Variation in Adults' Judgements about Relative Proportional Magnitude and Proportional Equivalence

Michelle A. Hurst (michelle.hurst@rutgers.edu)
Department of Psychology and Center for Cognitive Science
Rutgers University, New Brunswick, 152 Frelinghuysen Rd, Piscataway, NJ 08854

Paige Dadika (paige.dadika@rutgers.edu)
Department of Psychology and Center for Cognitive Science
Rutgers University, New Brunswick, 152 Frelinghuysen Rd, Piscataway, NJ 08854

Abstract

Proportional reasoning is critical for successful functioning across domains and development. However, proportional information is also complex, resulting in behavioral variation across contexts and tasks. In the current study, we systematically compare adults' proportion judgements on a proportion magnitude comparison task and an equivalent proportion matching task with both dot arrays and continuous rectangles. We find that the match-to-sample task is more difficult than the magnitude comparison task and dot arrays are more difficult than the rectangles. Interactions between task and format, as well as specific patterns of errors, provide additional insight into possible explanations for these patterns. Overall, findings provide theoretical insight into the cognitive processes involved in solving proportional tasks and methodological insight into how to best design and interpret performance on both comparison and match-to-sample proportion tasks.

Keywords: proportion; magnitude comparison; match-to-sample task

Introduction

Multiplicative relations between quantities, such as proportions and ratios, are ubiquitous in our environment and play a key role in everyday decisions, math education, and psychological theories (e.g., Bonawitz et al., 2014; Denison & Xu, 2014; Jara-Ettinger et al., 2016; Kalra et al., 2020). Despite their ubiquity, there is not a theory of proportional reasoning that makes consistent predictions across tasks and contexts. Part of the difficulty is that proportions are complex and have both quantitative and relational properties: they can be compared and ordered based on magnitude (e.g., ½ < ¾) *and* have equivalence classes across non-identical representations of the same proportion magnitude (e.g., 1/2 = 2/4). Given this, proportional reasoning is often measured using tasks tapping one or the other kind of thinking: magnitude comparison tasks (e.g., Jeong et al., 2007; Hurst & Cordes, 2018; Park et al., 2020) and equivalence matching tasks (e.g., Boyer et al., 2008; Hurst et al., 2021). Although qualitatively similar conclusions have been made using both kinds of tasks, informal comparisons across studies have revealed some differences (e.g., Hurst & Cordes, 2018). In the current study, we explicitly compare proportional reasoning and systematic errors on a proportion magnitude comparison task and an equivalent proportion matching task, with otherwise comparable stimuli and paradigms. Our central goal is to better understand the qualitative and quantitative differences across tasks, to gain both methodological and theoretical insight into the cognitive processes underlying proportional reasoning and how best to measure them.

One of the primary findings in both proportion magnitude comparison tasks and proportion matching tasks is that children and adults make more errors with discrete stimuli (e.g., a collection of dots, a divided pie chart) than with continuous part-whole stimuli (e.g., an undivided pie chart; Boyer et al., 2008; Hurst & Cordes, 2018; Hurst & Piantadosi, 2024; Jeong et al., 2007). These errors are typically attributed to an overuse of the absolute number in the numerator, leading to systematic errors such as deciding a divided annulus depicting 2/3 is less than a divided annulus depicting 4/9 because 2 < 4 (e.g., Jeong et al., 2007) or matching a divided rectangle depicting 2/3 with one depicting 2/9 instead of 6/9 because 2 = 2 (e.g., Boyer et al., 2008). Notably, however, this error is not unique to discrete countable stimuli: when stimuli are large sets or continuous area-based representations, adults make the same errors, systematically over relying on the size of the numerator (Hurst & Levine, under review). Despite being present for both discrete and continuous stimuli, the errors are typically larger for discrete stimuli than for highly regular continuous stimuli, such as pie charts (e.g., Hurst & Piantadosi, 2024).

Although this general finding has been found for both comparison and matching tasks, it may vary in strength. Hurst and Cordes (2018) found a qualitatively different pattern of performance with a probability comparison task than Boyer and colleagues (2008) found with a juice-mixing match-to-sample task. They speculated that it may be due to the increased difficulty of having three stimuli (in a match-to-sample task) vs. two (in a comparison task). A recent study, Boyer and colleagues (2024) explicitly compared the two paradigms with divided shapes – magnitude comparison of probabilities and match-to-sample of juice-mixing proportions – and similarly found better performance and more numerator interference on the magnitude comparison task than the match-to-sample task. However, the two tasks differed not only in their structure, but also in the domain they

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

were embedded in (i.e., probability vs. juice-mixing) and the shapes used in the task (i.e., pie charts vs. rectangles), making it difficult to interpret this finding.

Despite the difficulty interpreting this prior work, there are findings from the numerical cognition literature that might predict systematic differences between magnitude comparison and match-to-sample tasks. For example, children show more difficulty extracting numerical information from heterogeneous sets, in contrast to homogeneous sets, in the context of a matching task but not in a magnitude comparison task (Cantlon et al., 2007). This suggests that the magnitude comparison goal may facilitate numerical abstraction, whereas the matching goal may direct children to focus more on the superficial features of the set. In our proportion context, we could make a similar prediction: focusing on relative magnitude in the comparison task may facilitate extracting proportional information, which is obscured or inhibited in the matching task.

In the current study, we had two primary goals. First, how do typically studied format effects interact with task type, on an otherwise well-matched paradigm? We used the same proportion context with the same stimuli but differed whether the goal was to compare the relative magnitude of two stimuli or select which of those two stimuli matched a third stimulus. If the differences reported previously across and within studies was due, at least in part, to differences between a comparison and match-to-sample task, then we would expect to also find a difference in the current study, when stimuli and proportion context are well matched.

Our second goal is about specific kinds of errors on the two tasks. Most magnitude comparison tasks include trials where the numerator information is helpful (e.g., 2/3 vs. 1/8) and where the numerator information is harmful (e.g., 2/3 vs. 4/9). However, this makes it difficult to compare *proportional reasoning* across different formats; instead, we are capturing differences in people's tendency to use absolute information over proportional information. Thus, in the current study, we include numerator neutral trials (e.g., 2/3 vs. 2/5) so that we can ask whether having redundant numerator information is helpful and/or whether having interfering numerator information is harmful, relative to a baseline proportional comparison task.

Additionally, given the differences in task structure, there is also the possibility that different kinds of trials give rise to different opportunities for errors across the comparison and match-to-sample tasks. Thus, we introduced an atypical set of trial types on the match-to-sample task as well: exact matches that allow participants to match the identical proportion (e.g., 2/4 = 2/4) and equivalent match trials that require generalizing within a proportional equivalence class (e.g., 2/4 = 4/8). Additionally, we included three kinds of competing foils: the incorrect response could match the target on the numerator set, the total amount, or neither. In doing so, we can begin to better understand how format effects and task goals might interact to give rise to specific kinds of errors and proportional reasoning difficulties.

We report a sample of adults in the current study because even adults show difficulty with discrete proportional matching (e.g., Hurst et al., 2021) and comparison (e.g., Hurst & Piantadosi, 2024). Furthermore, this difficulty with proportional information extends into symbolic number domains, like fractions (e.g., Hurst & Cordes, 2014), and decision making (e.g., Thompson et al., 2020). These methodological and theoretical issues are also relevant for substantial research with children, making the current adult sample informative for understanding the complexities of task and context dependent proportional reasoning, as well as providing important insight for studying the development of proportional reasoning in future work with children.

Method

Participants

We recruited 201 adult participants from the USA on Prolific, M_{age} = 39 years, range: 19 to 63 years; 86 women, 111 men, 2 nonbinary, 2 not reported. The sample was 70% White, 18% Black or African American, and 5% Asian, and 4% one or more race. Additionally, 4% was Hispanic or Latine. Most participants (64%) had at least a four-year bachelor's degree.

We intended to recruit 200 participants with equal sample sizes in each of the between-subject versions, but due to a technical error and uneven drop out on Prolific, we ended up with 201 and a slight imbalance across versions: 98 participants were randomly assigned to compare continuous rectangles (47 completed the comparison task first and 51 completed the MTS task first) and 103 participants were assigned to compare dot arrays (52 completed the comparison task first and 51 completed the MTS task first).

Given the difficulty of estimating power and sample size needs for mixed effects models, we estimated our sample size needs using analogous traditional hypothesis testing methods. Based on simulations reported in Brysbaert (2019), 200 participants would provide 80% power to detect an interaction between task (within-subject) and format (between-subject) of d = 0.4.

Stimuli

Between subject, participants were assigned to make judgements about sets of dots (discrete) or integrated rectangles (continuous) and to use either orange or blue as the "numerator" color (see Figure 1 for example trials). Stimuli were generated using set sizes between 5 and 50 for each subset (i.e., orange and blue), resulting in total set sizes ranging from 13 to 60. The ratio between the two proportions (i.e., larger proportion/smaller proportion) varied from 1.15 to 1.70, with an average of 1.39.

In the discrete condition, the visual proportions were intermixed blue and orange dots, randomly placed within a grey rectangle, with a minimum of 10 pixels between the dots. The dots were all equal in size, so that cumulative surface area covaried with number.

In the continuous condition, the visual proportions were rectangles divided into the required proportion such that the

target color was on the bottom and the other color was on the top. The rectangle was then randomly placed within a grey rectangle. The rectangle dimensions were created so that the areas were matched to the discrete dot arrays. For example, for the stimulus depicting 8/12 orange, the orange portion of the rectangle would have the same surface area as the cumulative surface area of 8 dots. Similarly, a rectangle depicting 3/6 would have half the total area as a rectangle depicting 6/12, but the same proportion. Additionally, the heights and widths of the rectangles were varied by randomly selecting one dimension (within a restricted range) and calculating the remaining dimension as needed to meet the required area.

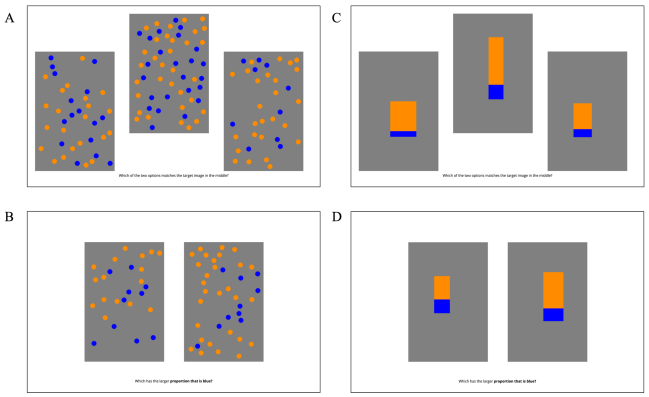

Figure 1: Example trials from both the match-to-sample task (Panels A and C) and magnitude comparison task (Panels B and D), with dots and rectangles. For these example trials, the goal is to decide which option matches the proportion that is colored blue in the target (MTS) or the image with the larger proportion colored blue (comparison).

The stimuli used in the comparison task were identical to the options used in the match-to-sample task. In the match-to-sample task, a third stimulus was generated to be proportionally equivalent to one of the two stimuli. On a quarter of the trials (6), the target was an exact value match to one of the options, but visually different (e.g., 12/20 and 12/20, but with a different dimension rectangle or different arrangement of dots). On 10 trials, the target was scaled down from the correct match (e.g., target: 12/24 vs. match: 24/48). On 8 trials, the target was scaled up from the correct match (e.g., target: 24/28; match: 12/24). The scale factor ranged from 1.5 to 2 (scale up) and 0.33 to 0.66 (scale down), with an average of 1.75 and 0.55, respectively.

Procedure

Participants completed two proportion tasks: a magnitude comparison task and a match-to-sample task, with order counterbalanced. Each task began with four practice trials, with feedback about which response was correct. After the practice trials, there were 24 trials in each task, with the order of the trials entirely random and no accuracy feedback. After each task, participants were asked to explain how they completed the task using an open text box. After both tasks, participants self-reported whether they were paying attention (multiple choice options: yes, sometimes, or no) and whether they had any technical or other issues.

Magnitude Comparison Task On the magnitude comparison task, participants were shown two visual proportions simultaneously, one on the left and one on the right, and asked to decide which displayed a larger proportion colored orange (or blue, counterbalanced) by selecting the left or right arrow key, respectively. The stimuli remained visible for a maximum of 1200ms, followed by a blank screen. Participants were encouraged to respond as quickly as they could, either while the stimuli were visible or after, and their response ended the trial. There were a total of 24 trials, with 8 each of three trial types: numerator incongruent, where the larger numerator and larger proportion corresponded to different proportions (e.g., 8/11 vs. 20/40), numerator congruent, where the larger numerator and larger proportion corresponded to the same proportion (e.g., 9/12 vs. 7/20), and numerator neutral trials, where the two options had the same numerator value, making it unusable (e.g., 10/34 vs. 10/22).

Match-to-Sample Task On the match-to-sample task (MTS), participants were shown three visual proportions: a target in the upper center of the screen and two options on the lower half of the screen, one on the right and one on the left. Participants were asked to decide which of the two options matched the target based on the proportion colored in orange (or blue, counterbalanced). They were asked to respond as quickly as they could using the left or right arrow key, corresponding to the left or right option. Stimuli remained visible for 1800ms, followed by a blank screen, and participants could respond while the stimuli were visible or after. There were 24 trials, with 6 each of four trial types: equivalent matches with a non-interfering foil, equivalent matches with a numerator foil, equivalent matches with a total foil, and exact matches (with varying foils).

On the three equivalent match trial types, the correct proportional match was an equivalent proportion to the target, but not an exact match (e.g., 10/12 and 5/6). These trial types differed in the features of the alternative non-proportional option. On non-interfering foil trials, the incorrect option did not match the target on the amount of orange, amount of blue, total amount (i.e., orange + blue), or proportion. On numerator foil trials, the incorrect option had the same size numerator as the target option (e.g., target: 10/12, incorrect: 10/20). On the total foil trials, the incorrect option had the same total amount as the target option (e.g., target: 10/12, incorrect: 6/12). On exact match trials, the target and the proportional match were identical proportion values (e.g., 10/12 and 10/12), but different visual instantiations (i.e., randomly mixed dots, differently placed rectangle). On these 6 trials, two each had the three possible foil types discussed above (i.e., non-interfering foil, numerator foil, total foil). However, we analyze these six trials as a group because of the small number of trials per subtype and because the presence of the exact match changes the interpretation of the

foils (e.g., on numerator foil exact match trials, all three options have the same numerator, making the foil not a more alluring alternative than the correct response).

Transparency, Openness, and Data Analysis

Data was analyzed in RStudio using R (Version 4.3.2; R Core Team, 2023) and R-packages from the *tidyverse* (Version 2.0.0; Wickham et al., 2019) for data wrangling, *ggdist* (Version 3.3.2; Kay, 2024) and *ggplot2* (Version 3.5.0; Wickham, 2016) for data visualization, and *lme4* (Version 1.1.35.1; Bates et al., 2015), *lmerTest* (Version 3.1.3; Kuznetsova et al., 2017), and *rstatix* (Version 0.7.2; Kassambara, 2023) for data analyses.

We used our lab's standard operating procedures for data exclusion: individual trials with reaction times less than 200ms or more than 10000ms were excluded (< 1% of trials in each task), entire participants when more than 50% of their trial level data was excluded (0), adults who self-reported not paying attention (0), and adults who reported substantial technical errors (0).

Throughout, we use primarily a generalized linear mixed effects model framework, using a binomial function given the binary outcome of accuracy, fixed effects determined for each analysis, and random effects of subject. When an interaction is included in the model, fixed effects of task (0 = comparison, 1 = MTS) and format (0 = continuous rectangle, 1 = dot arrays) are mean centered to facilitate interpretation.

Materials, data, and analysis scripts are available on the Open Science Framework: https://osf.io/kdm34.

Results

Overall Format and Task Effects

To investigate overall effects of format and task, we used a binomial mixed effects model with fixed effects of task, format, and their interaction. In addition to main effects of format and task, there was a significant interaction: $b = 0.50$, SE = 0.08, $p < .001$, $OR = 1.65$ (see Figure 2).

We investigate this significant interaction in two ways. First, using the same model approach within each of the two tasks separately, we find that participants scored significantly higher with rectangles than with dot arrays on both the magnitude comparison task, $M_{rect} = .67$, $M_{dots} = .50$, $b = -0.71$, SE = 0.09, $p < .001$, OR = 0.49, and the MTS task, $M_{rect} = .51$, $M_{dots} = .47$, $b = -0.19$, SE = 0.06, $p = .002$, OR = 0.83. Second, using one-sample t-tests on data aggregated at the participant level, participants scored significantly above chance with rectangles in the comparison task, $t(97) = 10.74$, $p < .001$, significantly below chance with dots in the MTS task, $t(102) = -3.80$, $p < .001$, but not significantly different from chance in the other two cases, $ps > .321$. Taken together, this suggests that although there were significant format effects in both tasks, they were quantitatively and qualitatively different. The effect was larger in the magnitude comparison task and driven by above chance performance with rectangles, whereas the effect was smaller and driven by below chance performance with dots in the MTS task.

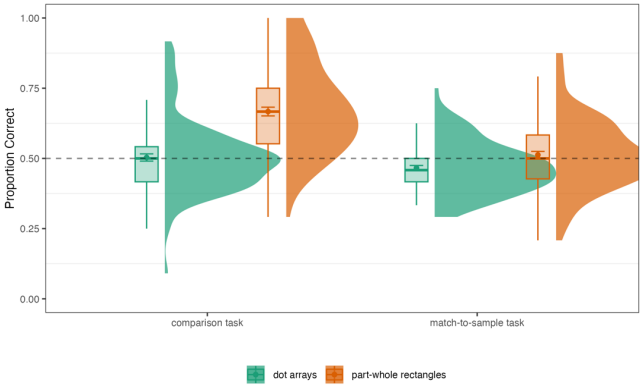

Figure 2: Proportion of all trials correct (y-axis) on the magnitude comparison task (left) and the match-to-sample task (right), when comparing dot arrays (green) or rectangles (orange). Box plots represent the interquartile range (box), full range (whiskers), and median (lines); points are means +/- one standard error. Half violin plots use a smoothed probability density function to display individual level data.

Additionally, performance on the two tasks was significantly correlated for rectangles, $r = .47$, 95% CI [.30, .61], $p < .001$, but not for dots, $r = .18$, 95% CI [-.01, .36], $p = .063$. These correlations were also significantly different from each other, $z = -2.29$, $p = .022$ (Figure 3).

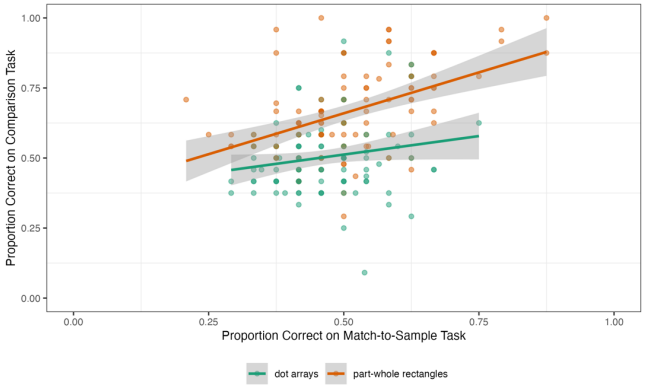

Figure 3: Proportion of all trials correct on the magnitude comparison task (y-axis) and match-to-sample task (x-axis) separated by format: dot arrays (green) and rectangles (orange). Each point is a single participant, and the line presents the line of best fit and 95% confidence bands.

Errors on the Comparison Task

We designed the magnitude comparison task to include three trial types: numerator neutral, numerator congruent, and numerator incongruent (Figure 4). Planned analyses included two kinds of comparisons: a) format effects within each trial type separately and b) within each format, the effect of having congruent or incongruent absolute information relative to neutral trials.

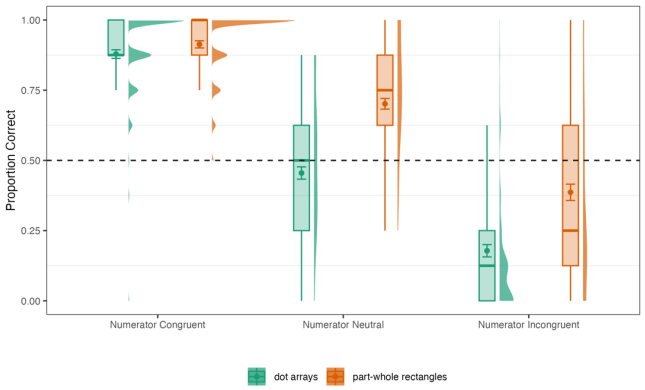

Figure 4: Proportion of all trials correct (y-axis) on the magnitude comparison task across the three trial types (x-axis), when comparing dot arrays (green) or rectangles (orange). Box plots represent the interquartile range (box), full range (whiskers), and median (lines). Means are represented using points, with error bars +/- one standard errors. Half violin plots use a smoothed probability density function to display individual level data. Data is presented summarized at the individual participant level for data visualization purposes only.

To compare performance on rectangles vs. dot arrays on each trial type separately, we used a series of binomial mixed effects models each with a fixed effect of format. On numerator congruent trials, there was not a significant effect of format, b = -0.42, SE = 0.23, p = .068, OR = 0.65, with adults' performance well above chance with both rectangles, M = .91, p < .001, and dots, M = .88, p < .001. On numerator neutral trials, there was a significant effect of format, b = -1.11, SE = 0.13, p < .001, OR = 0.33, with above chance performance with rectangles, M = .70, p < .001, and performance significantly below chance with dots, M = .46, p = .043. On numerator incongruent trials, there was again a significant effect of format, b = -1.40, SE = 0.23, p < .001, OR = 0.25, with below chance performance with both rectangles, M = .39, p < .001, and dots, M = .18, p < .001.

To compare performance within a format across trial types, we again use a series of binomial mixed effects models on each format separately, with trial type as a fixed effect using the neutral trials as the reference variable for congruent and incongruent trials. Relative to numerator neutral trials, performance was significantly higher on congruent trials for both rectangles, b = 1.63, SE = 0.15, p < .001, OR = 5.10, and for dots, b = 2.30, SE = 0.13, p < .001, OR = 9.98, and significantly lower on incongruent trials for both rectangles, b = -1.52, SE = 0.12, p < .001, OR = 0.22, and for dots, b = -1.45, SE = 0.12, p < .001, OR = 0.23.

Errors on the Match-to-Sample Task

We designed the MTS task to include four trial types: exact matches (regardless of foil type), equivalent matches with a non-interfering foil, equivalent matches with a numerator foil, and equivalent matches with a total foil (Figure 5). Planned analyses included two kinds of comparisons: format effects within each trial type separately and within each format, the effect of having a potentially interfering foil on equivalent match trials.

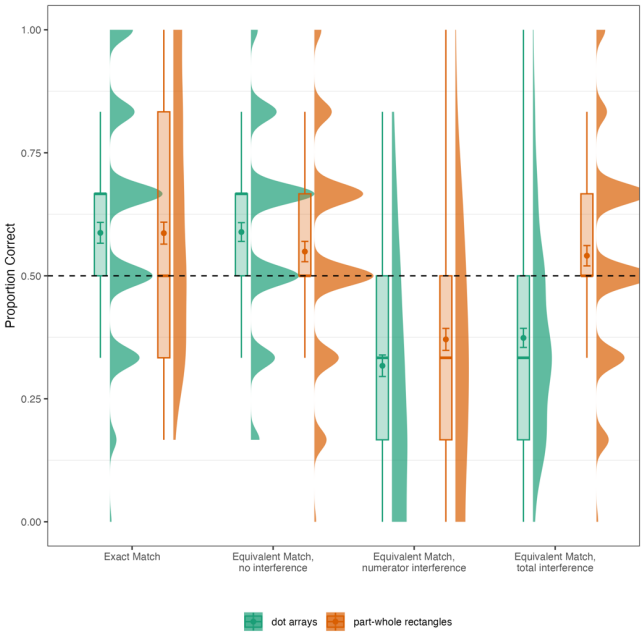

Figure 5: Proportion of all trials correct (y-axis) on the MTS task across the four trial types (x-axis), when comparing dot arrays (green) or rectangles (orange). Box plots represent the interquartile range (box), full range (whiskers), and median (lines); points represent means +/- one standard error. Half violin plots use a smoothed probability density function to display individual level data.

To compare performance on part-whole rectangles vs. dot arrays on each trial type separately, we again used a series of binomial mixed effects models with a fixed effect of format. There was not a significant effect of format for exact match trials, b = -0.003, SE = 0.13, p = .984, OR = 1.0, equivalent match non-interference trials, b = -0.16, SE = 0.12, p = .164, OR = 0.85, or equivalent match numerator interference trials, b = 0.25, SE = 0.14, p = .080, OR = 1.29. On exact match trials and equivalent match non-interference trials, adults are above chance for both dots, exact: M = .59, equivalent non-interference: M = .59, and rectangles, exact: M = .59, equivalent non-interference: M = .55, ps < .02. In contrast, adults were significantly below chance for both dots, M = .32, and rectangles, M = .37, on numerator interference trials, ps < .001. On equivalent match total interference trials, there was a significant effect of format, b = 0.68, SE = 0.12, p < .001, OR = 1.97, with performance significantly below chance with dots, M = .37, p < .001, and not significantly different from chance with rectangles, M = .54, p = .051.

To investigate the effect of having a potentially interfering alternative foil, we again use binomial mixed effects models with trial type as a fixed effect using the equivalent match non-interference trials as the reference variable for numerator

foil and total foil trials, on data from the dot arrays and rectangles separately. Relative to the non-interference trials, performance was significantly lower on numerator foil trials for rectangles, b = -0.73, SE = 0.12, p < .001, OR = 0.43, but not significantly different on total trials, b = -0.03, SE = 0.12, p = .768, OR = 0.97. In contrast, performance was significantly lower, again relative to the non-interference trials, on numerator foil trials, b = -1.13, SE = 0.12, p < .001, OR = 0.32, and total foil trials, b = -0.88, SE = 0.12, p < .001, OR = 0.42, for dots.

Discussion

In the current study, adults completed two kinds of proportion judgement tasks: magnitude comparison and equivalence matching, with two formats: part-whole rectangles and randomly arranged discrete dot arrays. Overall, we replicate prior work with both kinds of tasks: adults show more numerator interference with discrete dot arrays than with continuous rectangles. However, our comparison across tasks and specific kinds of errors also reveals both quantitative and qualitative differences in performance.

First, we highlight three interesting results about overall task performance: (1) the comparison task was generally easier than the MTS task, particularly for rectangles, (2) format effects were much larger for the comparison task than the MTS task, and (3) performance between the two tasks was correlated for rectangles, but not for dots.

The general difference between the magnitude comparison and MTS task is consistent with speculation reported in prior work interpreting differences across studies (e.g., Boyer et al., 2024; Hurst & Cordes, 2018). It may be that having a third stimulus, as opposed to just two, increases the difficulty by requiring more domain general resources. It is worth noting that participants were given more time with visible stimuli on the MTS task (we equated the tasks to 600ms per stimulus, resulting in 1200ms on the comparison task and 1800ms on the MTS task), and yet performance was still lower.

Alternatively, it may be that reasoning about *equivalence* between proportions is more difficult than reasoning about *relative magnitudes,* and that people are better able to extract proportional magnitudes from rectangles than dot arrays. Previous research comparing MTS and magnitude comparison tasks for absolute number suggests that numerical abstraction may be more difficult in MTS tasks relative to comparison tasks (Cantlon et al., 2007). Given that the proportion MTS task requires even further abstraction to equivalence classes of proportions, it seems possible that difficulty abstracting numerical information would also impede proportion matching. Although participants did poorly on the MTS task for both dots and rectangles, they performed even worse for dots. Thus, it may be that abstracting proportional information in the context of an equivalence task is difficult for both rectangles and dots, but exacerbated with the discrete stimuli, where extracting proportional information is also difficult even in a relatively simpler magnitude comparison task.

Finally, our tasks may have captured individual differences in this ability, specifically for rectangles: there was a significant correlation between performance on the magnitude comparison and MTS tasks. One possibility is that the shared individual difference is participants' ability to extract proportional information from rectangles. In contrast, the lack of correlation between tasks with dot arrays suggests that participants may be using fundamentally different strategies that tap different underlying skills across the two tasks. For example, it may be that individual differences with dot arrays are less attributable to proportional reasoning and instead due to idiosyncratic biases and errors.

Second, analyses of specific kinds of errors further highlight the difficulty with dot arrays. Replicating previous work, we find that when both numerator and proportional information is consistent, comparing both dot arrays and rectangles was well above chance, whereas when the absolute numerator is inconsistent with the proportion, performance was significantly worse and below chance for both the comparison task (e.g., Jeong et al., 2007; Hurst & Piantadosi, 2024) and the MTS task (e.g., Boyer et al., 2008; Hurst et al., 2021). However, our inclusion of the neutral trials on the comparison task and exact match and equivalent match no-interference trials on the MTS task provide further insight. First, adults have difficulty comparing proportions presented as dot arrays even when interfering information about the numerator is not available. That is, in the absence of consistent numerator information, participants were at chance comparing dot arrays; it's not *just* that dot arrays are more likely to result in interference from the numerator, but also that the proportional information itself is less readily available. This suggests that adults may have been behaving on the dot-version of this task as if it were a *numerical* comparison task, not a *proportion* comparison task. Second, on the MTS task, participants were susceptible to numerator interference for both dots and rectangles but only total interference from dots. Further, even on exact match trials and no-interference trials, performance on the MTS task was low (though, above chance) for both dots and rectangles. These findings suggest that the MTS task may have made it more difficult to extract proportional information effectively even from rectangles.

Future work could provide further insight into interpreting adults' low performance with dots (on both tasks) and on the match-to-sample task (with both formats) by comparing different stimuli, paradigms, or instructions that may encourage proportional abstraction and by investigating the processes underlying adults' extraction (or lack thereof) of proportional information.

In summary, adults showed lower proportional reasoning with a match-to-sample task (vs. a magnitude comparison task) and with dot arrays (vs. rectangles), suggesting that they may have difficulty extracting proportion magnitudes from these contexts. Further, interactions between task and format, as well as the specific patterns of errors, highlights the need to carefully consider task design when making inferences about both children's and adults' proportional reasoning.

Acknowledgements

Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number R00HD104990 to MAH. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bonawitz, E., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. Cognitive Psychology, 74, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003

Boyer, T. W., Bradley, L., & Branch Greer, N. (2024). Children's understanding of relative quantities: Probability judgement and proportion matching. Cognitive Development, 69, 101411. https://doi.org/10.1016/j.cogdev.2023.101411

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong. Developmental Psychology, 44(5), 1478–1490. https://doi.org/10.1037/a0013110

Brysbaert, M. (2019). How Many Participants Do We Have to Include in Properly Powered Experiments? A Tutorial of Power Analysis with Reference Tables. Journal of Cognition, 2(1). https://doi.org/10.5334/joc.72

Cantlon, J. F., Fink, R., Safford, K., & Brannon, E. M. (2007). Heterogeneity impairs numerical matching but not numerical ordering in preschool children. Developmental Science, 10(4), 431–440. https://doi.org/10.1111/j.1467-7687.2007.00597.x

Denison, S., & Xu, F. (2014). The origins of probabilistic inference in human infants. Cognition, 130(3), 335–347. https://doi.org/10.1016/j.cognition.2013.12.001

Hurst, M. A., Boyer, T. W., & Cordes, S. (2021). Spontaneous and directed attention to number and proportion. Journal of Experimental Psychology: Learning, Memory, and Cognition. https://doi.org/10.1037/xlm0001084

Hurst, M. A., & Cordes, S. (2016). Rational-number comparison across notation: Fractions, decimals, and whole numbers. Journal of Experimental Psychology: Human Perception and Performance, 42(2), 281–293. https://doi.org/10.1037/xhp0000140

Hurst, M. A., & Cordes, S. (2018). Attending to relations: Proportional reasoning in 3- to 6-year-old children. Developmental Psychology, 54(3), 428–439. https://doi.org/10.1037/dev0000440

Hurst, M. A., & Piantadosi, S. T. (2024). Continuous and discrete proportion elicit different cognitive strategies. Cognition, 252, 105918. https://doi.org/10.1016/j.cognition.2024.105918

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. Trends in Cognitive Sciences, 20(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Kalra, P. B., Hubbard, E. M., & Matthews, P. G. (2020). Taking the relational structure of fractions seriously: Relational reasoning predicts fraction knowledge in elementary school children. Contemporary Educational Psychology, 62, 101896. https://doi.org/10.1016/j.cedpsych.2020.101896

Kassambara, A. (2023). Rstatix: Pipe-friendly framework for basic statistical tests. Retrieved from https://CRAN.R-project.org/package=rstatix

Kay, M. (2024). ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. IEEE Transactions on Visualization and Computer Graphics, 30(1), 414–424. https://doi.org/10.1109/TVCG.2023.3327195

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Park, Y., Viegut, A. A., & Matthews, P. G. (2020). More than the sum of its parts: Exploring the development of ratio magnitude versus simple magnitude perception. Developmental Science. https://doi.org/10.1111/desc.13043

R Core Team. (2023). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Thompson, C. A., Taber, J. M., Sidney, P. G., Fitzsimmons, C., Mielicki, M., Matthews, P. G., Schemmel, E., simonovic, N., Foust, J., Aurora, P., Seah, T. H. S., Disabato, D., & Coifman, K. (2020). Math matters during a pandemic: A novel, brief educational intervention combats whole number bias to improve health decision-making and predicts COVID-19 risk perceptions and worry across 10 days [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/hukyv

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686