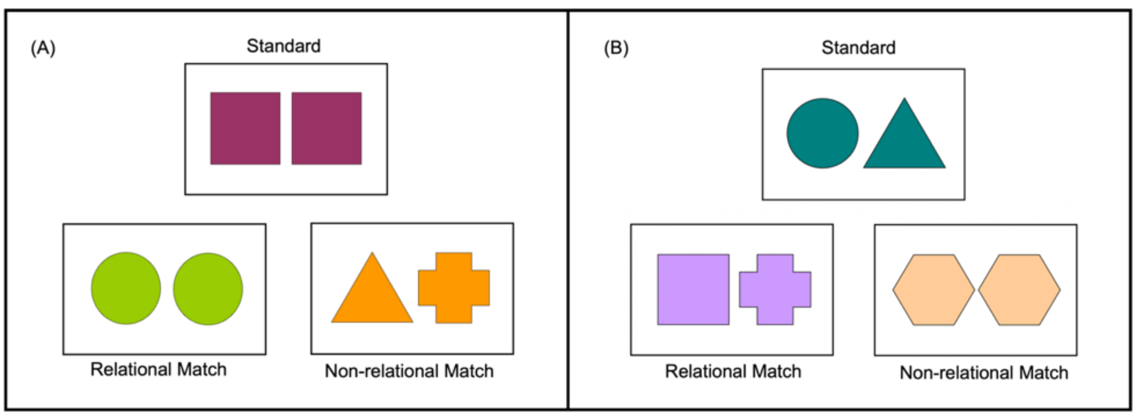
.60 (Faul et al., 2007). The resulting target sample size is approximately 32 children. This study was approved by the Institutional Review Board at a major US University (protocol number withheld for review). Families were given $20 as compensation for their participation and a child-sized T-shirt.

3.1.2 *Materials*

The materials were identical to those of Experiment 1 with two modifications. First, the stimuli were placed on a horizontal axis (Figure 5). Second, there were 16 trials.

Figure 5

Sample Triads from Experiment 2 with Stimuli Placed in a Horizontal Orientation. (A) Half the Standards Depicted the Same Relation (B) and the Rest were Different.



3.1.3 *Procedure*

The procedure was identical to that of Experiment 1. The only exception was that due to a mistake at the beginning of data collection, one participant was not administered any justification trials and five participants were administered only one of the two justification trials.

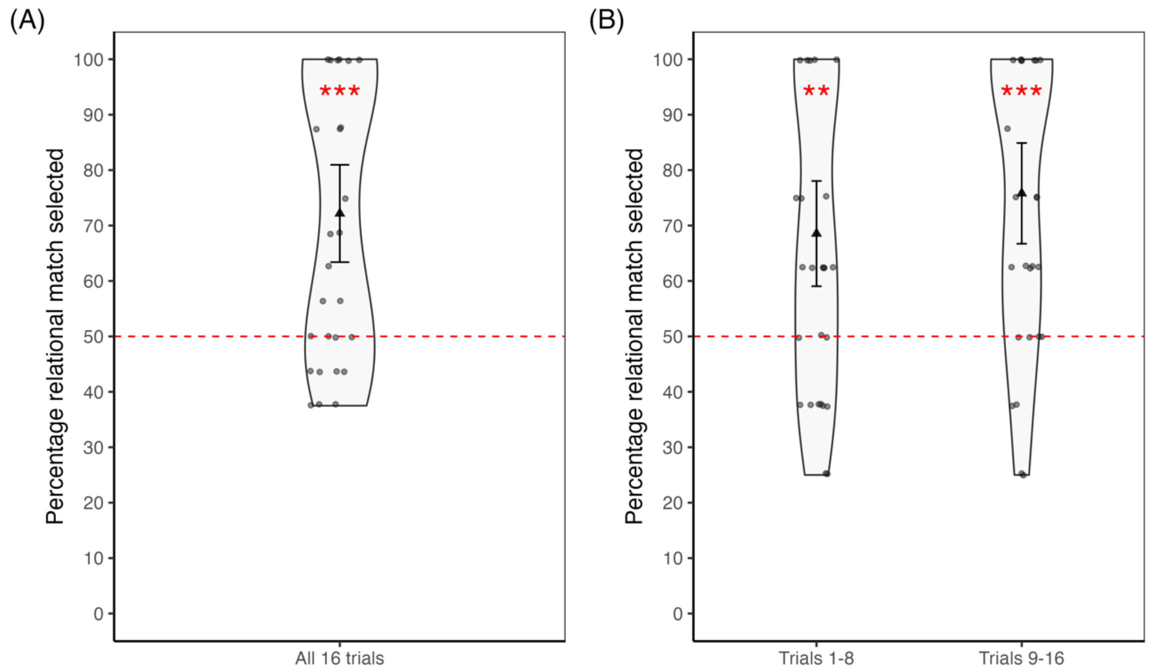
3.2 Results

3.2.1 *Performance on the RMTS task*

Similar to Experiment 1, we conducted a Wilcoxon's one-sample test and calculated its corresponding effect size (Wilcoxon's *r*) as the dependent variable (percentage of relational match selected) was not normally distributed. Again, 4-year-old children succeed on the RMTS task. Children performed at significantly above-chance levels (M = 72.18%, SD = 24.94, Z = 338.5, $p < .001$, Wilcoxon's $r = .62$, see Figure 6A). Examining each set of eight trials in turn, children performed at significantly above chance on the first set of eight trials (M = 68.55%, SD = 26.98, Z = 337, $p = .004$, Wilcoxon's $r = .88$) as well as on the second set of eight trials (M = 75.81%, SD = 25.81, Z = 350, $p < .001$, Wilcoxon's $r = .88$) (Figure 6B).

Figure 6

Percentage of Relational Match Selected in Experiment 2 (A) Across all 16 Trials and (B) Across Trials 1-8 and 9-16.



We also asked whether children's performance would be stable across trials of the RMTS task. We conducted a linear mixed-effects regression model with average percentage of relational match selected as the dependent variable, fixed effects of trial (first three or last three), age, and gender with random effects by participant. 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, b = 5.38, 95% CI [-2.87, 13.63], $t(56)$ = 1.31, p = .197, β = .17, 95% CI [-.09, .43] (see Table S4 in Supplementary Materials, Section 1.2). For comparability with prior studies which used 8 trials, we also analyzed the data from the first eight trials only. Here too, performance was comparable on the first three and the last three trials, b = 3.23, 95% CI [-7.66, 14.11], $t(56)$ = .59, p = .555, β = .10, 95% CI [-.24, .44] (see Supplementary Materials, Section 1.2, Table S4).

Analyses were also performed on children's performance on *same* vs. *different* trials and comparisons across Experiments 1 and 2. There were no significant differences in relational performance in any of these analyses (see Section 1.2 in Supplementary Materials).

3.2.2 *Effects of Relational Language*

Children who used the relational terms *same* and *different* in justifying their choices also had high accuracy on the RMTS task. We conducted a linear regression with percentage of relational match selected as the outcome variable and the number of *same-different* justifications (0, 1, or 2), age (in months), and gender (female or male) as predictors. Using the relational terms *same* and *different* on both justification trials, relative to justifying their choices with no relational terms, was a significant predictor of children's relational performance on the RMTS task, $b = 44.60$, 95% CI [23.89, 65.31], $t(26) = 4.43$, $p < .001$, $\beta = 1.79$, 95% CI [.96, 2.62]. Similarly, using *same* and *different* relational terms on one justification trial, relative to none, was also a significant predictor of children's relational performance on the RMTS task, $b = 29.66$, 95% CI [12.06, 47.25], $t(26) = 3.47$, $p = .002$, $\beta = 1.19$, 95% CI [.48, 1.89]. The descriptive statistics lend further details. Children who did not produce relational language revealed lower accuracy and considerable variability ($M = 58.0\%$, $SD = 20.4\%$) including some who performed well on the task. In contrast, children who produced relational language on both trials had a much higher accuracy and very little variability ($M = 97.9\%$, $SD = 5.1\%$).

3.3 Discussion

We found that 4-year-old children succeeded at the RMTS task with horizontal configurations of stimuli, just as they did with vertical configurations in Experiment 1. As predicted, children who produced *same* and *different* terms to justify their choices demonstrated significantly better relational performance than those who did not. These findings are consistent with the idea that command of the relevant relational concepts is associated with better performance on the RMTS, and that spontaneous production of the corresponding relational terms may be a better predictor of performance on the RMTS task than age. Consistent with this suggestion, children's performance in our studies was high and remained stable throughout trials—suggesting that they likely possessed the *same/different* concepts prior to participating in the experiment.

Experiment 2 found that spontaneous relational language production predicted success on the task. However, this is qualified by further inspection that revealed that producing these words was merely correlated with performance, it was not necessary. For example, there were children who produced relational language but failed the RMTS trials and the reverse; there were children who did not produce relational language but succeeded on the RMTS trials.

4. General Discussion

The results of these two experiments provide clear evidence that 4-year-old children can succeed on the pure RMTS task—that is, the RMTS without training or feedback. These experiments also show that the production of *same/different* language during the justification trials predicts success on the task. We suggest that command of the relational concepts *same* and *different* plays a key role in success on the RMTS.

Prior work in this area has also found support for this idea. Hochmann et al. (2017) asked children to justify their response on the last two trials (a technique we adopted in the present studies). Despite some differences in methods and in the age of success, they also found that the use of *same/different* language predicted performance on the RMTS task, and concluded that the ability to recruit *same* and *different* concepts is an important factor in success on the RMTS. Christie and Gentner (2014) took a direct approach by teaching younger children the terms *same* and *different*. They gave 2- and 3-year-olds—who normally fail the *same*-only RMTS—a pretraining task in which they labeled pairs of items as “same” or “different” and received feedback. With this training, 3-year-old children went on to pass the *same*-only RMTS. Other developmental studies have also found effects of relational language on cognition (Gentner et al., 2013; Gentner & Rattermann, 1991; Loewenstein & Gentner, 2005; Pyers et al., 2010; Yuan et al., 2024).

However, we are not arguing that age is irrelevant. Indeed, it is possible that there are limitations on when the concepts *same* and *different* can be understood. For example, in Christie and Gentner’s (2014) training study, there was an age effect in grasping the *same-different* language. Whereas 3-year-olds succeeded in the naming task after an average of 11 pairs, nearly all 2-year-olds failed the naming task, even after 24 pairs (and also failed the RMTS task). Thus, there could be maturational limits on when such concepts are possible. However, recent findings call the idea of age limits into question. For example, Walker and colleagues have found that 18-month-old infants can learn *same-different* relations when given a causal learning task (Carstensen et al., 2019; Walker et al., 2016; Walker & Gopnik, 2014). Indeed, even infants under a year old have shown the ability to form *same* and *different* concepts (Anderson et al., 2018; Ferry et al., 2015).

At this juncture, it may seem puzzling that 3-year-olds fail the RMTS without special support, yet 7- and 9-month-old infants, and even 3-month-olds, can form abstract concepts of *same* and *different* (Anderson et al., 2018; Ferry et al., 2015; Hespos et al., 2021). The explanation lies in the degree of support given. The method in the infant studies is habituation to just one relation (either *same* or *different*), instantiated over a small set of exemplars. The infants require perfect spatial alignment, multiple repeated trials, and short retention for successful learning. Under these conditions, 7-, 9- and even 3-month-olds can form an abstraction of *same*, as evidenced by significantly looking longer at a novel *different* pair compared to a novel *same* pair (or the reverse, if habituated to *different* pairs). These findings suggest that the mechanism for relational alignment is available early and is capable of abstracting relations through a series of readily comparable examples. However, we believe these abstractions are quite fleeting in the infants; we have no evidence that they retain the *same* or *different* concept after the study.

Hochmann et al. (2017) found that 3- and 4-year-old children, like pigeons, fail 2-item array-match-to-sample (RMTS) tasks but succeed with 16-item arrays. Success by preschool children and pigeons on the 16-item display is attributed to encoding entropy (defined as the internal variability of an array). In contrast, older children and adults can detect the relation of same/different between pairs of objects. Although entropy-based accounts have been proposed (e.g., Glorioso et al., 2021), in the pure version of the RMTS task used in this study, there is strong evidence that success cannot be achieved using entropy perception, as the entropy of a *same* pair (0) and that of a *different* pair (1) are not sufficiently different to be detected given our method (i.e., presenting eight RMTS trials without practice or feedback). There is debate over whether a human or pigeon could be trained to detect this difference, but there is general agreement that any such training would require many hundreds or thousands of trials. There is an excellent discussion of this issue in Hochmann et al. (2017) (p. 19 and following).

To succeed on the pure RMTS requires that the child have well-established representations of *same* and *different* that can be spontaneously recruited to encode novel stimuli (Gentner et al., 2021). Consistent with this claim, we saw no learning across trials in the present studies, even with 16 items – suggesting that children who succeeded had the relational concept of *same* and *different* prior to the experiment. In other words, the RMTS task tests whether the relational concept has become part of one's cognitive arsenal, sufficiently entrenched as to be spontaneously accessible.

Our findings are parallel to those of Premack (1983). In his studies, three chimpanzees taught to use a physical symbol system (which included symbols for *same* and *different*) passed the RMTS, while two animals not trained on symbols failed (see also Thompson et al., 1997). Premack's studies led him to argue that language (or some other symbol system) is critical in permitting a chimpanzee to engage in "…abstract judgment, as in analogical reasoning, matching proportions of physically unlike exemplars, and completing incomplete (external) representations of action." (Premack, 1983, page 125). We suggest that this comment is equally true for our species.

This paper is unusual in its focus on the simple unsupported RMTS. Despite the interest in relational learning, there are very few RMTS studies that have no training. As reviewed here, among primates, only humans and chimpanzees with extensive symbol training have shown success in the pure RMTS task (Experiments 1 & 2; Christie & Gentner, 2010, 2014; Premack, 1983; Thompson & Oden, 2000). Experiment 1 and 2 show that production of relational language is the best proxy for possessing this concept.

In the larger context, it is possible to argue that living in a society that embraces literacy and numeracy could constitute relational training. With this view, it becomes possible to argue that, like Premack's chimpanzees, the children in our studies have already been given training in relational symbol system prior to coming to our labs. This continuity view is consistent with the idea that relational learning in infancy through adulthood operates by a single underlying structure-mapping mechanism and the varied results represent a continuum in the type of training necessary to form the abstract relation. This view suggests that the change in cognition across development is one of elaboration and refinement.

4.1 Summary

In sum, the present findings demonstrate that preschoolers can succeed on the RMTS task without training. Whether the pairs were presented vertically or horizontally, 4-year-olds chose the relational alternative at above-chance rates. A key conclusion from these studies is the importance of having accessible *same/different* concepts in supporting this insight. The results from the present studies converge with those of Hochmann and colleagues (2017) in underlining the powerful role of relational terms. William James (1890) stated that the "sense of sameness is the very keel and backbone of our thinking" (p. 459). The present findings help clarify when and how that comes about.

5. 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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *arXiv:1406.5823 [Stat]*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carstensen, A., Zhang, J., Heyman, G. D., Fu, G., Lee, K., & Walker, C. M. (2019). Context shapes early diversity in abstract thought. *Proceedings of the National Academy of Sciences*, *116*(28), 13891–13896. https://doi.org/10.1073/pnas.1818365116

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. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373. https://doi.org/10.1080/15248371003700015

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

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

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

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), 1386–1405. https://doi.org/10.1111/cdev.12381

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), Article 1. https://doi.org/10.1037/0097-7403.33.1.55

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*(8), 650–657. https://doi.org/10.1037/amp0000082

Gentner, D., Özyürek, A., Gürcanli, Ö., & Goldin-Meadow, S. (2013). Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition*, *127*(3), 318–330. https://doi.org/10.1016/j.cognition.2013.01.003

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on Language and Thought* (1st ed., pp. 225–277). Cambridge University Press. https://www.cambridge.org/core/product/identifier/CBO9780511983689A017/type/book_part

Gentner, D., Shao, R., Simms, N., & Hespos, S. (2021). Learning same and different relations: Cross-species comparisons. *Current Opinion in Behavioral Sciences*, *37*, 84–89. https://doi.org/10.1016/j.cobeha.2020.11.013

Glorioso, G. C., Kuznar, S. L., Pavlic, M., & Povinelli, D. J. (2021). Still no solution to non-verbal measures of analogical reasoning: Reply to Walker and Gopnik (2017). *Cognition*, *214*, 104288. https://doi.org/10.1016/j.cognition.2020.104288

Green, C. T., Bunge, S. A., Briones Chiongbian, V., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143. https://doi.org/10.1016/j.jecp.2016.12.005

Hespos, S., Gentner, D., Anderson, E., & Shivaram, A. (2021). The origins of same/different discrimination in human infants. *Current Opinion in Behavioral Sciences, Same-Different Conceptualization*, *37*, 69–74. https://doi.org/10.1016/j.cobeha.2020.10.013

Hespos, S., & Spelke, E. S. (2004). Conceptual precursors to language. *Nature*, *430*(6998), Article 6998. https://doi.org/10.1038/nature02634

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

James, W. (1890). *The principles of psychology* (Vol. 2). Henry Holt and Company. https://doi.org/10.1037/11059-000

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: Analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing*, *14*(2), 175–187. https://doi.org/10.1007/s10339-013-0551-7

Kassambara, A. (2023). *rstatix: Pipe-friendly framework for basic statistical tests* (Version 0.7.2) [Computer software]. 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

Kuwabara, M., & Smith, L. B. (2012). Cross-cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology*, *113*(1), Article 1. https://doi.org/10.1016/j.jecp.2012.04.009

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. https://doi.org/10.1016/j.cogpsych.2004.09.004

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. 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*(5), 443–457. https://doi.org/10.1037/xhp0000726

Mix, K. S., Bower, C. A., Hancock, G. R., Crespo, S., & Smith, L. B. (2024). Place value learning in kindergarten. *Journal of Educational Psychology*, *116*(7), 1129–1152. https://doi.org/10.1037/edu0000866

Murphy, A. N., Zheng, Y., Shivaram, A., Vollman, E., & Richland, L. E. (2021). Bias and sensitivity to task constraints in spontaneous relational attention. *Journal of Experimental Child Psychology*, *202*, 104981. https://doi.org/10.1016/j.jecp.2020.104981

Posit team. (2025). *RStudio: Integrated development environment for R* [Computer software]. Posit Software, PBC. http://www.posit.co/

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

Pyers, J. E., Shusterman, A., Senghas, A., Spelke, E. S., & Emmorey, K. (2010). Evidence from an emerging sign language reveals that language supports spatial cognition. *Proceedings of the National Academy of Sciences*, *107*(27), 12116–12120. https://doi.org/10.1073/pnas.0914044107

R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Simms, N. K., Matlen, B. J., Jee, B. D., & Gentner, D. (2023). Spatial alignment supports comparison of life science images. *Journal of Experimental Psychology: Applied*, *29*(4), 747–760. (2023-60337-001). https://doi.org/10.1037/xap0000471

Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*(3), 626–638. https://doi.org/10.1016/j.cognition.2008.05.002

Thompson, R. K. R., & Oden, D. L. (2000). Categorical perception and conceptual judgments by nonhuman primates: The paleological monkey and the analogical ape. *Cognitive Science*, *24*(3), 363–396. https://doi.org/10.1016/S0364-0213(00)00029-X

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. https://doi.org/10.1037/0097-7403.23.1.31

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), 161–169. 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), 3–22. 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

Yuan, L., Novack, M., Uttal, D., & Franconeri, S. (2024). Language systematizes attention: How relational language enhances relational representation by guiding attention. *Cognition*, *243*, 105671. https://doi.org/10.1016/j.cognition.2023.105671

Zentall, T. R., Andrews, D. M., & Case, J. P. (2018). Sameness may be a natural concept that does not require learning. *Psychological Science*, *29*(7), 1185–1189. https://doi.org/10.1177/0956797618758669

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

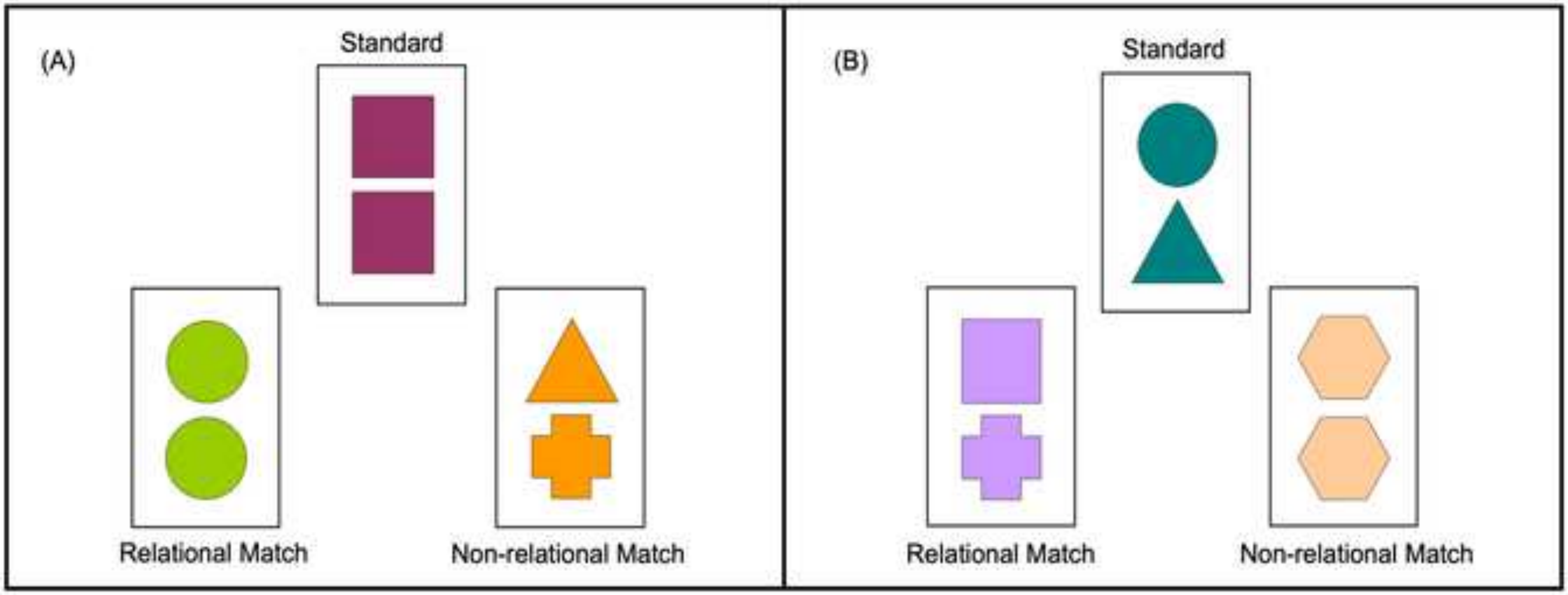


Figure 1

Sample Triads for the RMTS Task Used in Experiment 1 Depicting (A) Same and (B) Different Relations. For the Full Version of the RMTS Task, Half the Standards Depicted the Same Relation and the Other Half Depicted Different.

Figure 2

Click here to access/download;Figure;Figure 2.png

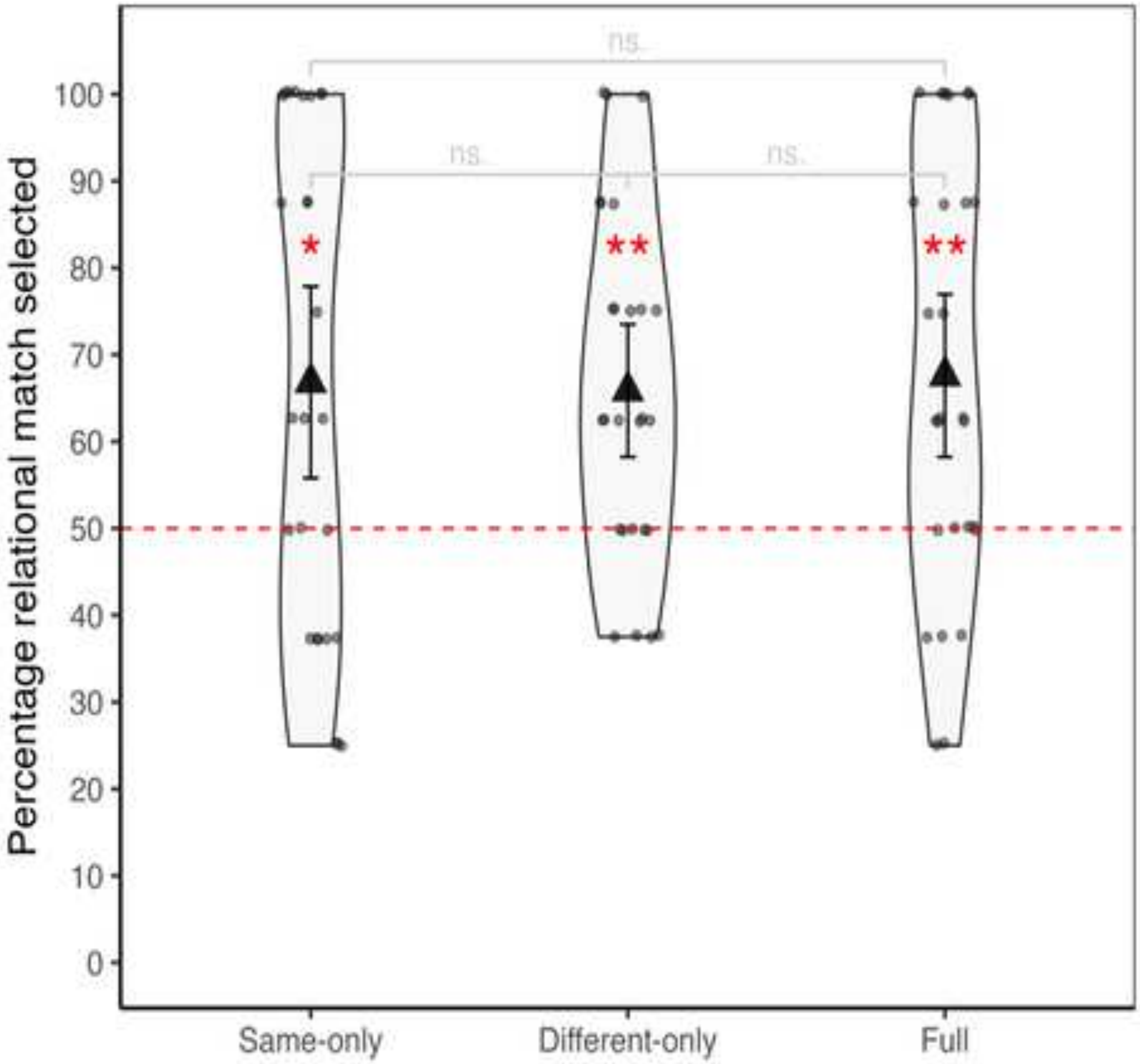


Figure 2

Percentage of Relational Match Selected in Experiment 1.
Note. 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 various points of the dependent variable.
* $p < .05$, ** $p < .01$, *** $p < .001$.

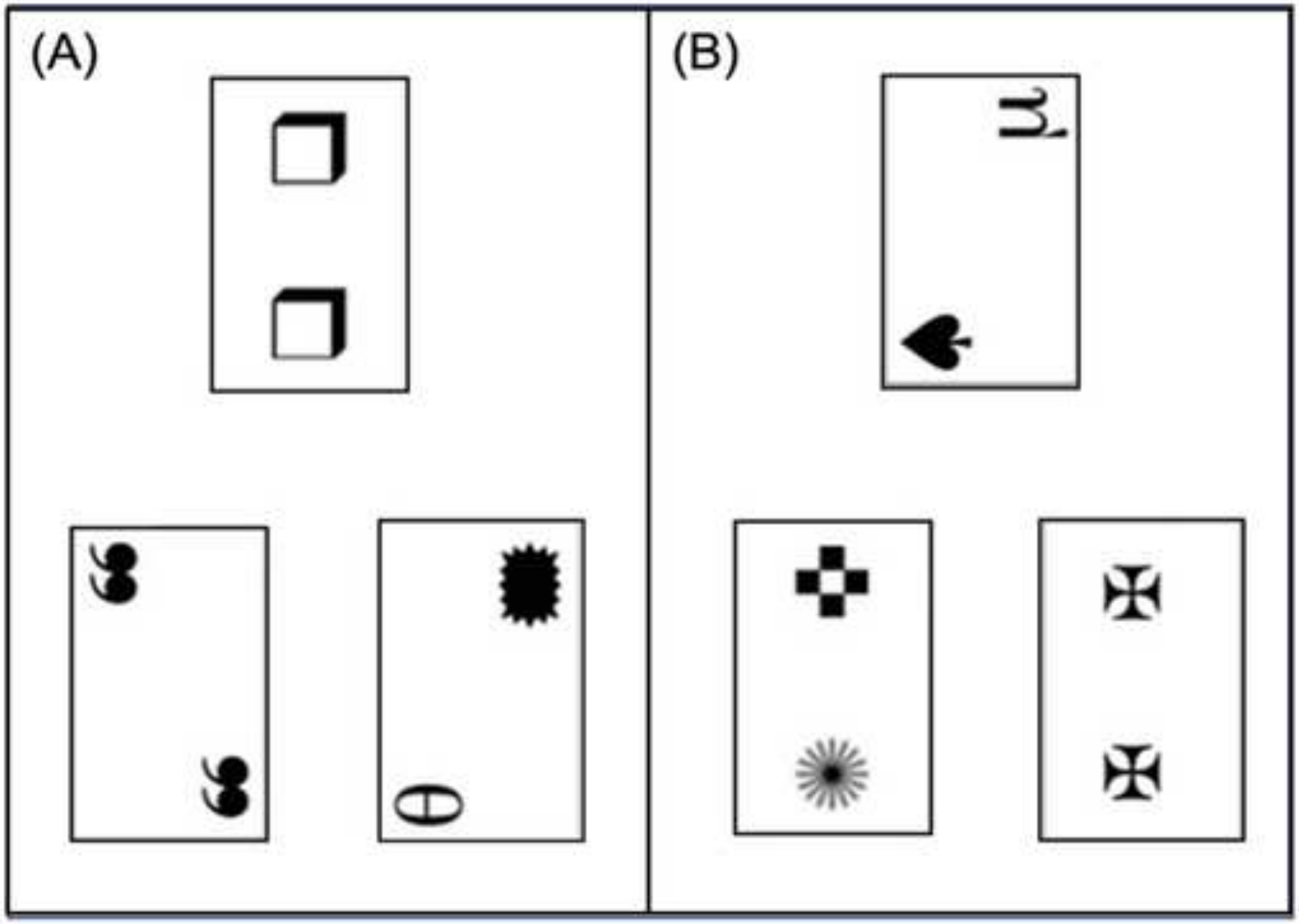


Figure 3

A Sample Triad of the (A) Same and (B) Different Relations of the Full RMTS Task as Used in Hochmann et al. (2017).

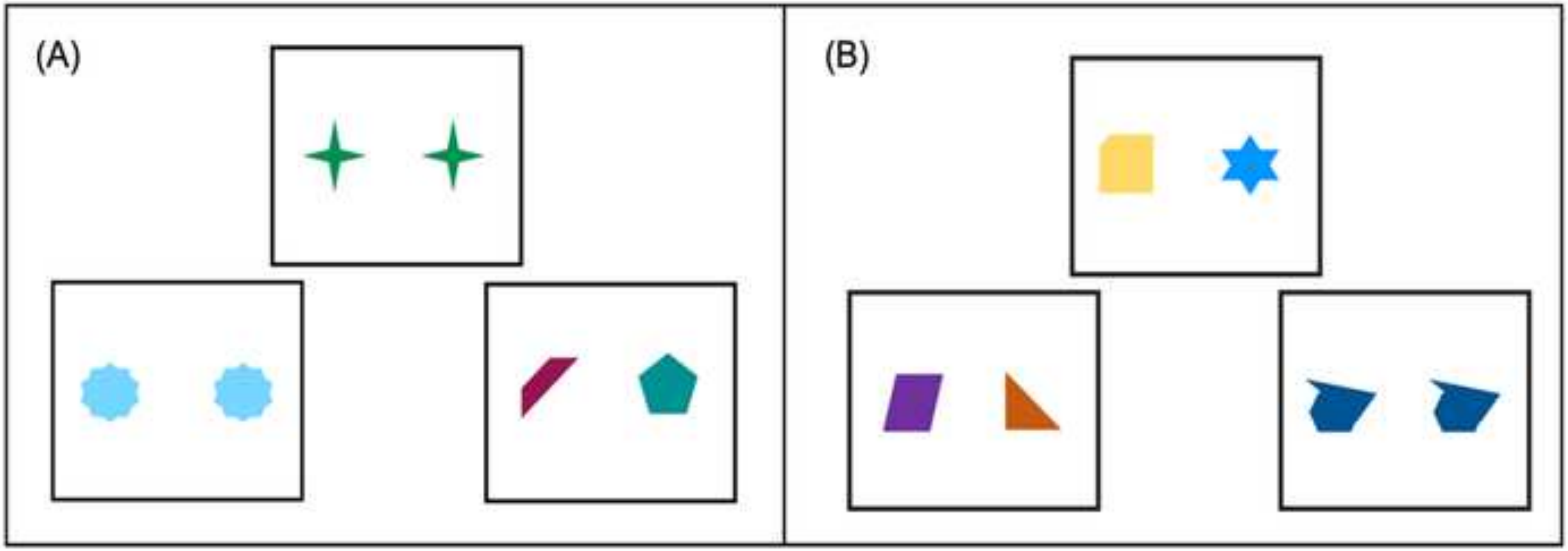


Figure 4

A Sample Triad of the (A) Same and (B) Different Relations of the RMTS Task as Used in Kroupin and Carey (2022).

Click here to access/download;Figure;Figure 5.png

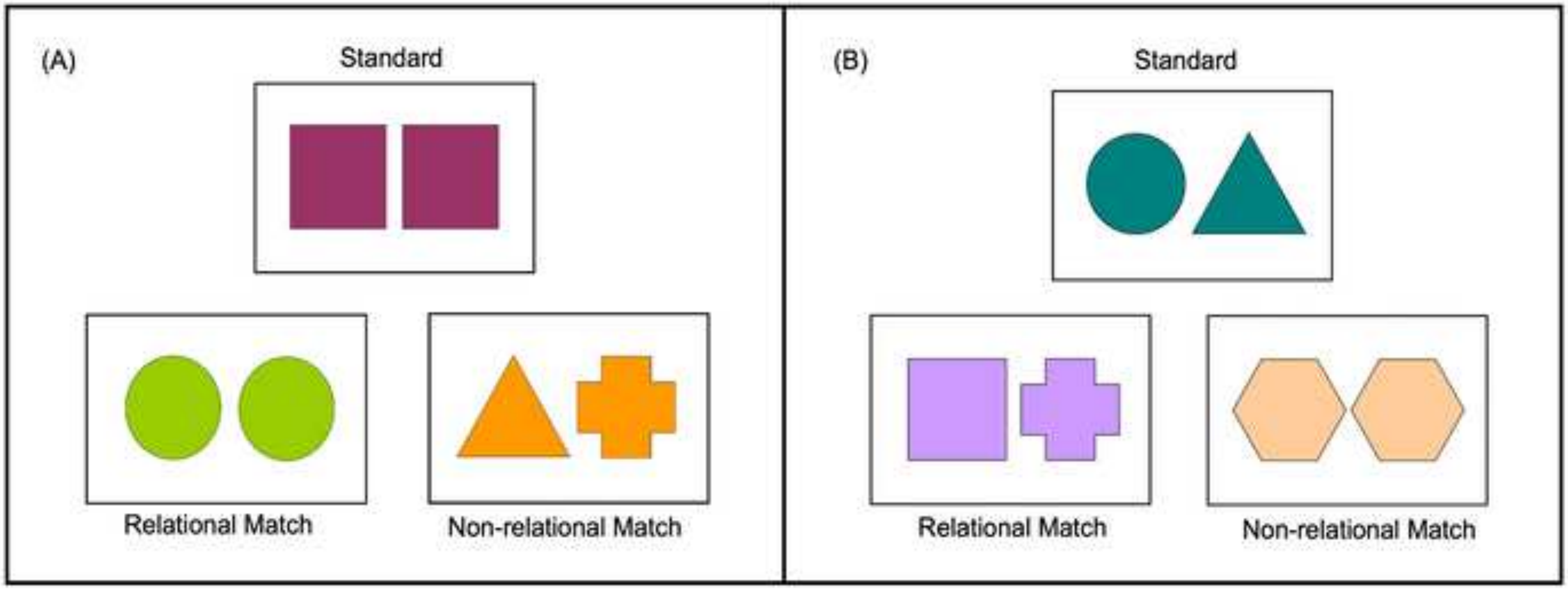


Figure 5

Sample Triads from Experiment 2 with Stimuli Placed in a Horizontal Orientation. (A) Half the Standards Depicted the Same Relation (B) and the Rest were Different.

 Click here to access/download;Figure;Figure 6.png

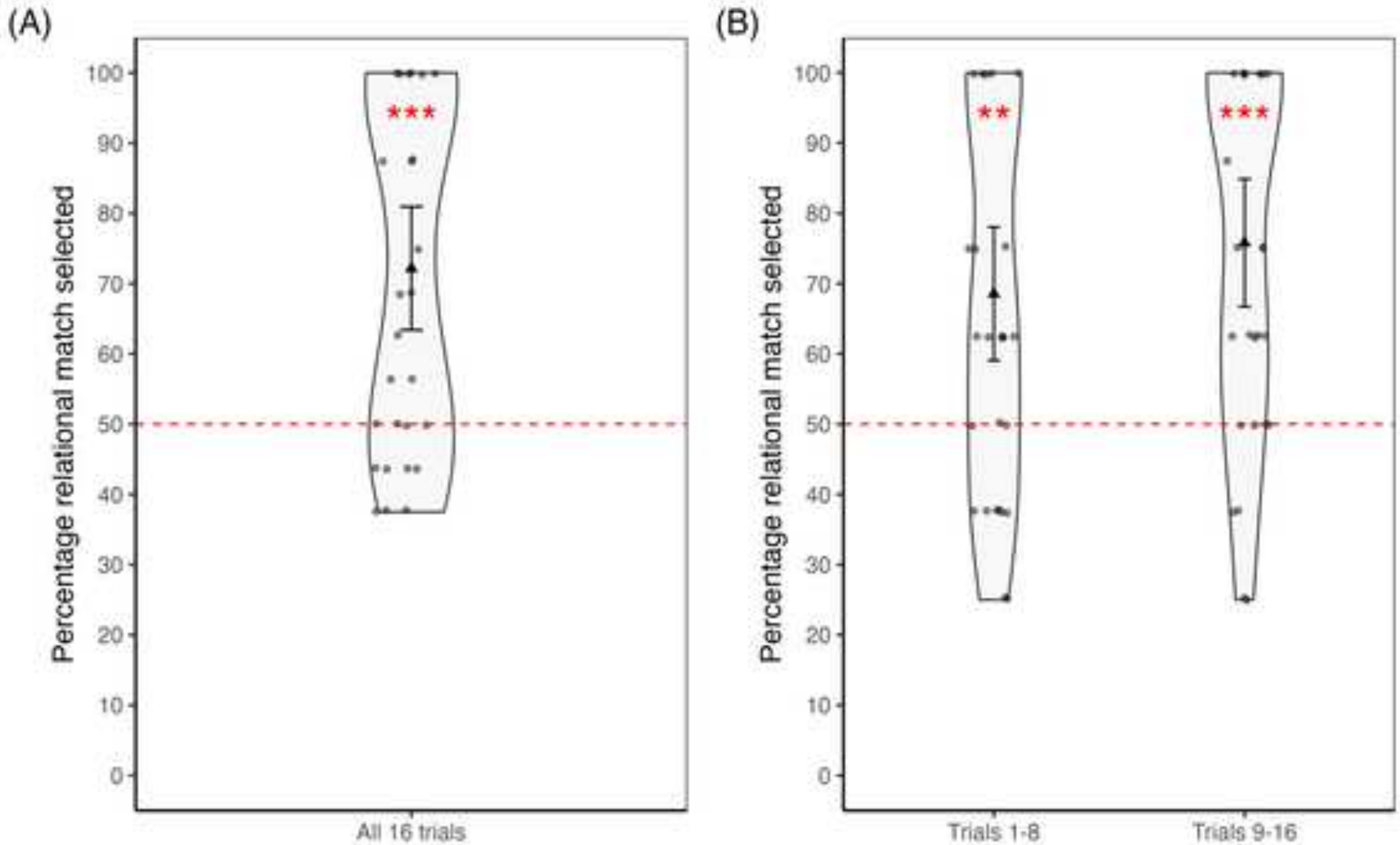


Figure 6

Percentage of Relational Match Selected in Experiment 2 (A) Across all 16 Trials and (B) Across Trials 1-8 and 9-16.

Note. 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 various points of the dependent variable.
* $p < .05$, ** $p < .01$, *** $p < .001$.