Quantitative reasoning is facilitated by cross-format comparison

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Singh, Amritpal

Howell, Madison

Hurst, Michelle A

Publication Date

2026

Quantitative reasoning is facilitated by cross-format comparison

Amritpal Singh (as3695@cornell.edu)[1,2], Madison Howell[2], Michelle A. Hurst[1,2]
[1]Department of Psychology, Rutgers University, New Brunswick
[2]Rutgers Center for Cognitive Science, Rutgers University, New Brunswick

Abstract

Proportional reasoning is a widespread form of quantitative thinking. Proportion also provides an opportunity to investigate the mechanisms that determine what algorithms we use to reason about quantity. Prior work shows that the visual features of proportional information (e.g., whether it is depicted continuously vs. discretely) influence whether thinkers adopt more optimal or suboptimal strategies. However, these studies investigated strategy selection when all proportions were presented in a single visual format. We hypothesize that comparing proportions *across* formats (continuous "blobs" vs. discrete dot arrays) will discourage suboptimal numerator strategies by making numerical information unavailable for comparison across proportions and encourage using more reliable part-whole relations within each proportion. Here, we report a study that tested this hypothesis. Participants completed a proportion comparison task where they compared proportions either within- or across-formats. Results support the hypothesis that cross-format comparison deters suboptimal numerator strategies and promotes more optimal part-whole quantitative reasoning strategies.

Keywords: Quantitative Reasoning; Proportion; Relational Reasoning; Reasoning Strategies

Introduction

Proportional reasoning is a kind of quantitative thinking that is crucial for a wide variety of everyday situations, ranging from adjusting a dinner recipe to deciding upon an equitable distribution of resources. Proportion also provides a useful testbed for investigating the forces that shape how we reason about quantity and quantitative relations. In most situations, a thinker can deploy any one of multiple strategies to reason about proportional information. Perhaps surprisingly, prior work demonstrates that people often adopt suboptimal reasoning strategies. For example, if asked to judge whether 5/9 or 3/4 is larger in a comparison task, children and adults often mistakenly judge 5/9 to be larger than 3/4 because 5 > 3, even though 55% < 75%. In this case, people appear to use a strategy that over-relies on isolated information about the numerator, while ignoring the part-whole relation between the numerators and denominators. But what factors determine the algorithms we use to reason about quantitative information?

Prior work suggests that the visual features of proportional information (i.e., its visual format) can shape the strategies individuals use in those different perceptual contexts. For example, proportions can be depicted with discrete, multicolored arrays of dots or with contiguous shapes like rectangles, pie charts, or amorphous blobs. In tasks when people must judge which of two proportions is larger, they are more likely to use a suboptimal numerator strategy when both proportions are depicted as discrete sets of dots or discretized rectangles (e.g., rectangles with lines that demarcate countable units) than when the proportions are depicted as continuous, integrated visual forms, like rectangles and pie charts (Abreu-Mendoza et al., 2023; Boyer et al., 2008; Hurst & Levine, 2026; Hurst & Piantadosi, 2024)[1]. However, previous studies have only investigated how visual features, like discreteness, influence whether people use numerator-based or proportion-based strategies when both proportions involved in the comparison are presented using the same visual format (i.e., a set of dots vs. a set of dots; a rectangle vs. a rectangle).

It is possible that people are more likely to adopt a suboptimal numerator strategy when the numerators of both proportions appear in a discrete visual format because numerical information is available and easily comparable relative to continuous visual formats. However, a proportional reasoning strategy that relies on comparing part-whole relations from within each proportion would not require the two proportions to appear in the same visual format. For example, the proportions 75% and 50% are readily comparable, even if the 75% is depicted with amorphous blobs and the 50% is depicted with dot arrays. If this reasoning is true, then proportional information presented in a mixture of distinct visual formats will cue more optimal reasoning strategies than when proportions are presented in a uniformly discrete format. That is, reasoning about proportions that appear in *distinct* visual formats may discourage the use of suboptimal numerator strategies and encourage the use of more reliable part-whole proportional strategies.

[1]In proportional reasoning tasks that use non-symbolic depictions of proportions (e.g., rectangles, dot arrays, and blobs), participants must compare which of two visual representations show a greater proportion of a particular target color, such as orange (e.g., a greater proportion of the area that is orange or a greater proportion of the number of dots that are orange). We use the term "numerator" to refer to the component of the stimulus that is this target color.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

We report here a study that investigated this hypothesis by having participants complete a proportional reasoning task in which we manipulated the visual format of proportion stimuli within-subject. Participants completed a proportion comparison task in which they judged which of two proportions was larger when (1) both proportions were depicted by continuous, amorphous "blob" areas, (2) both proportions were depicted by discrete multicolored dot arrays, or (3) one proportion was depicted as a blob, and the other proportion was depicted as a dot array.

To investigate if participants over-rely on numerator, part information, we include trials that are both numerator congruent (where the larger proportion contains a larger numerator, such as 5/9 vs. 8/10) and numerator incongruent (where the larger proportion contains a smaller numerator, such as 5/9 vs. 3/4). If participants use a numerator-driven strategy, there will be a larger difference between their accuracy in the congruent and incongruent trials, which we characterize as greater numerator interference.

We predict that, if our hypothesis is correct, we should observe less numerator interference (i.e., a smaller difference in accuracy between numerator congruent and incongruent trials) for cross-format proportion comparisons than for comparisons where both proportions appear as dots.

Materials and Methods

Participants

Participants were recruited from a university student research pool and compensated with class credit. One hundred and thirty-four participants completed the consent form. Ten participants did not move on to the study from the consent form. From the 124 participants for whom we have data, we excluded data from: 2 participants who did not complete the study, 11 participants who reported not paying attention during the study, and 3 participants for whom we had to exclude more than 50% of the trials (i.e., more than 32 trials) from any one format condition for responses that were either faster than 200ms or slower than three standard deviations beyond the mean response time for trials in that format condition. At the trial level, we excluded 42 trials (0.20%) where responses were faster than 200ms and 221 trials (1.07%) where responses were slower than three standard deviations beyond the mean response time within that format condition. Data from 108 participants were analyzed. Demographics for participants were: M_{age} = 18.64 YO, $Range_{age}$ = 18-22 YO; 65 female, 42 male, 1 nonbinary; 39 Asian, 28 White, 15 Black or African American, 13 Hispanic or Latino/a/e, 10 multiracial, 2 Middle Eastern, 1 missing data; 77 with high school degree or GED, 26 with at least one year of college, 1 with associate's degree, 3 with Bachelor's degree, 1 missing data.

Materials

The study was programmed using jsPsych (de Leeuw et al., 2023), and implemented using Pavlovia.org, an online behavioral research platform. The consent form and an optional demographics form were administered using Qualtrics.

Stimuli The Python Image Library was used to create the stimulus images used for the proportion comparison task. Each 400x600 pixels image presented a proportion as an array of discrete orange and blue dots or as a continuous irregular area with orange and blue subcomponents (blobs, see Fig. 1). For example, 8/27 with orange as the target color could be presented as 8 orange dots and 19 blue dots, for a total of 27 dots. Alternatively, 8/27 could be presented as an irregular continuous blob in which the ratio of the orange component's area to the blue component's area would be 8:19 and the surface area was equated to the cumulative surface area of the corresponding number of dots. That is, the area covered by 8 orange dots when presenting 8/27 was equal to the area of the orange component in a blob stimulus presenting 8/27. Because all dots were equal in size, set numerosity correlated with set area such that a set of 18 dots covered twice the area of a set of 9 dots.

We randomly generated a list of proportions such that any subset had a minimum of five discrete elements and a maximum of fifty-five discrete elements. This list of proportions was used to construct sixty-four unique stimulus pairs for test trials. The ratio between the two proportions (i.e., larger proportion / smaller proportion) varied between 1.5 and 3.0 (with an equal number of pairs falling into the following ranges: (a) 1.5 to 1.875, (b) 1.875 to 2.25, (c) 2.25 to 2.625, and (d) 2.625 to 3.0. We also constructed six stimulus pairs of proportions for practice trials.

Thirty-two of the pairs were numerator congruent pairs (i.e., the larger proportion also contained the larger numerator), and thirty-two were numerator incongruent (i.e., the larger proportion contained the smaller numerator). In both numerator congruent and numerator incongruent pairs, sixteen trials had the correct answer (larger proportion) appear on the left side of the screen, and half had the correct answer (larger proportion) appear on the right. For each stimulus image, we created one version in which the numerator components appeared as orange dots or blob areas and the non-numerator components appeared as blue dots or blob areas. A second version swapped which component was which color so that we could counterbalance across participants whether the numerator was orange or blue.

Stimulus pairs were depicted as (1) two irregular blob areas (i.e., blobs-only condition), (2) two dot arrays (i.e., dots-only condition), or (3) one blob area and one dot array (i.e., cross-format condition; see Fig. 1). The blobs and dots always appeared against a gray background. For each participant, thirty-two trials in the cross-format block were randomly assigned to have the left proportion appear as blobs, and thirty-two trials to have the right proportion appear as blobs.

Procedure

We used a similar procedure and trial structure to previous studies investigating proportional reasoning strategies (Hurst & Levine, 2026; Hurst & Piantadosi, 2024). Participants were

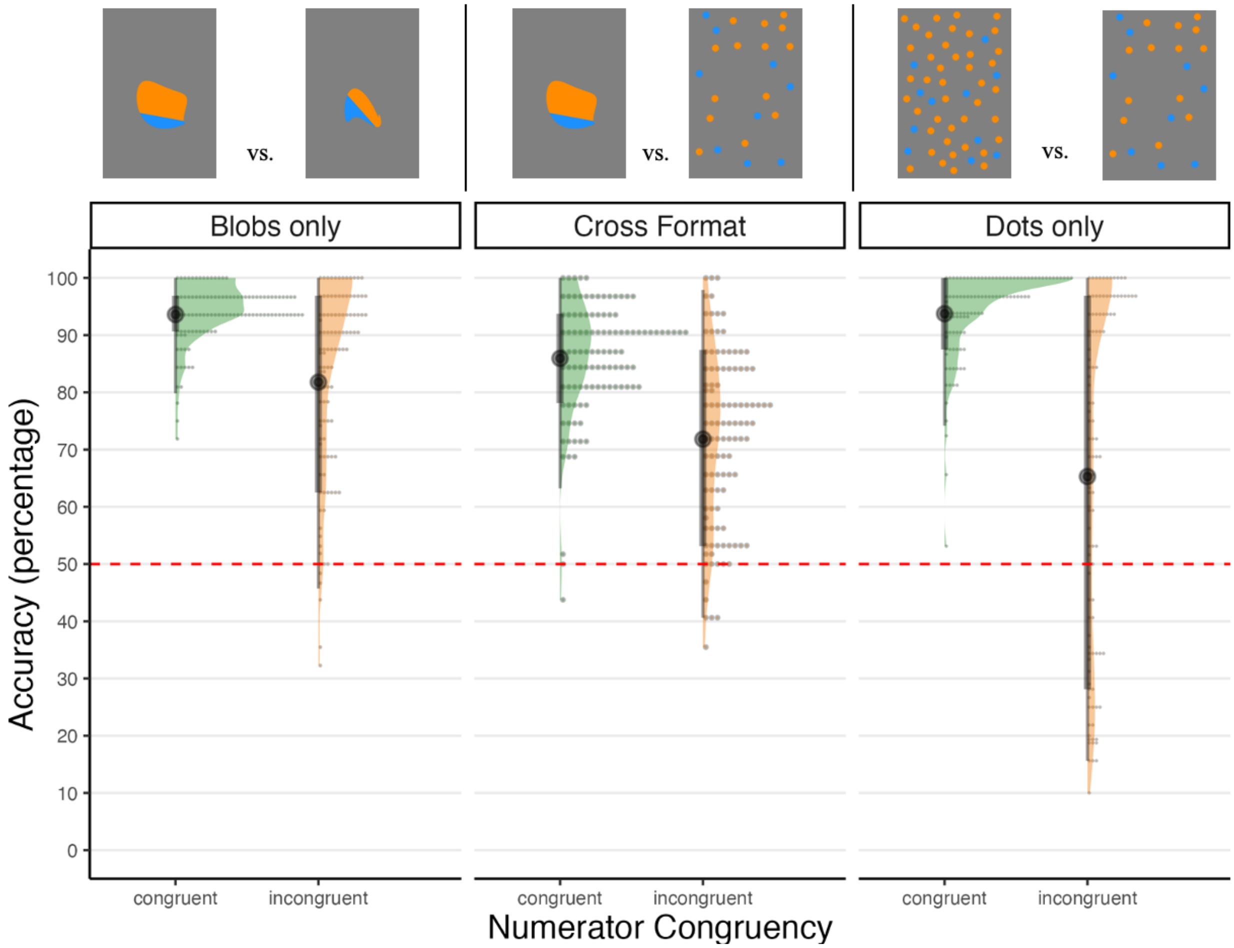

Figure 1. Average accuracy across within-subject conditions (blobs-only vs. cross-format vs. dots-only) and trial type (numerator congruent vs. numerator incongruent). Each colored dot represents a participants' average accuracy for those trials. The shaded regions represent the distributions of subject-wise accuracy (green for numerator congruent and orange for numerator incongruent trials). The black dots represent the mean accuracy for that format condition and trial-type. Black bars represent the interquartile range. Above each facet are examples of stimuli from that condition.

pseudo-randomly assigned to complete the study in one of the six possible orders using the Qualtrics randomizer and its capacity to present conditions roughly an equal number of times. Similarly, participants were assigned to see the numerator as orange and the non-numerator as blue, or vice-versa.

After filling out the consent and demographics forms, the experiment launched in full screen on the participant's machine. Instructions with examples of the stimuli explained that the participant's task was to judge which of the two stimulus images in each pair displayed the larger proportion of a target color (orange or blue) with example images on the screen. Participants completed three blocks of six practice trials and sixty-four test trials each. Each block corresponded to one format condition (i.e., either blobs-only, dots-only, or cross-format condition). Thus, participants completed 192 total test trials with visual format of proportion comparisons manipulated within-subject.

In each trial, a mask was presented for 500ms. Then the proportion stimuli were presented for up to 1200ms along with a prompt to choose which image displayed the larger proportion of the numerator color by pressing either the left or right arrow key. If the participant responded before the 1200ms passed, the trial ended, and the task would progress. If the participant took longer than 1200ms from stimulus onset, the stimulus images would disappear and only the prompt would remain on the screen until the participant made their choice.

When the participants completed the first and second block, a message would appear informing the participant about how far they had progressed and how much further they had to go. Once the participant was ready, they pressed the spacebar to move on to the next block. After the third block, participants answered a question about whether they paid attention throughout the study (yes, sometimes, or no as multiple-choice options). Finally, participants were asked to share any comments or problems they encountered during the study before being returned to the university's SONA page for them to receive course credit for their participation.

Data Analysis

Data were analyzed using the data wrangling packages from the tidyverse (version 2.0.0; Wickham et al., 2019), the plotting package ggplot2 (version 3.5.1; Wickham, 2016), and the statistical packages lme4 (version 1.1-35.5; Bates et al., 2015) and emmeans (version 1.10.4; Lenth, 2024) using R (version 4.4.2; R Core Team, 2024) in rStudio (Posit team, 2024) to run mixed-effects linear models (Baayen et al., 2008). We use the Holm-Bonferroni method (Holm, 1974) to control for family-wise error. We report the Holm-Bonferroni corrected α criterion when the correction was applied.

For all analyses reported, we used model comparison methods. We would test if a fully specified model was significantly different from a reduced model that was identical to the full model except that it would exclude a particular term of interest. For example, a full model might analyze accuracy as a function of the fixed effects of visual format condition, numerator congruency, and their interaction, as well as a random intercept for subjects. The reduced model would contain the fixed effects of visual format condition and numerator congruency, as well as the random intercept for subjects but exclude the interaction term from the full model.

Table 1: Means and standard deviations of subject-wise accuracy for each visual format and numerator congruency trial type.

Format Condition	Numerator Congruent	Numerator Incongruent
Blobs	0.94 (0.05)	0.82 (0.17)
Cross Format	0.86 (0.10)	0.72 (0.15)
Dots	0.94 (0.08)	0.65 (0.30)

Results

First, when comparing a model that analyzed accuracy (binary, 0, 1) as a function of numerator congruency with a random intercept for subjects to a reduced model that excluded the numerator congruency term, we found that numerator congruency contributed significantly to the model in each format condition after correcting for 3 comparisons. These results reflect significant numerator interference (i.e., lower accuracy on numerator incongruent trials than congruent trials) in all three conditions: Blobs-only, $\chi^2(1) = 246.59$, $p < .001$, OR = 3.58, 95% CI [3.03, 4.23], Holm-Bonferroni $\alpha = .025$; dots-only, $\chi^2(1) = 1111.2$, $p < .001$, OR = 12.9, 95% CI [10.8, 15.4], Holm-Bonferroni $\alpha = .0167$; cross-format, $\chi^2(1) = 222.6$, $p < .001$, OR = 2.54, 95% CI [2.24, 2.88], Holm-Bonferroni $\alpha = .05$ (see Table 1).

Replicating prior work, we find that numerator interference was significantly stronger in the dots-only condition than the blobs-only condition, $\chi^2(1) = 77.87$, $p < .001$, $OR_{interaction} = 2.89$, 95% CI [2.29, 3.65], Holm-Bonferroni $\alpha = .025$. However, the critical analyses that test our hypothesis compare numerator interference between the cross-format condition and the two within-format conditions (i.e., test for significant two-by-two interactions of numerator congruency and format condition when comparing the cross-format and blobs-only condition and when comparing the cross-format and dots-only condition).

Numerator interference was significantly weaker in the cross-format condition than in the dots-only condition, $\chi^2(1) = 159.25$, $p < .001$, $OR_{interaction} = 0.28$, 95% CI [0.22, 0.34], Holm-Bonferroni $\alpha = .0167$, but significantly greater than in the blobs-only condition, $\chi^2(1) = 7.38$, $p = .005$, $OR_{interaction} = 0.75$, 95% CI [0.61, 0.92], Holm-Bonferroni $\alpha = .05$ (see Figure 1). This pattern suggests that being required to compare two proportions of different formats decreased reliance on numerator information relative to when both proportions were presented discretely. At the same time, however, comparing proportions across formats did not reduce numerator interference to the same degree as presenting fully continuous blob stimuli. Numerator interference was slightly (two percentage points) but significantly larger in the cross-format condition than the blobs-only condition.

Interestingly, accuracy was significantly lower for cross-format comparisons than blobs-only comparisons in *both* numerator congruent, $\chi^2(1) = 112.93$, $p < .001$, $OR_{interaction} = 2.43$, 95% CI [2.05, 2.87], Holm-Bonferroni $\alpha = .025$, and numerator incongruent trials, $\chi^2(1) = 104.92$, $p < .001$, $OR_{interaction} = 1.87$, 95% CI [1.65, 2.10], Holm-Bonferroni $\alpha = .025$ (see Figure 1, Table 1). Participants' accuracy was also significantly lower for cross-format congruent trials than for dots-only congruent trials, $\chi^2(1) = 122.16$, $p < .001$, $OR_{interaction} = 0.39$, 95% CI [0.33, 0.46], Holm-Bonferroni $\alpha = .0167$. Further, accuracy did not differ between the blobs-only congruent trials and the dots-only congruent trials ($p = .80$). The lack of a difference between blobs-only and dots-only congruent trials, and the significant lower accuracy in cross-format congruent trials than in the congruent trials of the other two conditions suggests that that participants were not simply relying more on numerator information in the cross-format condition than in the blobs-only condition. Instead, participants may have faced greater difficulty comparing proportions across formats than comparing proportions depicted as blobs due to other factors beyond numerator interference (see Discussion).

Discussion

The primary goal of the study reported here was to test whether comparing proportional information across distinct formats would deter people from using suboptimal component-based reasoning heuristics and instead foster more reliable part-whole-based reasoning strategies. Our study results provide evidence in support of this hypothesis by finding that participants experienced significantly less numerator interference in the cross-format comparison condition than in the discrete dots-only condition. These results suggest that participants were less likely to deploy a numerator strategy when comparing proportions of two different visual formats, presumably because numerical information was not available for comparison across stimuli.

Instead, participants may have used a more optimal strategy in which they compared part-whole relations from within each proportion.

A significant interaction of format by numerator congruency suggested that cross-format comparisons still showed more numerator interference on average than comparisons of area-based blob stimuli that contained no discrete information. However, the difference between numerator congruent and incongruent trials (i.e., the numerator interference effect) differed by only two percentage points between cross-format condition than the blobs-only condition. We will test for this interaction in a pre-registered replication with a larger sample size to determine if the interaction observed in this initial study is reliable and replicates. However, in this initial study participants were less accurate for the cross-format incongruent trials than the blobs-only incongruent trials. Participants were also less accurate for the cross-format congruent trials than the blobs-only or the dots-only congruent trials. This pattern in the congruent trials is notable because a numerator strategy would produce high accuracy on congruent trials, not low accuracy. Thus, the overall pattern of results we observe is not readily explained by participants having relied on numerator information in the cross-format condition.

An alternative explanation is that using a proportional part-whole strategy may be more difficult with dots than blobs per se. This may be the case if cross-format comparisons require participants (1) to implement a proportion strategy that is more complex than a numerator-based strategy because it incorporates information about both parts and wholes, and (2) to inhibit attention to numerical information in the discrete dot stimulus that is irrelevant due to the lack of numerical numerator information in the blob stimulus. If this idea is correct, then it should be more difficult to deploy a proportion strategy in the cross-format comparisons (which contained one dot array per trial) than in the blobs-only comparisons (which contained no dot arrays). In this way, comparing proportions across discrete and continuous formats may have impeded performance more generally than comparing two continuous stimuli, leading to the differences we observed between our blobs-only and cross-format conditions, while still deterring participants from using a numerator-based strategy.

If true, this explanation would dissolve what might appear to be a potential conflict between our results and a previous developmental study (Boyer et al., 2008) that investigated children's proportional reasoning. Boyer et al. (2008) reported no accuracy differences between conditions where children compared (1) two proportions depicted as continuous integrated rectangles (i.e., rectangle analogs to our blob stimuli) and (2) a continuous integrated rectangle to a continuous *discretized* rectangle (i.e., a continuous rectangle with lines drawn on it to make numerical information available). If it is true that it is more difficult to use proportional information for fully *discrete* entities than *continuous* ones, then Boyer et al. (2008) may have observed no difference between their continuous vs. continuous matching task and continuous vs. discretized matching task because all proportions in both conditions were depicted with integrated rectangles. The difference we observed between blobs-only and cross-format conditions would be due to our use of fully discrete dot arrays in the cross-format condition. However, additional work is needed to establish if it is truly more difficult to attend to proportional information with fully discrete stimuli than continuous stimuli.

Further, our study differs from Boyer et al. (2008) along multiple dimensions that may also impact quantitative reasoning (e.g., the use of a comparison task versus a match-to-sample task, differences between our populations, discriminability of specific proportions used, and structure of the stimuli used across experiments), so other factors may also have played a role in producing differences between our study results and theirs.

This initial experiment motivates future studies to further investigate the perceptual cues and reasoning strategies used by participants to compare proportions. For example, a study could vary the size of individual dots in the dot stimulus images to eliminate participants' ability to use cumulative surface area when comparing proportions. Such a study could then establish if participants compare cross-format proportion representations by comparing the relative surface areas of surface components or by representing and comparing the proportion depicted in each image. Further, computational modeling approaches could be used to compare how likely it is that participants relied on more relational proportion-based strategies or numerator-based strategies in cross-format comparisons relative to when the proportions to be compared are represented in a single visual format (e.g., using methods similar to Hurst & Piantadosi, 2024).

These results provide evidence that visual cues can lead people to use different reasoning algorithms even when the underlying quantitative information they need to reason about is identical across visual formats. Specifically, the different visual formats in our task appeared to lead participants to adopt either a suboptimal numerator strategy that was more likely to produce reasoning errors, or a more optimal proportion strategy that was more accurate due to its incorporation of part-whole relational information. Participants adopted different strategies even though the underlying proportions being depicted were identical across format conditions, thus producing variation in their judgements and resulting reasoning accuracy. Thus, this study also contributes to a growing body of behavioral (Abreu-Mendoz et al., 2023; Boyer et al., 2008; Hurst et al., 2020) and computational work (Hurst & Piantadosi, 2024) that highlights the need to systematically investigate the causes of variation in reasoning and behavior. The findings of this study, and the previous work, suggest that such variation may be driven by concrete visual features cueing different cognitive strategies in post-perceptual reasoning.

Altogether, these results support the view that comparing proportions across formats discourages component-based strategies and facilitates the use of part-whole proportion relations. However, requiring a cross-format comparison did not eliminate numerator interference to the same degree as comparing two continuous visual representations. Overall, this study sheds new light on the mechanisms by which determine how adults reason about different forms of quantitative information.

References

Abreu-Mendoza, R. A., Powell, A. B., Renninger, K. A., Rivera, L. M., Vulic, J., Weimar, S., & Rosenberg-Lee, M. (2023). Middle-schoolers' misconceptions in discretized nonsymbolic proportional reasoning explain fraction biases better than their continuous reasoning: Evidence from correlation and cluster analyses. *Cognitive Psychology*, *143*, 101575. https://doi.org/10.1016/j.cogpsych.2023.101575

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

Bates D. M., Mächler M., Bolker B., & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1–48. doi:10.18637/jss.v067.i01.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Boyer, T. W., Levine, S. C., & Huttenlocher, J. (2008). Development of proportional reasoning: where young children go wrong. *Developmental Psychology*, *44*(5), 1478.

De Leeuw, J. R., Gilbert, R. A., & Luchterhandt, B. (2023). jsPsych: Enabling an open-source collaborative ecosystem of behavioral experiments. *Journal of Open Source Software*, *8*(85), 5351.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. *6*(2): 65–70.

Hurst, M. A., & Piantadosi, S. T. (2024). Continuous and discrete proportion elicit different cognitive strategies. *Cognition*, *252*, 105918.

Hurst, M. A., Shaw, A., Chernyak, N., & Levine, S. C. (2020). Giving a larger amount or a larger proportion: Stimulus format impacts children's social evaluations. *Developmental Psychology*, *56*(12), 2212.

Lenth, R. (2024). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.10.4, https://rvlenth.github.io/emmeans/.

Posit team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, **4**(43), 1686. doi:10.21105/joss.01686.