Beyond words and actions: what implicit measures reveal in preschoolers' performance on the RMTS task

Antonia Götz (a.goetz@westernsydney.edu.au)
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Genevieve Quek (g.quek@westernsydney.edu.au)
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Zoe-Vasilia Fountotos (Z.Fountotos@westernsydney.edu.au)
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Laura Fimmano (20127321@student.westernsydney.edu.au)
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Susan Hespos (s.hespos@westernsydney.edu.au)
MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

Abstract

This study investigates relational reasoning in preschoolers using the Relational-Match-To-Sample (RMTS) task, which tests their ability to match "same" and "different" relations. We investigate (1) whether 4-year-old children can succeed in the RMTS task and (2) whether verbal justifications of relational language predict success. Forty-nine children participated (M_{age}=54.97 months), and their performance was measured both behaviourally and through eye-tracking. Results show children identified relational matches above chance. Children who used relational language selected relational matches more often. Eye-tracking data revealed distinct temporal looking patterns during relational and non-relational choice trials, with children preferring relational matches after a brief comparison phase. A cluster-based analysis confirmed that children looked longer at relational than non-relational matches. These findings suggest that relational reasoning in preschoolers involves a dynamic comparison process, and eye-tracking provides valuable insight into this implicit cognitive process.

Keywords: relational match to sample, eye-tracking, preschoolers

Introduction

Among all species, humans excel in using abstract representations: for example, you read a map by mentally aligning the image on your phone with real-life landmarks and streets as you navigate. This ability to perceive relations between objects, events, or ideas, and to compare those relations across situations, is termed analogical reasoning, and it is the foundation of effective abstract learning in both unstructured and formal education settings. This is particularly so in mathematics and science, where analogical comparisons underlie the development of basic arithmetic problem solving (mapping written number representations onto countable physical tokens) and hypothesis testing (operationalising intuitive predictions as measurement patterns (Simms et al., 2023). As a cornerstone of higher reasoning, clarifying how analogical abilities emerge and develop during early childhood is key to understanding why some children learn more readily than others, and to developing effective interventions for scaffolding conceptual learning.

To study relational ability at its core, we focus on *same* and *different* relations – the fundamental building blocks of relational matching. Specifically, we focus on 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. With children, there are many ways to facilitate performance on this task through labels (Christie & Gentner, 2010, 2014), setting the task within a causal context (Carstensen et al., 2019; Walker et al., 2016; Walker & Gopnik, 2014), feedback during the task (Hochmann et al., 2017), and relational training (Kroupin & Carey, 2022). Together these tasks demonstrate that there are many paths to success on the RMTS task when training and feedback support performance.

In the Premack (1983) RMTS task, participants are not given pretraining before the task, nor do they receive feedback during the task (hereafter called the *pure* RMTS task). We focus on this *pure* RMTS version of the task because it addresses the critical question of whether there is a convergence in process once children possess the ability to succeed on the task. This investigation could lend new insights about why humans excel in their analogical reasoning skills compared to other species. There are only a

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

few studies that have performed the *pure* RMTS task on children and the results are mixed. Some find success at 4 years of age (Christie & Gentner, 2014; Shivaram et al., 2023) and other studies find children do not pass at 4 years, but do pass at 5.5 years (Hochmann et al., 2017; Kroupin & Carey, 2022).

One consistent predictor of performance that has emerged is it might not be age that determines success, but rather the interaction between producing relational language and possessing *same/different* concepts. For example, studies with children between 4 and 6 years of age found that success on the *pure* RMTS task was predicted by whether children produced the words "same" and "different" to justify why they chose a pair of cards (Hochmann et al., 2017; Shivaram et al., 2023). This indicates that an interaction between relational language and possessing *same/different* concepts may be a better proxy than age for this ability (Shivaram et al., 2023).

In this paper focus we focus on the interaction between relational language and concepts in the *pure* RMTS task. However, to take the next step requires a step-change in analogical research methods. At present, our ability to gain traction on this problem is frustratingly limited by the current state-of-the art in developmental psychology, which relies on group level inference that does not consider individual variability, dichotomous outcomes (e.g., did/did not discriminate the relation), and overt behavioural responses in children who do not yet possess clear communication skills.

To achieve the needed step-change, we propose an eye-tracking measure to complement our behavioural measure. Eye-tracking offers a powerful, language-independent window into children's cognitive processing. It provides real-time data on where and when children allocate their attention as they engage with relational concepts. This is particularly valuable for investigating structure mapping—the process by which children align elements across representations and map corresponding relational structures (Gentner, 1983). Eye-tracking enables us to examine how and when children shift their gaze toward relationally relevant aspects of a visual scene, offering fine-grained temporal insights into how the mapping process unfolds. For example, earlier fixations may reflect perceptual salience or object-based strategies, while later fixations can reveal deeper relational reasoning and decision-making processes. Tracking the time course of gaze behaviour allows us to differentiate between children who rapidly identify relational correspondences and those who require more time or follow different attentional pathways before arriving at a decision.

Moreover, the time-resolved nature of eye-tracking data allows us to go beyond outcome measures (e.g., where the child ends up looking) to examine the process of relational interpretation—how children weigh competing options, switch attention between items, and gradually build up a relational match. This dynamic view is critical for capturing variability within and across individuals, especially when explicit responses are absent or ambiguous.

The necessary first step is to validate this method by collecting simultaneous behavioural and eye-tracking data, enabling us to assess how well children's overt responses align with their real-time processing patterns, and to explore whether eye movement patterns can serve as early indicators of relational understanding.

Experiment

The goal of this experiment is to examine (1) whether children can succeed on the RMTS task at the age of 4 years and (2) whether justifying their answers with "same" and "different" predicts success on the RMTS. We will assess performance with behavioural measures of performance and complement these measures with eye-tracking data.

Method

Participants Forty-nine children participated in the study (25 female, 24 male; M_{age} = 54.97 months, SD = 2.85, range = 4;3-5;3). Parents reported that the children had no known significant auditory, visual, or developmental impairments or disabilities. An additional 15 children were excluded from the analysis due to failing the catch trial (n = 13, described below) or refusal/non-compliance (n = 2).

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 target sample size was determined based on the effect size (d) derived from the performance of 4.5-year-olds who achieved success on Shivaram et al.'s (2023) RMTS task at above-chance levels. Children were recruited using an established, voluntary participant database. Approximately 88% of primary caregivers had a university degree or higher, 7% started university but did not complete their degree/diploma, and 5% completed high school. Parents provided informed consent before the experiment. All procedures were approved by the Ethics Committee. Parents were given $30 as compensation for their participation, and their child was given a graduation certificate and a gift. For the osf registration, see here: osf.io/75462

Materials The materials were identical to Shivaram et al. (2023) study, except we presented the stimuli on a screen instead of laminated cards placed on a table. There were sixteen triads. Each of the three cards in a triad had two coloured geometric shapes placed horizontally. The individual shapes were familiar figures such as triangles and circles; each appearance of a given shape had a unique colour. Each triad had a standard card depicting a *same* or *different* relation and two alternatives – one depicting *same* and the other depicting *different*, as shown in Figure 1. To avoid object matches, neither alternative included any shape that was also present in the standard. Left/right placement of the alternatives was counterbalanced. Three catch trials were given after the end of the test trials to verify that children understood the task. In these triads, the standard was a single object (e.g., blue fish) and the alternatives were a card with an object that was highly similar to the standard (e.g., red fish) and another card with an object that was highly dissimilar to the standard (e.g., yellow cup). A Tobii X120 eye tracker (Tobii Technology BA, Stockholm, Sweden) was used to capture children's eye movements at a 60 Hz sampling rate.

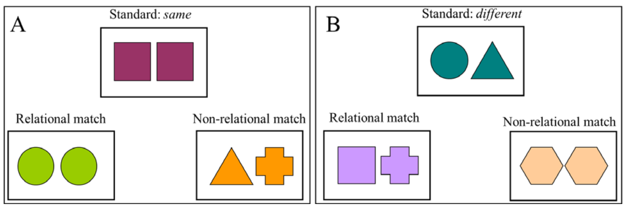

Figure 1: Sample triads for the RMTS task depicting (A) *same* and (B) *different* relations. The standard appeared first and when the participant was looking at the screen, the experimenter tapped to make the two alternatives appear simultaneously.

Procedure After obtaining parental consent and the child's verbal assent, the experimenter invited the child to a quiet room. The child was seated on a chair so that their eyes were approximately 60 cm from the eye tracker and centred in front of the computer screen. The experimenter was seated to the side. Once the eye-tracker could detect the participant's eyes, the child did a 5-point calibration. Children were taught to raise their left or right hand depending on the side on which the icon appeared.

Next, the experimenter began the first eight RMTS trials. The trials were presented in a pseudorandomised order with *same* or *different* standards. Before the start of each trial, an attention-getter was displayed on the screen to attract the child's attention. Once their attention was focused on the screen, the experimenter presented the standard at the top of the screen and said, "*Can you see this one?*". When the child acknowledged, the experimenter presented the two alternatives in the bottom corners of the screen and said, "*Can you see these two? Now, which of these two at the bottom is more like the one at the top?*". The child indicated with response by lifting their left or right arm for choosing the left or right match and the experimenter wrote down the answer on a clipboard before going on to the next trial. After the first eight trials, the child was given a short break. Then, the experimenter did the next eight trials, followed by three catch trials presented in the same manner. Failure to correctly match any of the three catch trials led to the participant's data being excluded, as being unable to match two identical objects indicates that the child did not understand the task.

Following the catch trials, there were two RMTS justification trials. The justification trial was identical to the RMTS trials except after the child selected their answer, the standard and the card the child chose were presented in a column and the experimenter said, "*Why do you think the bottom one is more like the top one?*". The experimenter wrote down the child's answer (for more information see also Shivaram et al., 2023). Neutral encouragement was provided throughout the experiment to keep the child motivated and engaged; no feedback was given.

Analysis Plan

Behavioural measures. All statistical analyses were performed using R (version 4.4.0, R Core Team, 2024).
Two types of behavioural analyses were conducted. First, as the dependent variable (percentage of relational match chosen) was not normally distributed, Wilcoxon's one-sample *t*-test and its corresponding effect sizes (*r*) were calculated to test whether children's performance was significantly different from chance. Second, to examine whether children's relational language was a predictor for the performance of the experiment, a logistic mixed-effect regression was conducted with accuracy (0 for incorrect and 1 for correct responses) as the dependent variable and relational language, age (in months) and sex (male or female) as fixed factors.

Eye-tracking measures. The areas of interest (AOIs) were defined as rectangular regions surrounding the standard, relational match, and non-relational match (see Figure 1, ranged from 320-528 pixel). The raw eye-tracking data were first organised and formatted using the eye-tracking R package (version 0.2.1, Dink & Ferguson, 2015). The proportion of looking was calculated based on fixations within the defined AOIs, ignoring fixations outside these areas to focus the analysis on how looks were distributed between relevant AOIs.

The time window of interest was the first 3000 ms after the two alternatives appeared on the screen modeled on Yuan et al. (2024). Within this time window, trials with more than 40% track loss were removed. Participants with excessive track loss (defined as more than half of their data) were excluded from the analysis to maintain the reliability of the dataset. After applying these criteria, the dataset retained an average of 9.94 trials per participant (SD = 3.11). For the eye-tracking analysis, 17 more participants were excluded because children had less than 50% of trials after trackless rejection. That means we have a sample of 32 children (M_{age} = 55.09 months (SD = 2.99), for our eye-tracking analysis

We used a cluster-based permutation analysis because eye-tracking studies face additional challenges due to the lack of a standardised method for data analysis, particularly in how statistical models like ANOVAs or generalised mixed-effects models handle time as a variable. A common strategy involves dividing the trial into several time windows (e.g., every 10 time bins) and analysing each window separately. However, this method has its drawbacks. First, it introduces artificial boundaries in the processing timeline. Second, studies employing this approach often neglect to adjust for multiple comparisons when analysing repeated eye-movement data, increasing the likelihood of Type I errors. To overcome these limitations, a non-parametric permutation analysis was utilised (Chan et al., 2018; Maris, 2012; Maris & Oostenveld, 2007). This method generates a permutation distribution by resampling the observed data, making it especially effective for identifying processing events in a data-driven manner.

The permutation procedure involved calculating the differences in gaze behaviour between relational vs. non-relational matches at each time point (here every 17 milliseconds, corresponding to eye-trackers sampling at 60 Hz). A t-statistics was computing comparing the two conditions, using the available (within-subject, and non-missing) data. The data were then resampled 1,000 times to produce a distribution of test statistics under the null hypothesis, which accounted for variability and assessed the significance of observed differences in gaze behaviour.

Subsequently, significant differences across contiguous time points were grouped into clusters (the sum *t*-distribution), representing periods where gaze patterns consistently diverged between conditions. The significance of each cluster was evaluated by comparing the observed test statistics against the permutation distribution.

Results

Behavioural Results

The main question was whether children would succeed at the RMTS task. The results showed that children chose the relational match at the above-chance levels (M = 55.63 %, SD = 16.48, Wilcoxon's r = .28, p = .042, see Figure 2).

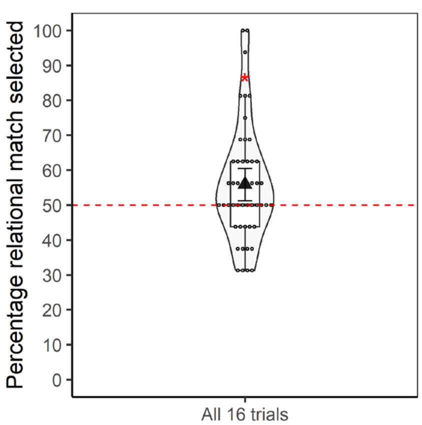

Figure 2: Percentage of times children chose the relational match. For this figure, the black triangle represents the mean, and the whiskers are standard error bars. The grey dots represent the individual data points, and the violin plot indicates the density of the distribution at different points of the dependent variable. The red dashed line indicates chance performance. * means $p < .05$

Children who used the relational terms 'same' and 'different' in justifying their choices (in one or two of the justification trials, n =13) had high accuracy on the RMTS task (see also Shivaram et al., 2023). We conducted a logistic regression with the accuracy (0 incorrect, 1 correct) as the outcome variable and used whether the children had relational language (producing the words *same-different*) during the justifications as a predictor.

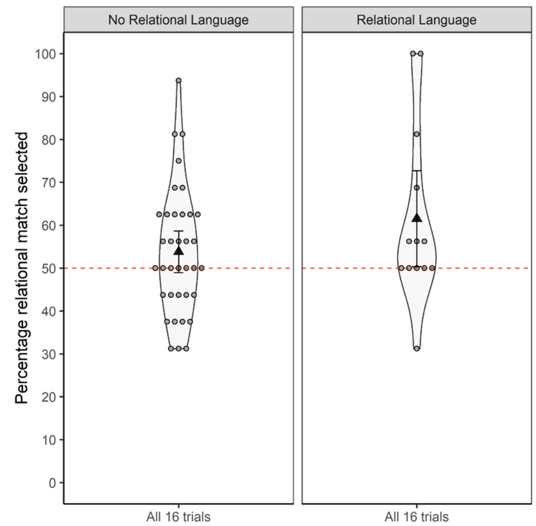

Figure 3: Percentage of times children chose the relational match and did not produce *same- different* relational language (left) and children who produced relational language (right). For this figure, the black triangle represents the mean, and the whiskers are standard error bars. The grey dots represent the individual data points, and the violin plot indicates the density of the distribution at different points of the dependent variable. The red dashed line indicates chance performance.

Although the trend was in the predicted direction, using the relational terms *same* and *different* on justification trials, relative to justifying their choices with no relational terms, was a not a significant predictor of children's relational performance on the RMTS task (β = 7.99, SE = 5.24, t = 1.53, p = .13), see Figure 3.

Eye-tracking Results

The main question was whether children look longer at the relational match compared to the non-relational match. The answer is yes. Figure 4 shows the mean proportion of looks to the relational match, non-relational match and standard. The results from the cluster-based permutation task revealed that children looked more to the relational than the non-relational match in the time window from 2227ms to 2839ms (*sum t* = 117, p = .003). For the second analysis, we examined whether the proportions of looks differed statistically based on the production of relational language (n = 6) versus no relational language (n = 24). Specifically, we compared the differences in proportions of looks to the relational match and non-relational match across relational language use (relational vs. non-relational language) within the time window identified by the cluster-based permutation task. There were no statistical differences in the proportions of looks between relational language use ($\beta(SE)$ = 0.015 (0.031), t = 0.501, p = .995). Note /that there were only six out of the 13 children who produced relational language and good eye tracking data so it is likely that the analysis is underpowered.

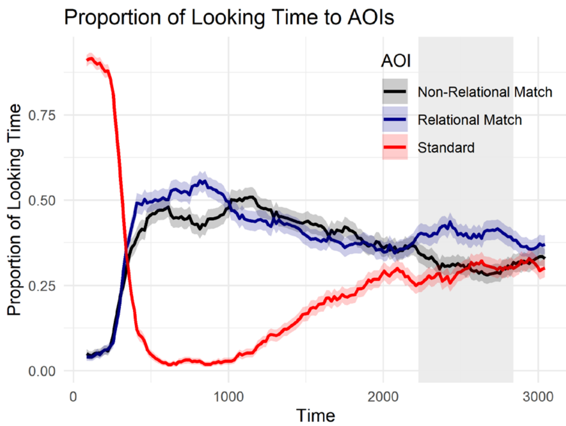

Figure 4: Average proportion of look time to the Relational Matches, Non-Relational Matches, and Standard. The grey rectangular shading indicates the time bins which were found to be significantly different between the Relational and Non-Relational Match in the permutation analysis.

As part of an exploratory analysis, we investigated how looking patterns varied on an individual trial basis. We aimed to explore whether children use distinctive patterns for trials in which alternative they chose. The eye-tracking data were divided into trials where children chose the relational match vs. non-relational match (see Figure 5). This approach allowed us to explore the speed with which children made their decisions. Again, we applied a cluster-based permutation task to find a difference in looking proportion to relational and non-relational matches. For trials in which children chose the relational match (n = 226), we identified two clusters with a higher proportion of looking to relational than non-relational matches. The first cluster was from 527ms-952ms (*t sum* = 99, *p* = .001), and the second cluster was from 2227ms-2924ms (*t sum* = 160, *p* < .001). In contrast, for trials where children chose the non-relational match (n = 190), we identified one cluster from 1020ms-1326ms (*t sum* = -64, *p* = .032).

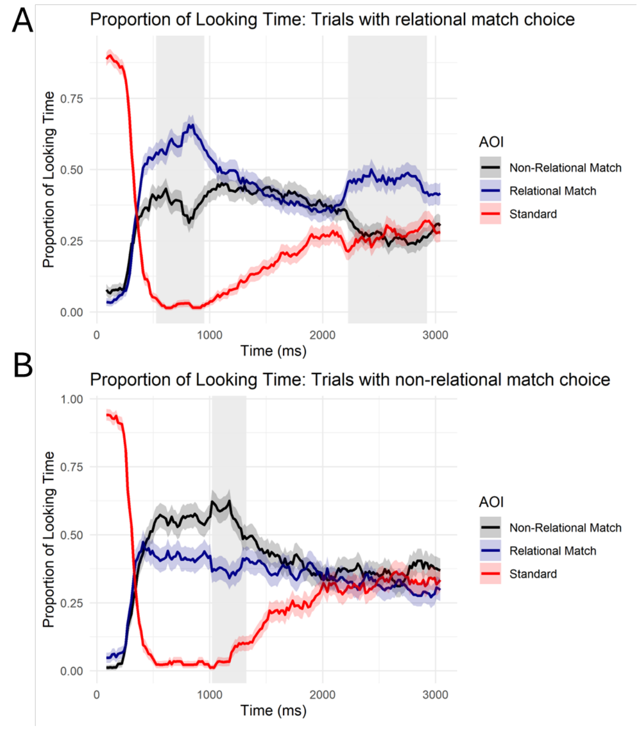

Figure 5: Average proportion of looking time to the Relational Match, Non-Relational Match, and Standard. Divided by trials where children chose the relational match (A) and children who chose the non-relational match (B). The grey rectangular shading indicates the time bins found to be significantly different between the Relational Match and Non-Relational Match in the permutation analysis.

Group-level time series data show that children spend a higher proportion of time looking at the relational match compared to the non-relational match, starting around 2200 ms post-stimulus. Prior to this divergence, looking proportions for both conditions overlap. In the context of this study, children appear to base their decision-making on the relative comparison of relational versus non-relational stimuli. This prompted an analysis of fixation transitions. On average, children shifted their gaze (bidirectionally) from the standard to the relational match 11.10 times (SD = 6.39), from the standard to the non-relational match 10.65 times (SD = 5.98), and from the relational match to the non-relational match 14.67 times (SD = 6.81). A series of chi-square tests revealed significant differences in how often children shifted their gaze between different areas of interest. Children shifted their gaze between the relational match and the non-relational match significantly more often than between the standard and the non-relational match, $\chi^2(1) = 10.18$, $p = .001$. Similarly, transitions between the relational and non-relational match occurred more frequently than between the standard and the relational match, $\chi^2(1) = 6.17$, $p = .013$. In contrast, there was no significant difference in the frequency of gaze shifts between the standard and relational match versus the standard and non-relational match, $\chi^2(1) = 0.31$, $p = .575$. Figure 6 confirms the development of fixation transitions over time. After an initial comparison between the standard and relational match and the standard to the non-relational match, children focused on the comparison between alternatives.

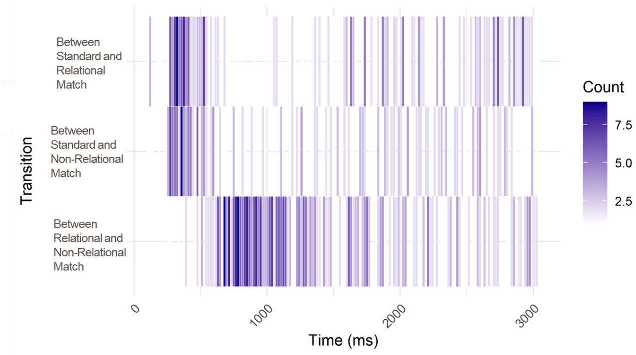

Figure 6: Number of fixation transitions over time.

Discussion

The behavioural results from our experiment show that Australian preschoolers choose the relational match at rates that are significantly above chance. These findings provide evidence that of relational reasoning because the children detected the *same/different* relation without training or feedback. This finding provides converging evidence with other studies in this age group (Shivaram et al., 2023), but our study is new in that the task was presented on a screen instead of physical cards presented on a table.

The behavioural results on the production of relational language were less clear. There was a trend in the predicted direction in that children who produced relational language tended to choose the relational choice (M = 61.53%) more often than children who did not produce relational language (M = 53.65%). However, this difference did not reach statistically significant levels. One possible explanation is that very few children produced relational language ($N = 13$) compared to those that did not produce relational language (N = 36). Consequently, the null result could be due to the unequal sample sizes. A factor that could have contributed to fewer children who produced relational language during justification trials is that approximately half of the participants were from families that speak a heritage language other than English in the home. In turn, many of the children in the study did not have fluency in English as they have not yet started school. One solution to this problem is to test older children, and this is currently underway.

While the behavioural data capture a transitional period in relational reasoning, there are three interesting findings from the eye-tracking data. First, there is global convergence between the behavioural and eye-tracking results (see Figure 4). A cluster-based, permutation analysis demonstrates that across all trials children looked significantly longer at the relational match compared to the non-relational match 2000-3000 ms after the alternatives were presented. Second, the eye-tracking measure makes is possible to separate trials based on whether the child chose the relational or non-relational match in the behavioural outcome. These eye-tracking data demonstrate that there is a different temporal pattern for relational and non-relational choices. For trials in which children chose the relational match, the time course can be divided into four distinct phases. First (0–500 ms), there is an initial increase in looking proportion toward both the relational and non-relational choices, while attention to the standard decreases. Second (500–1000 ms), children show a clear preference for the relational choice. Third (1000–2000 ms), a comparison phase emerges, indicated by equal proportions of looking time for the relational and non-relational choices. Finally (2000–3000 ms), the preference for the relational choice re-emerges.

In contrast, non-relational choice trials follow a three-phase pattern. First (0–500 ms), as in relational choice trials, children initially increase their looking toward the alternatives while decreasing looking to the standard. Second (500–1300 ms), they show a preference for the non-relational choice. Third (1300–3000 ms), a comparison phase occurs, with similar proportions of looking time directed toward both the relational and non-relational choices.

These findings suggest that children's selection of relational versus non-relational matches is influenced by a comparison process between the two alternatives. This interpretation is further supported by the number of fixation transitions throughout the trial. Within the first 500 ms, children exhibit the highest number of transitions between the standard and the two alternatives. However, after this initial phase, their transitions primarily focus on comparing the relational and non-relational choices.

These patterns of looking align with the structure mapping theory that the process of comparison facilitates detecting the abstract pattern (Gentner et al., 2021). Future studies could track individual children before and after training to see if their eye-tracking patterns change once they learn the relation.

Taken together, these findings offer promising new avenues for research on the RMTS task. They contribute an existence proof that eye-tracking with preschoolers can be done with the RMTS task. This initial attempt demonstrates that there is converging evidence between our overt behavioural measures and more implicit non-verbal responses. The complementary nature of behavioural and eye-tracking data allows for a more nuance description of developmental changes. These changes move us away from dichotomous answers (e.g., success vs. failure) to allow for analyses of individual differences and possibly insights about the variance when there is more than one path to development.

Acknowledgements

This research was partially supported by the Australian Government through the Australian Research Council's Discovery Projects funding scheme (project DP250101325) awarded to Susan Hespos and Genevieve Quek. The views expressed herein are those of the authors and are not necessarily those of the Australian Government or the ARC.

References

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

Chan, A., Yang, W., Chang, F., & Kidd, E. (2018). Four-year-old Cantonese-speaking children's online processing of relative clauses: A permutation analysis. *Journal of Child Language*, *45*(1), 174–203.

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

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis*.

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.

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., Shao, R., Simms, N., & Hespos, S. (2021). Learning same and different relations: Cross-species comparisons. *Current Opinion in Behavioral Sciences*, *37*, 84–89.

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

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

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology*, *49*(4), 549–565. https://doi.org/10.1111/j.1469-8986.2011.01320.x

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

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

R Core Team. (2024). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* [Computer software]. https://www.R-project.org/

Shivaram, A., Shao, R., Simms, N., Hespos, S., & Gentner, D. (2023). When do Children Pass the Relational-Match-To-Sample Task? *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45). https://escholarship.org/uc/item/27j6n7jm

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. https://doi.org/10.1037/xap0000471

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.

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

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