Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Exp Psychol Gen. Author manuscript; available in PMC 2026 January 27.

Systematic Variation in Proportion Judgments: Spatial features impact adults' strategies and decisions

Michelle A. Hurst[1], **Susan C. Levine**[2]

[1]Rutgers University, New Brunswick

[2]University of Chicago

Abstract

Proportional information is important for a range of everyday actions, from infants' and toddler's probabilistic inferences to adults' medical and financial decisions. Unfortunately, children and adults frequently make systematic errors in some proportional reasoning contexts. For example, people tend to focus more on the numerators, rather than the proportional relations, when proportions are discrete (i.e., with enumerable units) or when the sub-components are spatially separated. Importantly, it is not that people cannot reason proportionally, as they do not make these same errors with continuous proportions presented as part of a single coherent whole. Although format-dependent variation has been shown across many studies with both children and adults, no work has systematically manipulated multiple aspects of visual, non-symbolic proportional stimuli simultaneously to better understand which spatial factors impact proportional reasoning, and how. Here, we manipulate proportional stimuli in three ways: the availability of enumerable units (i.e., discreteness), predictability of the proportional information, and spatial separateness of the proportion sub-components. We also formalize competing strategy explanations using mathematical models to infer people's strategies. Overall, we find that discreteness, predictability, and spatial separatenesss (as operationalized here) significantly impact adults' performance and strategies. Furthermore, all features interact with each other, and qualitative patterns suggest that spatial separateness and predictability may be particularly important, despite being less well studied. By systematically varying the spatial features of proportions, we provide insight into the mechanisms that underlie proportional reasoning and highlight important interactions between spatial, numerical, and relational information.

Keywords

proportion; strategies; mixture model; variation

Understanding how people reason about proportional information is critical, both theoretically, for cognitive theories of development and learning, and in practice, such as in math education. For example, non-symbolic proportions lay the foundation for

Correspondence concerning this article should be addressed to Michelle A. Hurst, 152 Frelinghuysen Rd, Piscataway, NJ 08854. michelle.hurst@rutgers.edu.

The authors made the following contributions. Michelle A. Hurst: Conceptualization, Data curation, Formal analysis, Funding acquisition, Writing - Original Draft Preparation, Writing - Review & Editing; Susan C. Levine: Conceptualization, Resources, Writing - Review & Editing.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

understanding symbolic fractions and ratios (Matthews et al., 2016) and play a key role in people's ability to make probabilistic inferences from proportional information (Denison & Xu, 2012). However, children and adults are not consistent in how they reason about proportion. Instead, research has revealed substantial variation, including variation across contexts, such as probability vs. juice mixing (Boyer et al., 2024) and across stimuli, such as proportional displays with numerical information available vs. no numerical information (Boyer et al., 2008; Hurst & Piantadosi, 2024) or spatially separated components vs. a single integrated whole (Hurst et al., 2020; Möhring et al., 2016). However, this variation is often studied with isolated contrasts between a small number of contexts or stimuli, limiting our ability to build a complete explanation of the underlying cause(s) of these difficulties (see Figure 1 for common contrasts used in previous studies). This issue is important because without understanding the cause(s) of this variation, we cannot have a complete understanding of the processes and mechanisms involved in proportional reasoning, which in turn limits our ability to build theories that depend on proportional thinking or to develop interventions for targeting proportion learning. In the current set of experiments, we aim to increase understanding of the cause(s) of variation in proportional reasoning by systematically modifying perceptual features of the proportion display and measuring format-dependent performance and strategy use.

Variation Across Discrete and Continuous Proportion

Substantial work has compared proportional reasoning with discrete stimuli, where numerical units are available, to proportional reasoning with continuous stimuli where numerical units are not available. The way discreteness is instantiated has varied across studies in a few ways, including divided rectangles, divided annuli, and entirely discrete sets of dots (see Figure 1, Panel A). Additionally, for some studies, number and area can both be used within the same stimulus comparison; that is, because the numerical unit is a consistent size, more units will also correspond with more area, either within a part-whole shape or cumulative area across a set of dots (Boyer & Levine, 2015; Boyer et al., 2008; Hurst & Piantadosi, 2024; Hurst et al., 2020). In other cases, number and area are prevented from co-varying so that the proportional information must be based on number alone and not cumulative surface area (Hurst & Piantadosi, 2024). Across these many studies, and subtle differences in instantiation, it is typically found that people show worse proportional reasoning with discrete stimuli, compared to continuous stimuli. For example, when asked to compare two circular game spinners and decide which has a higher probability of resulting in a red (vs. blue) outcome, 6-year-old children succeed when the spinners are not divided but score significantly lower when the spinners are divided into unit sized pieces (e.g., divided into eighths with six units colored in) (Hurst & Cordes, 2018; Jeong et al., 2007). Importantly, this decrease in performance with discrete stimuli arises specifically when the numerical information in the numerator is in conflict with the proportional information; that is, when the option with more red pieces (i.e., highest number) has a lower proportion colored in red (i.e., lower proportion/probability). This same pattern has been found in a range of contexts that involve comparing proportional information, including matching juice and water mixtures (Boyer et al., 2008), making social evaluations based on resource distributions (Hurst et al., 2020), and interpreting the quantifier word "most" (Hurst &

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Levine, 2022). Across contexts, decreased proportion responding with discrete stimuli is often attributed to children using a numerical strategy that focuses on the number of items in the numerator (i.e., the most salient subset), potentially due to an overuse of counting (Boyer et al., 2008).

However, a counting-based explanation is likely insufficient for fully explaining the difference between discrete and continuous stimuli, as a similar pattern of errors is found when adults make judgements about proportions created from large sets of intermixed dots (Fabbri et al., 2012; Hurst et al., 2021), which are presumably less likely to be counted. An alternative possibility is that adults are encoding and comparing an *approximate* representation of the number of items in a subset as a heuristic (Hurst & Piantadosi, 2024). Importantly, unlike counting, which is unique to discrete number, using an approximate representation of absolute magnitudes could, at least in principle, be used for all proportional values based on *any* type of magnitude (e.g., area, length, time). For example, it could be that having a larger numerator in terms of area would similarly bias proportional judgements. Thus, one possibility is that some aspects of typical discrete stimuli make the numerator information more salient, and therefore this kind of error more likely, but that other visual or spatial features would similarly elicit this behavioral pattern.

Coherent Whole vs. Separated Components

There is some evidence that non-numerical spatial features do impact proportional reasoning: spatially separating the components involved in a proportion decreases attention to the proportional information, compared to when the quantities are part of an integrated whole (see Figure 1, Panel B). For example, when asked to make social evaluations based on resource distribution (i.e., who is “nicer” after seeing how much/many of their resource(s) they shared), when the non-shared and shared resources were spatially separated children made more numerator-based errors than when the components were spatially integrated, despite no numerical information being available in either case (Hurst et al., 2020). Additionally, children’s ability to mentally scale visual proportions presented as part of an integrated whole was significantly correlated with their fraction ability, but their mental scaling of spatially separated visual proportions (i.e., proportions as two separate subsets) was not (Möhring et al., 2016). Considered together, these findings suggest that spatially separating the quantities involved in a proportion decreases attention to proportional information.

Other Visualspatial Features

The prior work has primarily considered differences between discrete vs. continuous proportion or spatially separated vs. integrated components of proportions, without explicitly comparing across them, examining how they interact, or examining the impact of other visual-spatial features (which are sometimes varied incidentally, such as presenting proportional information in the context of rectangles vs. annuli). There is some evidence to suggest that other visual-spatial features also impact behavior (Hurst & Piantadosi, 2024; Park et al., 2020). For example, Hurst and Piantadosi (2024) investigated children’s and adults’ strategy use when comparing proportion or ratio-based stimuli in different forms

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that included variation within continuous (e.g., line lengths, pie charts, blobs) and discrete (e.g., dots that were all equal in size vs. varied in size) stimuli. Their goal was to compare mathematical models of strategy use across discrete and continuous stimuli, and their results are broadly in line with this distinction. However, there was also variation *within* continuous and discrete categories. For example, there was a weaker dominant strategy (and in some cases, not a clearly preferred strategy) when continuous stimuli were irregular blobs (vs. pie charts) and when dots were all equal in size (vs. varied in size). This variation suggests that other spatial features beyond the distinctions of discrete vs. continuous and spatially separated vs. integrated components may contribute to variation in proportional reasoning processes and behavior.

Operationalizing Visual and Spatial Features of Proportional Displays

In summary, the prior work suggests that focusing on only broad-strokes distinctions one at a time may be leading to incomplete and/or incorrect conclusions about the cause(s) of variation. To move toward a more complete understanding of the processes and mechanisms involved in proportional reasoning, we need more general principles about how perceptual and contextual features cause variation in the cognitive strategies we use to compare proportions. Here, we aimed to address this issue by systematically modifying proportional stimuli in multiple ways: the availability of enumerable units (i.e., discreteness), having spatially separated vs. integrated sub-components (i.e., separate parts to be integrated into a “whole” vs. a single “whole” divided into two parts), and the predictability or ease of extracting proportional information from the whole object or scene, due to structural regularities of the format (see Figure 2 for examples of all formats used in the current studies). Although we operationalize these as three features for the sake of simplifying the stimulus space, it is worth noting that we do not conceive of these as natural or true categories with strong boundaries. Instead, they likely interact with each other in ways that lead to different affordances and different biases. Our goal is to use these features to systematically investigate the stimulus space and better understand what these distinct affordances and biases may be, how they interact with each other, and how they may arise.

As in prior work, we operationalized discreteness as the availability of discrete enumerable units. Previous studies have typically used two kinds of discrete proportional stimuli: entirely separate objects, such as an array of dots (Fabbri et al., 2012; Hurst & Piantadosi, 2024) or demarcated units within an object, sometimes called discretized stimuli (Abreu-Mendoza et al., 2023; Boyer et al., 2008). Additionally, discrete proportional stimuli have typically been generated from small and countable sets (Boyer et al., 2008) and/or large sets that are not countable in practice given the timeframe (Hurst & Piantadosi, 2024). For the purposes of the current experiment, the presence of enumerable units is the defining feature of discrete stimuli, regardless of whether they are countable in practice and/or entirely separable objects. Additionally, in the current study, we investigate the presence of numerical information in addition to surface area information. Although substantial research investigating absolute magnitude representations has focused on distinguishing between the processing of numerical information from other magnitude dimensions, such as area and contour (Cantrell & Smith, 2013; Gebuis & Reynvoet, 2012; Leibovich et al., 2017; Lourenco & Aulet, 2022; Starr et al., 2017), we do not address this issue here. In

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

fact, although numerical information is available in our discrete stimuli, it is possible for participants to rely exclusively on (cumulative) surface area. Importantly, however, they cannot rely on *absolute area* – just as they cannot rely on *absolute number* - instead, proportional reasoning is required, whether it be based on the relative areas or the relative numbers.

We operationalized spatial separateness based on whether the two sub-components that get combined to create a proportional whole were spatially integrated or separated. By spatially separating the two sub-quantities, we may be deemphasizing the whole (i.e., the amount the proportion is in reference to) and/or emphasizing a single sub-quantity (e.g., the numerator amount) making it more difficult or less likely to process the proportional information. Although this pattern has been shown for rectangles (Hurst et al., 2020; Möhring et al., 2016), it is an open question of whether the availability of the whole vs. emphasis on separate parts also drives behavior with other formats (e.g., discrete sets) and whether this spatial feature interacts with other spatial features.

Finally, although substantial prior work has investigated behavior with continuous area-based proportion, these stimuli can be quite variable, including round stimuli, such as pie charts or annuli (Abreu-Mendoza et al., 2020; Hurst & Cordes, 2018; Jeong et al., 2007; Mock et al., 2018), rectangles (Abreu-Mendoza et al., 2023; Boyer & Levine, 2015; Boyer et al., 2008), or amorphous blobs (Hurst & Piantadosi, 2024; Park et al., 2020). No work, to our knowledge, has systematically compared how the shape of the continuous proportional amount impacts proportional reasoning. We operationalize this as predictability[1] because blobs, rectangles, and pie-charts (as circles or annuli) vary in structural ways that may impact the ease of extracting proportional information, as well as in their familiarity, which may similarly impact predictability or ease of processing. Blobs are unpredictable because they are irregular; in other words, it is not possible to attend to only one feature to calculate the proportion. Instead, we must fully encode the area of each component and/or the area of the entire blob. In contrast, the more structured rectangles and canonical pie charts may allow us to encode the proportion value by focusing on only one dimension, simplifying the extraction of proportional information. For example, because the two subcomponents within a rectangle are standard in width, you could encode the proportion by just looking at the height of the two components and/or the height of one component relative to the total height – ignoring the width. Similarly for pie charts or annuli, the angles of the slices can be compared, ignoring the size of the circle and the overall area. Pie charts are also highly familiar, and it's possible that people can quickly encode and compare the proportional values from only one angle (i.e., the angle of the numerator slice), automatically encoding it in relation to the complete circle. In this case, pie charts may be even more predictable than rectangles.

[1]In our pre-registration, we refer to this feature as regularity instead of predictability. Although regularity can contribute to predictability, predictability better captures our interpretation of the differences between these stimuli.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

The Current Study

In the current study, we aimed to systematically investigate stimulus driven variation in proportional reasoning. To do so, we present three experiments with adult participants using a novel probability comparison task with 12 different stimulus formats (see Figure 2). Although each experiment targets specific contrasts within and across our operationalized features, we had three overarching research questions:

1. For what kinds of stimuli are adults sensitive to the congruency between the size of the numerator and the overall proportion?

2. How do each of operationalized stimulus features impact proportion comparison behavior? Furthermore, are these effects independent, overlapping, or interacting with each other?

3. How does strategy use vary across formats?

In Experiment 1, we target all three research questions with 8 distinct formats: integrated and separated blobs, integrated and separated rectangles, integrated and separated dot clouds, and integrated and separated dot grids. This design allows us to contrast behavior with proportional displays that are discrete and continuous, have spatially integrated and separated sub-components, and differ in their shape or structure in ways that contribute to the predictability of the proportional information – as well as several distinct combinations of these features. In Experiment 2, we incorporate a commonly used continuous stimulus that was omitted from Experiment 1, the pie chart, to contrast behavior with proportions displayed as pie charts (continuous, spatially integrated, highly predictable) and proportions displayed as intermixed dots (discrete, spatially intermixed, unpredictable). In Experiment 3, we take a more careful look at discreteness and variations across shape by investigating the added presence of discrete numerical units within an integrated shape and potential differences in proportional reasoning with pie charts and rectangles.

Previous work has addressed these kinds of questions using proportional comparison tasks with direct instructions (e.g., "which has a higher proportion colored in red?") or using a probability cover story based on a sampling outcome (e.g., "which has a higher probability of the spinner landing on red?", "which has a higher probability of pulling a red ball from the machine?"). Here, we had two constraints that made these typical approaches inappropriate. First, we wanted to use a paradigm that would be appropriate for future work with children, to better understand the development of strategy use. Second, we wanted to use a paradigm and procedure that would be equally applicable for different stimuli with identical instructions. Thus, we could not use the direct instructions approach because young children do not know what "proportion" means and often require a cover story to understand the instructions. We also could not use common probability scenarios that apply only to specific visual information (e.g., spinning a spinner only applies to circles, pulling a ball randomly from a box only applies to discrete sets). Instead, we opted to modify the typical probabilistic sampling approaches to be applicable to (almost) any visual presentation of proportional information. Specifically, we introduce participants to a "magic ball" that only lands on the two colors in the proportion and ask them which of two images we should use if we want the magic ball to land on a given color. Although we did not use these words, this

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

is akin to asking: if we were to select a pixel at random from each image (i.e., a randomly selected landing spot for the magic ball), which picture is most likely to result in a pixel of the desired color (we discuss the cases that this paradigm cannot easily cover as a limitation in the General Discussion).

Experiment 1

In Experiment 1, we compare all three features by operationalizing: sets of dots as discrete and undivided shapes as continuous; rectangles and grids as more predictable than blobs and dot clouds, respectively; and the blue and orange sub-parts being spatially separated contrasted with intermixed or contiguous colors as part of a single set (discrete) or shape (continuous).

Based on prior work, we predicted that people would show interference from the size of the numerator for all formats but would have lower performance and rely less on proportion-based strategies with discrete stimuli than with continuous stimuli, spatially separated stimuli than integrated stimuli, and the less predictable stimuli than the more predictable stimuli. Importantly however, we also explored how these features may interact in nuanced ways, though without specific a priori predictions.

Method

Participants: Using Prolific, we collected data from 400 adults (M_{age} = 35.48 years, SD = 11.48) in the United States. Following our pre-registered data exclusion procedure, 3 adults were excluded, leaving a final sample of 397 (data exclusion details provided in Data Analysis). Although we pre-registered collecting data from 50 participants per condition for our between subject design, because of uneven drop out on Prolific, the sample size varied slightly across the eight conditions, as follows: 49 compared separated blobs, 47 compared integrated blobs, 51 compared separated rectangles, 50 compared integrated rectangles, 48 compared separated random dot clouds, 49 compared integrated random dot clouds, 49 compared spatially separated dot grids, and 54 compared integrated dot grids.

Through built-in prolific screeners, we required participants to be fluent in English (self-reported) and to have not participated in our previous related pilot studies. Based on self-report (see demographic questions in Supplemental Materials), the sample was 60% men, 38% women, and 2% nonbinary or genderfluid. Most participants reported having at least a bachelor’s degree or equivalent (58%). Race and ethnicity was reported as 77% white, 9% Asian, 6% Black or African American, 1% American Indian or Alaskan Native, and 7% selecting more than one race. Additionally, 4% reported being Hispanic or Latine.

Stimuli and Material: During the task, adults were shown two images side by side that included one or more blue and orange shapes on a grey background. All stimuli were created programmatically using a lab designed Python script and were the same across participants within the same condition (see Figure 2, Panel A for an example of each kind of stimulus). Regardless of the visual format, the same proportion comparison pairs were used across all conditions (see Appendix for a list of all stimulus values). To capture differences in both the proportion magnitudes and the magnitudes of individual components, surface area of

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

the continuous stimuli was matched to the cumulative surface area summed across dots on the equivalent discrete trial. To do this, we use a labeling convention that assigns one dot as a single unit and the area corresponding to one dot as the single unit for continuous stimuli. For example, the proportion 2/5 would have half the total amount as the proportion 4/10 (i.e., 5 vs. 10) and half the numerator amount (i.e., 2 vs. 4), in terms of area for both continuous shapes and discrete sets, as well as in terms of number for discrete sets. Importantly though, both 2/5 and 4/10 would have the same relative distribution between orange and blue (i.e., 40% the numerator color).

Stimuli were created so that total set size (based on the above convention for determining "size" of area-based stimuli) could range from 14 to 53, with each subset ranging from 4 to 41. Proportions could range from 0.167 to 0.864. There were 60 unique test trials, each shown twice (once with the correct answer on the left and once with the correct answer on the right). On half the trials, the numerator was congruent with the proportional response (e.g., 10/21 vs. 25/43) and on half the trials, the numerator was incongruent with the proportional response (e.g., 12/20 vs. 15/29). The trials were selected so that the ratio between the numerators was similar to the ratio between the proportions on average (1.52 and 1.44, respectively) and so that on 50% of the trials there was a larger difference between numerator magnitudes than proportion magnitudes, and on the remaining 50% there was a larger difference between proportion magnitudes than numerator magnitudes. An additional four trials were used during practice and were repeated during the test blocks but not included in the analyses.

Integrated and Separated Dot Clouds (Discrete, Unpredictable). Dot clouds were sets of dots, where each dot represented one unit and was colored either orange or blue. The dots were all the same size, meaning that the proportion of dots that were a specific color could be determined based on the number of dots or the cumulative surface area. Importantly however, to correctly compare proportions the relative proportional information was still necessary, regardless of whether the underlying source quantities were based on number and/or cumulative surface area. In the integrated dot clouds, the dots were randomly placed within the image boundary, at least 10 pixels apart, with the color of each dot randomly assigned so that the blue and orange subsets were intermixed. For the spatially separated dot clouds, the set of blue dots and set of orange dots were spatially separated so that the numerator set (i.e., the set in the target color) was on the left side of the image and the remaining set was on the right side of the image with a gap between them.

Integrated and Separated Dot Grids (Discrete, Predictable). Dot grids were created by organizing the dots into a grid with rows of 10. The grid was then randomly placed within the image boundaries. For the integrated dot grid, the dots were randomly distributed in terms of color so that they were intermixed within and across rows. In the spatially separated dot grids, the target color (i.e., numerator) was always used first, followed by the other color, which began immediately after the numerator set was completed (i.e., the transition from orange to blue, or vice versa, could occur at any point within a row). Notably, this instantiation of spatial separateness within dot grids is distinct from the dot

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

clouds, rectangles, and blobs, which we come back to in the Discussion when interpreting the pattern of results.

Integrated and Separated Blobs (Continuous, Unpredictable). Blobs were created by generating a random set of points on an x-y grid and using the Graham's scan algorithm to find the convex hull of those points, then smoothing along this outer edge and scaling the blob on the x-y plane to create the required approximate total area. For the integrated blob, the orange and blue components were presented as part of a single blob that was randomly placed within the image boundaries. The two sub-components were created by adding a straight line between two points (the start point was randomly selected, and the endpoint was determined based on the required proportion). For the separated blobs, two blobs were created, one orange and one blue, and placed next to each other, with a set space between them. The numerator color was always placed to the left of the other color and the blobs were center aligned.

Integrated and Separated Rectangles (Continuous, Predictable). Rectangles were created by randomly selecting a width between a set minimum and maximum value and then calculating the required height to match the required total area for that stimulus. For the integrated rectangle, the two sub-components were stacked on top of each other to create a single rectangle with the bottom and top of the rectangle colored the assigned numerator color and the remaining color, respectively. The rectangle was then randomly placed within the image boundaries. For the separated rectangles, two rectangles were created, one orange and one blue, and placed next to each other with the numerator on the left, a set space between them, and aligning on the lower edge

Procedure: The study was completed entirely online. We required participants use a device with an external keyboard (i.e., a laptop or a desktop) but did not restrict screen size. Consent and demographics occurred on Qualtrics and the proportion comparison task was programmed in jsPsych and presented via cognition.run. In Qualtrics, adults were randomly assigned to one of two colors to use as the numerator when making judgements (blue or orange) and one of the eight visual formats.

After the consent and demographic survey, participants were redirected to the proportion task where their screen was automatically converted to full screen. During the instruction phase, adults were told about a magic ball that could only land on blue and orange. Adults were shown an image in their condition-specific format with 85% colored the assigned numerator color and told about the probabilistic relation between the proportion display and the magic ball outcome. For example, if the participant was assigned to have orange as the primary color, then 85% of the colored portion was orange and the remaining 15% blue. They would then be told that if the magic ball was thrown at the image 100 times, it would land on orange about 85 times, and that we would say the image had an 85% probability of resulting in an orange outcome. Adults were then told they would see two images simultaneously and their task was to select which image had a higher probability of an orange outcome as quickly as they could. After the instruction phase, adults completed eight practice trials with feedback (i.e., told if they were correct or incorrect). The eight practice trials included four very easy problems (11/14 vs. 2/20; 2/40 vs. 46/50; 38/40

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

vs. 3/50; 3/14 vs. 18/20) and four problems randomly selected from the test trials. After practice, adults completed 124 total trials (the 120 test trials, plus the four easy practice trials) randomly separated into two blocks of 62 trials each, with a pause screen in between if participants wanted a break. The four easy problems used during practice were repeated during the test block to provide a measure of sustained attention. On each trial, the two images remained visible for a maximum of 1200 ms, but participants had unlimited time to respond (they could respond while the images were visible or after). After responding, a blank screen was displayed for 500 ms and the next trial began (Figure 3).

After the proportion comparison task, participants were asked to self-report whether they were paying attention (multiple choice: yes, sometimes, or no), to recall what decision they were asked to make (open ended text box), what strategy they used (open ended text box), and any comments about the study or their experience (open ended text box).

Transparency and Openness: We transparently report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The sample size, design, and analysis plan were preregistered on the Open Science Framework (https://osf.io/vb5dm/) and all materials, data, and code is also available (https://osf.io/zqxg7/). All analyses that were not included in the preregistration are described as exploratory. All procedures for Experiments 1 and 2 were approved by the University of Chicago Institutional Review Board (IRB17–1599, "Relational Math Reasoning"), and participants provided informed consent prior to participation.

Data Analysis. We used R (Version 4.4.2; R Core Team, 2024) and the R-packages *broom.mixed* (Version 0.2.9.6; Bolker & Robinson, 2024), *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *flextable* (Version 0.9.7; Gohel & Skintzos, 2024), *forcats* (Version 1.0.0; Wickham, 2023a), *ggdist* (Version 3.3.2; Kay, 2024), *ggplot2* (Version 3.5.1; Wickham, 2016), *ggpubr* (Version 0.6.0; Kassambara, 2023), *lubridate* (Version 1.9.3; Grolemund & Wickham, 2011), *purrr* (Version 1.0.2; Wickham & Henry, 2023), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *sjPlot* (Version 2.8.17; Lüdecke, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.4; Barth, 2023) for all data wrangling and visualization. The manuscript was written as a reproducible report in RMarkdown using *papaja* (Version 0.1.3; Aust & Barth, 2024). We fit mixed effects models using *lme4* (Version 1.1.35.5; Bates, Mächler, Bolker, & Walker, 2015) with restricted maximum likelihood and calculated p-values using *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017). Odds ratios are provided as estimates of effect sizes by taking the exponential of the raw coefficients and confidence intervals of the odds ratios were calculated using the Wald method. Additionally, we used *rstan* (Version 2.32.6; Stan Development Team, 2024) to run the Bayesian strategy models.

As pre-registered, we excluded individual trials that were less than 200ms or more than 10000 ms (0.22% of trials). We excluded entire participants from analyses if they self-reported not paying attention ($n = 1$), reported substantial technical issues ($n = 1$), or had 50% or more of their trial-level data excluded ($n = 1$).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

To address our first two research questions, we use mixed effects models, with random intercepts for participant and item and fixed effects as noted to address each research question. To address our third research question, we use model-based strategy analysis.

Strategy Models. We mathematically formalize three possible strategies: a guessing strategy, a numerator comparison strategy, and a proportion comparison strategy. Guessing is a baseline strategy, defined as a 0.5 probability of selecting the correct response on any given trial, regardless of the trial features. The numerator comparison strategy is based on models of the approximate number system (ANS) (Feigenson et al., 2004; Odic & Starr, 2018). For this strategy, we assume that people encode each of the two *numerators* separately (n_1 and n_2), which are each represented as a normal distribution centered on the true value with a standard deviation of $w * n$, where w, often referred to as the Weber's fraction, models the precision of the individual's ANS. The probability of selecting the larger numerator is then computed as the difference between these two normal distributions (Equation 1; Φ is the cumulative normal distribution). We estimated w with an exponential(1) prior and restricting $w > 0$ (Piantadosi, 2016).

$$\Phi\left[\frac{|n_1 - n_2|}{w\sqrt{(n_1^2 + n_2^2)}}\right]$$

(Equation 1)

To formalize a proportion comparison strategy, we used a one-cycle power model for estimating the value of each of the two proportions (Hollands & Dyre, 2000), then took the difference between these estimates as a predictor of the probability of success using a binomial function (Equation 2). This model includes two parameters. The first parameter, β, is modeled using an exponential(1) prior and is based on the psychophysics of Stevens' power law, which claims that the perceived value of a stimulus can be best represented as a power of its true value (Stevens, 1957). This same parameter is applied to each of the two subsets - the numerator amount and the remaining non-numerator amount - and then the two components are combined using basic arithmetic (numerator estimate/(numerator estimate + non-numerator component estimate)). The second parameter, B_1 is modeled with a normal(0, 3) prior and is a scale parameter multiplied by the absolute difference between the two proportion stimulus estimates. The Hollands and Dyre (2000) proportion model is a model of proportion estimation, and there is not a clear way to model the *comparison* of two proportions. Thus, the difference between the two estimates provides a simple approach and the scale parameter captures the strength of the association between this difference and accuracy. Previous work using this model on similar tasks has found that when comparing the fit of four proportion strategy models the strategy based on the one-cycle power model fit similarly to the other proportion models in most cases (Hurst & Piantadosi, 2024).

$$B_1\left|\frac{n_1^{\beta}}{r_1^{\beta} + n_1^{\beta}} - \frac{n_2^{\beta}}{r_2^{\beta} + n_2^{\beta}}\right|$$

(Equation 2)

To formalize our strategy comparison, we used a mixture model that simultaneously estimated the parameters required for each of the three models described above and a vector of three probability weights, one for each model. We model the accuracy on each trial using a Bernoulli function with the probability of success as a weighted probability of success marginalized over the three strategy models, using a softmax function. Parameters were inferred using a No-U-Turn sampler (Hoffman & Gelman, 2014) sampling four chains each with 5000 iterations, using 50% as warm up. We ran our mixture model on each individual participants' data separately, resulting in estimated probability weights on a person-level. Model diagnostics showed appropriately large effective sample sizes (minimum = 2,184.32) and all $\widehat{R} = 1$, suggesting the models did converge (Stan Development Team, 2024). One participant showed one divergent transition in the spatially integrated blobs condition, but we have opted to retain this participant in the analyses, as the overall results are identical when this individual is included vs. excluded.

Results

Participants scored very well on the "easy" items, both when presented during practice with feedback, $M = 0.97$, and during the test trials without feedback, $M = 0.99$. For the remaining primary analyses, we only include the trials during the two test blocks (i.e., without feedback) and do not include these "easy" trials (both with and without feedback). Following our pre-registered data analysis plan, we use four distinct sets of analyses to address our research questions, as well as additional exploratory analyses as noted.

For what kinds of stimuli are adults sensitive to the congruency between the size of the numerator and the overall proportion?: First, we compared behavior on trials where the numerator was congruent with the correct proportional response (e.g., 3/4 vs. 2/3) and trials where the numerator was incongruent with the correct proportional response (e.g., 4/9 vs. 2/3), using logistic mixed effects models on each of the eight visual formats separately, with congruency as a fixed effect and random intercepts for participant and item. We predicted a significant effect of numerator congruency in each of the eight formats.

In line with our predictions, there was a significant effect of numerator congruency in each of the eight formats, with higher scores when the numerator was congruent with the correct proportional response compared to when it was incongruent with the correct proportional response (see Table 1 for a summary, and Supplemental Tables S1–S8 for full results with each format). However, the size of this difference varied dramatically across formats (Figure 4, Panel A). In exploratory follow-up analyses, we compared behavior on the numerator incongruent trials to chance using similar mixed effects models, but with only a fixed intercept (i.e., a significantly positive intercept suggests a higher probability of being correct and a significantly negative intercept suggests a higher probability of being incorrect, on average). We are interpreting significance based on a Bonferroni corrected alpha for the family of eight tests (adjusted alpha = .006). Overall, we find that on the numerator incongruent trials, when comparing spatially separated blobs and spatially separated dot

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

clouds, participants were significantly more likely to select the numerator response than the proportional response (i.e., performance significantly below chance). In contrast, when comparing both integrated and separated dot grids, participants were significantly more likely to select the proportional response than the numerator response (i.e., above chance on numerator incongruent trials). In all other cases, the intercept was not significantly different from zero (note that comparing integrated rectangles was significant at an alpha of .05, which did not hold up to corrections for multiple comparisons).

How do each of the operationalized stimulus features impact proportion comparison behavior?: Second, we separately address whether each of the three features we operationalized (discreteness, spatially separateness, and predictability) impact behavior and interact with numerator congruence. To do so, we used separate logistic mixed effects models with the target feature, numerator congruency, and their interaction as fixed effects, and random intercepts for participant and item. Notably, although these three features are not natural categories and are likely to interact, investigating each independently allows us to simplify the stimulus space and to align with prior research that has considered these features in narrower comparisons. After investigating each feature independently, we turn to the full model that incorporates interactions directly, as well as analyses of inferred strategies within each format separately.

Discreteness. Both discreteness (0 = discrete, 1 = continuous; centered to be −0.50 and 0.50, respectively) and numerator congruence (0 = incongruent, 1 = congruent; centered to be −0.50 and 0.50, respectively) were centered to facilitate interpretation (full model results are in Supplemental Table S9). There was a significant main effect of discreteness, $b = -0.19$, SE = 0.09, $OR = 0.83$, 95% CI[0.70, 0.98], $z = -2.15$, $p = .031$, with lower accuracy on average for continuous trials, $M = 0.69$ ($SD = 0.15$), than for discrete trials, $M = 0.72$ ($SD = 0.16$). There was also a significant main effect of numerator congruence, as above (i.e., higher average accuracy on numerator congruent trials than incongruent trials), $b = 2.19$, SE = 0.13, $OR = 8.91$, 95% CI[6.88, 11.55], $z = 16.58$, $p < .001$. However, there was not a significant interaction, $b = 0.06$, SE = 0.05, $OR = 1.07$, 95% CI[0.96, 1.18], $z = 1.26$, $p = .207$. Thus, the pattern for discreteness was not as hypothesized and in fact is the opposite pattern than what is typically reported, with overall lower accuracy on continuous trials than discrete trials.

Spatial Separateness. Both spatial separateness (0 = separate, 1 = integrated; centered to be −0.49 and 0.51, respectively[2]) and numerator congruence were centered to facilitate interpretation (full model results are in Supplemental Table S10). There was a significant interaction between numerator congruence and spatial separateness, $b = -0.60$, SE = 0.05, $OR = 0.55$, 95% CI[0.50, 0.61], $z = -11.61$, $p < .001$. Specifically, although the predicted effect of numerator congruency (i.e., congruent > incongruent) was significant for both spatially separated and integrated stimuli (see Table 1), the effect was larger for spatially separated stimuli. Analyzed another way, for numerator congruent stimuli, there was not a significant difference between spatially separated, $M = 0.88$ ($SD = 0.07$), and integrated

[2]Because of uneven sample sizes and/or trial numbers (after exclusions), some centering is not exactly at −0.5 and 0.5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

stimuli, M = 0.87 (SD = 0.08), b = −0.10, SE = 0.07, OR = 0.91, 95% CI[0.79, 1.05], $z = -1.34$, $p = .182$. In contrast, for numerator incongruent stimuli (i.e., cases where the smaller proportion had the larger numerator) performance was significantly lower for spatially separated stimuli, M = 0.48 (SD = 0.24), than for integrated stimuli, M = 0.58 (SD = 0.23), b = 0.51, SE = 0.13, OR = 1.67, 95% CI[1.30, 2.14], $z = 4.05$, $p < .001$. Together, these complementary interpretations of the interaction suggest that the negative effect of having an incongruent numerator was larger when the stimuli were spatially separated, and put another way, the effect of having spatially separated components was larger when the numerator was incongruent.

Predictability. Both predictability (0 = less predictable, 1 = more predictable; centered to be −0.51 and 0.49, respectively) and numerator congruence were centered to facilitate interpretation (full model results are in Supplemental Table S11). As with spatial separateness, there was a significant interaction between numerator congruency and predictability, b = −0.53, SE = 0.05, OR = 0.59, 95% CI[0.53, 0.65], $z = -10.34$, $p < .001$. Again, although the predicted effect of numerator congruency (i.e., congruent > incongruent) was significant for both less predictable and more predictable stimuli (see Table 1), the effect was larger for less predictable stimuli. Unlike spatial separateness however, the simple comparisons are all significant. That is, there is a significant, though small, difference between more predictable and less predictable stimuli for numerator congruent trials, with higher performance on the more predictable stimuli, M = 0.88 (SD = 0.08), than the less predictable stimuli, M = 0.86 (SD = 0.07), b = 0.21, SE = 0.07, OR = 1.23, 95% CI[1.07, 1.42], $z = 2.85$, $p = .004$. Similarly, there was a significant and larger difference for numerator incongruent stimuli, again with more predictable stimuli resulting in higher performance, more predictable: 0.60 (SD = 0.23), less predictable: 0.46 (SD = 0.23), b = 0.81, SE = 0.12, OR = 2.25, 95% CI[1.77, 2.87], $z = 6.62$, $p < .001$. In other words, when the numerator was congruent with the proportion magnitude, performance was high regardless of the predictability of the stimulus, with only a small difference. In contrast, when the numerator was incongruent with the proportion magnitude, performance was lower for unpredictable stimuli and higher for predictable stimuli (though, still not as high as for numerator congruent stimuli).

Are the effects of stimulus features independent, overlapping, or interacting with each other?: Third, we use model comparisons to test whether the effects of discreteness, spatial separateness, and predictability explain unique variance or whether a more parsimonious model including only a subset of these features (and/or their interactions) explains maximal variance. To do so, we built a complete model with all main and interaction effects and compared that model to simplified models removing each feature (results of this full model are provided in Supplemental Table S12). When each feature was removed, the simplified model did not include any of the main or interaction effects involving that feature. Overall, we find that the full model is a better fit to the data (using chi-square tests, AIC, and BIC), than any of the simplified models (see Table 2 for model comparisons). This result suggests that the dimensions of discreteness, separateness, and predictability (as operationalized here) each explain additional variance in behavior over and above each other. Furthermore, in the full model, the four-way interaction

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

between numerator congruency, discreteness, separateness, and predictability is statistically significant, $b = 1.07$, SE = 0.21, $OR = 2.92$,95% CI[1.95, 4.39], $z = 5.17$, $p < .001$. Although the exact meaning of this interaction is difficult to interpret, it does suggest that the effect of each spatial feature cannot be interpreted separately from the others, and instead unique combinations of these features, instantiated within a single format, may drive behavior.

How does strategy use vary across formats?: In our final data analysis approach, we focus on three strategies adults may be using (guessing, numerator comparison, proportion comparison) and use a Bayesian mixture model to estimate the probability that each individual is relying on each of the three strategies in each of the stimulus formats separately. In prior work, we used a similar approach to mathematically formalizing individual strategies and used model comparison to infer which strategy best fit people's behavior (Hurst & Piantadosi, 2024). Here, we use a mixture model to derive a more continuous estimate of the probability that each individual is relying on each of the three modeled strategies, relative to each other. We also report exploratory binomial exact tests comparing the number of individuals whose numerically highest probability weight was on the numerator strategy vs. the proportion strategy (Table 3).

Estimated probability weights for each of the three strategies at an individual and summarized level are provided in Figure 4, Panel B and Table 3. Overall, we find that the probability weights on the guessing strategy were low across all formats, with no participants having this strategy as their most probable strategy. In contrast, the relative probability of the numerator and proportion strategies varied across formats. Specifically, when comparing separated blobs and separated dot clouds, average probability weights are much higher for the numerator strategy than the proportion strategy and the distribution of individual level data suggests that the numerator strategy had the highest probability weight for most participants. In contrast, when comparing dot grids, both integrated and separated, the average probability weight is higher for the proportion strategy (vs. numerator), and again the individual level distribution is skewed with most participants having a dominant proportion strategy. In the remaining format conditions (integrated blobs, separated and integrated rectangles, integrated dot clouds), the probability weight is, on average, also higher for the proportion strategy than the numerator strategy. Notably, however, the difference is less extreme, and the distribution of individual level data shows some bimodality, with the number of participants being grouped as proportion vs. numerator strategy users not significantly different than 50:50.

Although our primary interest is in the estimated mixture weights, both the numerator heuristic strategy and the proportion comparison strategy included additional parameters that were simultaneously estimated in the mixture model. Here, we broadly discuss the overall patterns (all parameter estimates are reported in Supplemental Tables S23 and S24). Estimates of the weber fraction from the numerator heuristic strategy ranged from 0.28 to 0.80 for the full sample, similar to what has been reported previously (Hurst & Piantadosi, 2024) and higher than is typically reported for weber fractions generated from absolute magnitude comparisons specifically (Odic et al., 2013). However, when looking at just the subset of the sample where the highest strategy estimate was the numerator heuristic strategy, the weber fractions were much lower, closer to what is typically reported for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

adults in an absolute magnitude comparison task (Odic et al., 2013): when comparing blobs, rectangles, or dot clouds, the weber fractions were between 0.13 – 0.16; when comparing dot grids, the estimates are slightly higher at 0.20 and 0.35. The proportion comparison strategy included a β parameter (used in the calculation of the proportion estimate) and a B_1 parameter (used as a scale parameter on the association between the difference between the two proportion estimates and the overall accuracy). As found previously in Hurst and Piantadosi (2024), β parameter estimates were larger than expected, all above one, ranging from 1.27 to 2.09. The B_1 parameter estimates ranged from 2.93 to 7.48, though this value is difficult to interpret theoretically (we discuss this more in the General Discussion).

Discussion

There are two key findings from Experiment 1. First, regardless of the format of the stimulus, adults' behavior was modulated by the congruency between the size of the numerator and the value of the overall proportion magnitude: adults made more errors when the numerator comparison was incongruent with the proportion comparison. Although many prior studies have investigated numerical interference in discrete proportional reasoning, here we reveal that this is a much broader phenomena: even when the numerator is entirely non-numerical and based on continuous area, the relative size can negatively impact proportional reasoning.

Second, despite all visual formats revealing main effects of numerator congruency, estimates of strategy use from the mathematical models revealed variability in the preferred strategy across formats. Most adults relied primarily on a numerator strategy with separated blobs and separated dot clouds – stimuli that we operationalized as having spatially separated components and less predictable proportional structure. In contrast, most adults relied on a proportion strategy with dot grids (both integrated and separated), stimuli that we operationalized as more predictable and discrete. One possible explanation for both patterns of results is that proportional strategies are inhibited (and/or numerator strategies are more salient) when the whole in the part-whole proportional relation is less clearly accessible (as in separated blobs and dot clouds) but are facilitated when the whole is readily available and predictable (as in the dot grids). The remaining formats did not show group level preferences, suggesting variability potentially due to individual differences.

At first glance, it may seem contradictory that the formats with the lowest rates of proportional behavior were spatially separated (blobs, dot clouds), but the formats with the highest rates of proportional behavior also included spatially separated variations (dot grids). However, it's worth noting an important distinction in how we operationalized separateness across formats. For dot clouds and blobs, spatial separateness was created by putting the blue and orange subcomponents next to each other with an actual space between them. In contrast, for dot grids, spatial separateness was created by putting the blue and orange subcomponents sequentially in the grid formation – making them not integrated (and therefore contrasting with the integrated dot grid), but also not spatially separated to the same extent as in the other formats. This issue arises because dot grids allow for three levels of integration/separateness that are not as easy to instantiate in either dot clouds or single shapes: entirely spatially separated (as in the separated dot clouds and blobs/rectangles),

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

sequential but not separated (as in the integrated rectangles/blobs), and fully intermixed (as in the integrated dot clouds); however, in the current study we chose only the latter two for dot grids, which are more closely aligned with different instantiations of "integrated" in the blobs/rectangles and dot clouds, respectively. It may be that the predictability of the dot grid, which remained a single contiguous dot grid for both separated and integrated variations, drove the proportional behavior found in both cases.

This explanation, however, does not explain why people were not consistently proportional with the rectangle stimuli. We predicted low errors and high use of a proportion strategy for stacked rectangles (i.e., more predictable, integrated, and continuous stimuli), but instead found evidence of numerator congruency effects on par with other formats, numerically lower use of a proportion strategy than some discrete representations, and some bimodality suggesting individual differences. Moreover, our operationalization of spatial separation for dot grids means that the *integrated* rectangles were in some ways more perceptually similar to the *separated* dot grids, and yet people were consistently proportional with the separated dot grids and inconsistent (i.e., no dominant strategy) with the integrated rectangles. One possibility is that the varied widths of the rectangles, but not the dot grids (consistent rows of 10), decreased the predictability of these stimuli, making proportional information more difficult or less salient. Previous work has shown high performance with continuous proportions that are highly predictable and familiar, such as pie charts and annuli (Hurst & Piantadosi, 2024; Jeong et al., 2007) and consistent width rectangles (Abreu-Mendoza et al., 2023; Boyer et al., 2008). One possibility is that the rectangles used here were in some way less predictable than these previously used stimuli. We probe this issue further in Experiments 2 and 3.

Experiment 2

When analyzed in terms of discreteness, Experiment 1 revealed a surprising pattern of results: participants performed more proportionally with discrete stimuli (on average, across instantiations of discrete) than continuous stimuli (on average, across instantiations of continuous). This finding is counter to what has been previously reported for both adults and children (Boyer et al., 2008; Hurst & Piantadosi, 2024; although there are some cases involving symbolic fractions where discrete stimuli lead to better performance than continuous stimuli: DeWolf et al., 2013; Hurst et al., 2022). In Experiment 2, we contrast behavior with pie charts, a highly predictable and canonical continuous representation, with intermixed dot clouds. This comparison allows us to ensure that the surprising patterns found in Experiment 1 are not due to incidental features of the paradigm (e.g., the novel "magic ball" probability scenario), data collection method (Prolific samples), or data analysis approach, and instead may be due to nuances in the format of the visual proportions.

Thus, in Experiment 2 we address the same three research questions as in Experiment 1, but isolated to the specific contrast of pie charts vs. integrated dot clouds

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Method

Participants: We collected data as in Experiment 1 from 100 adults (M_{age} = 41.33 years, SD = 14.62) in the United States, with 50 randomly assigned to the pie chart condition and 50 to the integrated dot cloud condition. Based on self-report, the sample was 45% men, 54% women, and 1% nonbinary. Just over half the participants reported having at least a bachelor's degree or equivalent (56%). Race and ethnicity was reported as 76% white, 5% Asian, 15% Black or African American, and 3% selecting more than one race. Additionally, 12% reported being Hispanic or Latine.

Stimuli, Material, and Procedure: We used an identical paradigm as Experiment 1 but administered via Pavlovia (https://pavlovia.org) and with only two between subject formats: pie charts and dot clouds. Thus, half the participants were randomly assigned to the discrete condition, which was identical to the integrated dot cloud condition in Experiment 1. The remaining participants were randomly assigned to the pie chart condition. In this condition, the stimuli were pie charts with one segment colored the numerator color and the rest colored the other color (see Figure 2). As in Experiment 1, the two colors were blue and orange and which was the numerator color was counterbalanced across participants. Other aspects of the pie charts were determined as described for the blobs in Experiment 1 and the pie charts were randomly oriented and placed within the image boundaries. The radius was determined so that the area of the pie charts matched the cumulative area of the corresponding dot stimulus.

Transparency, Openness, and Data Analysis: Experiment 2 was also preregistered (https://osf.io/uyx5s/) and all data, materials, and code are provided on the same OSF project page as Experiment 1 (https://osf.io/zqxg7/).

We used the same exclusion criteria, which resulted in no participants being excluded at the participant level and 0.07% of trial level data being excluded for having reaction times that were too short or too long. We used the same general analytical framework, but without model comparison, given that there are only two formats under consideration, resulting in easily interpretable parameters.

Results

Participants scored very well on the "easy" items, both when presented during practice with feedback, $M = 0.96$, and during the test trials without feedback, $M = 0.98$. For the remaining primary analyses, we only include the trials during the two test blocks (i.e., without feedback) and do not include these "easy" trials (with or without feedback).

For what kinds of stimuli are adults sensitive to the congruency between the size of the numerator and the overall proportion?: As in Experiment 1, we find a significant effect of numerator congruence, with higher performance on congruent trials than incongruent trials, for both the pie chart comparisons and the intermixed dot comparisons (see Table 1 for summary and Supplementary Tables S13, S14 for full results).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

How does performance vary between pie charts and dot clouds?: As in Experiment 1, we used a logistic mixed effects model with format (0 = discrete dots, 1 = continuous pie charts; centered to be −0.50 and 0.50, respectively) and numerator congruence (0 = incongruent, 1 = congruent; centered to be −0.50 and 0.50, respectively) and their interaction as fixed effects (Figure 5, Supplementary Table S15). In contrast to Experiment 1, but consistent with prior work, we find a significant main effect of format, with higher performance for pie charts compared to dot clouds, b = 0.81, SE = 0.15, OR = 2.26, 95% CI[1.68, 3.03], $z = 5.41,\ p < .001$. There was also a significant interaction between format and numerator congruence, b = −1.00, SE = 0.11, OR = 0.37, 95% CI[0.30, 0.45], $z = -9.35,\ p < .001$, revealing that the size of the numerator congruence effect (i.e., congruent > incongruent) was much larger for the discrete dot clouds than the continuous pie charts.

How does strategy use vary across format?: Estimated probability weights for each of the three strategies at an individual and summarized level are provided in Figure 5, Panel B and Table 3 (parameter estimates are provided in Supplementary Tables S23, S24 and show a similar pattern to Experiment 1). As predicted, when comparing pie charts, the average probability weights are much higher for the proportion strategy than the numerator strategy and the distribution of individual level data is highly skewed, suggesting little individual variability. When comparing dot clouds, the average probability weight is also higher for the proportion strategy than the numerator strategy, but this difference is substantially smaller and the distribution of individual level data is somewhat bimodal, indicative of individual differences (replicating Experiment 1). Specifically, it may be that some individuals are relying on a proportion strategy over a numerator strategy and others are relying more on a numerator strategy, rather than a proportion strategy when processing discrete dot clouds. Furthermore, using a t-test to directly compare the use of a proportion strategy across the two formats, we find that adults were significantly more likely to use the proportion strategy with pie charts than with the dot clouds, $\Delta M = -0.24$, 95% CI $[-0.34, -0.15],\ t(85.51) = -5.04,\ p < .001$.

Discussion

Findings from Experiment 2 suggests that the lower-than-expected performance and reliance on a proportion strategy for stacked rectangles in Experiment 1 is unlikely to be the result of the paradigm and instead are due to features of the specific formats. Notably, pie charts, rectangles, and blobs vary in terms of what information is needed to generate the proportional value. Pie charts can be interpreted using only one piece of information: the angle of the segment, while the size and radius of the pie chart can be ignored. Additionally, pie charts are likely to be familiar for adult participants, providing them with substantial practice using the angle information to infer proportions. For rectangles (in Experiment 1), the heights of both the blue and orange subcomponents (or one of the components and the whole height) must be integrated together to generate a proportion value. Finally, for blobs (in Experiment 1), the entire surface area was needed to generate a proportional value because they lacked a predictable structure entirely: each was different and varied in both the horizontal and vertical axis, making it impossible to use a single dimension alone. Thus, it may be that the predictability of common continuous representations, including pie charts

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and annuli (Hurst & Piantadosi, 2024; Jeong et al., 2007; Mock et al., 2018) is what drives the typically reported benefit of continuous representations of proportion.

However, it is still an open question as to why continuous and integrated rectangles resulted in lower proportional reasoning than we expected based on prior work. One difference between previous work and the current study is that previous studies use rectangles with consistent widths across stimuli (Abreu-Mendoza et al., 2023; Begolli et al., 2020; Boyer et al., 2008), but in the current study, the widths varied across stimuli. For example, when widths are equal a rectangle of denominator 10 will be taller than a rectangle of denominator 8, which may simplify comparisons across proportions; whereas when the width varies (as they did in Experiment 1), a rectangle of denominator 10 will have a larger surface area than a rectangle of denominator 8 but may not be taller. Although the width can be ignored to just encode the proportion from a single rectangular proportion, the comparison process *across* rectangles of different widths may still add an additional level of difficulty.

Experiment 3

Experiments 1 and 2 provide an implicit comparison of rectangles vs. pie charts, revealing more consistent proportional behavior for pie charts (in Experiment 2) than rectangles (in Experiment 1), however we did not directly compare these in the same experiment. In Experiment 3, we explicitly compare annuli (a modified donut-like pie chart) and rectangles. We choose annuli because substantial prior work has used them with both children and adults (Abreu-Mendoza et al., 2020; Hurst & Cordes, 2018; Jeong et al., 2007), and they may make it more difficult to rely on angle alone (i.e., which of the two pie charts has the larger angle).

Additionally, prior work with children has shown a difference between continuous and *discretized* shapes (i.e., rectangles with added unit demarcation lines) (Boyer et al., 2008; Hurst & Cordes, 2018; Hurst et al., 2020; Jeong et al., 2007). That is, when continuous rectangles or circles are divided into unit size pieces - and otherwise unchanged - children make systematic number-based errors that they don't make when the shapes are undivided. However, these same effects have not been shown in adults, making it unclear whether the patterns of discreteness typically reported for children using discretized stimuli may be qualitatively different than those shown for adults using fully discrete stimuli. Thus, in Experiment 3, we ask whether adults are sensitive to the *mere presence* of discrete information in *discretized* formats (i.e., continuous formats with divisions, as opposed to fully separate discrete entities such as dot clouds).

In summary, in Experiment 3, we address the same three research questions as in Experiments 1 and 2 about numerator congruency effects, format differences in performance, and format differences in strategy use. However, we include new contrasts between rectangles versus annuli and undivided (i.e., continuous) versus divided (i.e., discretized) shapes to ask more nuanced questions about the nature of format effects.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Method

Participants: We collected data as in the previous experiments from 200 adults (M_{age} = 36.02 years, SD = 11.07) in the United States. Because of uneven drop out on Prolific, the sample size varied slightly in each of the four conditions, as follows: 50 compared non-divided donut pie charts, 49 compared divided donut pie charts, 49 compared non-divided rectangles, and 52 compared divided rectangles. The sample was 45% men, 52% women, and 3% nonbinary. A little less than half the sample reported having at least a bachelor's degree or equivalent (43%). Race and ethnicity was reported as 79% white, 8% Asian, 10% Black or African American, and 3% selecting more than one race. Additionally, 8% reported being Hispanic or Latine.

Stimuli, Material, and Procedure: The paradigm was identical to that used in Experiments 1 and 2 except that adults were assigned to one of four conditions that varied along two dimensions: shape (annulus vs. rectangle) and dividedness (i.e., divided lines making it discretized or no divided lines). Rectangles were created as in Experiment 1 by randomly selecting a width and then calculating the required height. However, because the width and heights were calculated for the divided and non-divided rectangles separately, the same stimulus pairs in each of the two conditions likely had different dimensions. Annuli were created as in the pie charts in Experiment 2, but with a small circle removed from the center to make the angle less transparent (see Figure 2). The divided stimuli were created by also including divided lines to demarcate the units. Thus, like the discrete conditions (i.e., dots) in Experiments 1 and 2, the divided condition made numerical information about the subsets available through counting or estimating the number of units. However, the part-whole information typical of integrated continuous formats was *also* available, because the whole shape remained a single intact whole. Additionally, we used lighter shades of blue and orange than in Experiments 1 and 2 so that the unit lines were more visible. All other aspects of the paradigm and procedure were identical to the previous experiments.

Transparency, Openness, and Data Analysis: Due to an oversight, Experiment 3 was not preregistered. Instead, we follow the planned sample size guidelines, exclusion criteria, and analysis plan preregistered for Experiments 1 and 2. All data, materials, and code are provided on the same OSF project page as the other experiments (https://osf.io/zqxg7/).

We used the same exclusion criteria, which resulted in no participants being excluded at the participant level and 0.24% of trial level data being excluded for having reaction times that were too short or too long. Experiment 3 was approved by the Rutgers University Institutional Review Board (Pro2023001930, "Characterizing Quantitative Development").

Results

Participants scored very well on the "easy" items, both when presented during practice with feedback, $M = 0.98$, and during the test trials without feedback, $M = 0.98$. For the remaining primary analyses, we only include the trials during the two test blocks (i.e., without feedback) and do not include these "easy" trials (with or without feedback).

For what kinds of stimuli are adults sensitive to the congruency between the size of the numerator and the overall proportion?: There was a significant effect of numerator congruence in each of the four formats, with higher performance when the numerator was congruent with the correct proportional response compared to when it was incongruent with the correct proportional response (see Table 1 and Supplementary Tables S16–S19). However, the size of this difference varied across formats (Figure 6, Panel A). As in Experiment 1, we compared behavior on the numerator incongruent trials to chance and are interpreting significance based on a Bonferroni corrected alpha for the family of four tests (adjusted alpha = .012). Overall, we find that when asked to compare divided and non-divided annuli and divided rectangles, participants were significantly more likely to select the proportional response than the numerator response on numerator incongruent trials. When comparing non-divided rectangles, the intercept was not significantly different from zero (note that, as for Experiment 1, comparing non-divided, integrated rectangles was significant at an alpha of .05, though did not hold up to corrections).

How does each stimulus feature (discreteness & shape) impact proportion comparison behavior?

Discrete Units. We used a logistic mixed effects model with discreteness (0 = discrete, 1 = continuous; centered to be −0.50 and 0.50, respectively) and numerator congruence (0 = incongruent, 1 = congruent; centered to be −0.50 and 0.50, respectively) and their interaction as fixed effects (see Supplementary Table S20). There was only a significant main effect of numerator congruence, as above (i.e., higher average accuracy on numerator congruent trials than incongruent trials), b = 1.83, SE = 0.16, OR = 6.23, 95% CI[4.55, 8.52], $z = 11.44$, $p < .001$. There was not a significant main effect of discreteness, b = 0.18, SE = 0.11, OR = 1.19, 95% CI[0.97, 1.47], $z = 1.65$, $p = .098$, or an interaction, b = 0.07, SE = 0.08, OR = 1.08, 95% CI[0.93, 1.25], $z = 0.96$, $p = .338$.

Shape. We again used a logistic mixed effects model this time with shape (0 = rectangle, 1 = annulus; centered to be −0.49 and 0.51, respectively) and numerator congruence (coded as previously) and their interaction as fixed effects (see Supplementary Table S21). There was a significant interaction between numerator congruency and shape, b = −0.51, SE = 0.08, OR = 0.60, 95% CI[0.52, 0.70], $z = -6.72$, $p < .001$. Specifically, although there was a significant effect of numerator congruency for both rectangle and annulus stimuli (as discussed above), the effect was larger for rectangles. Analyzed another way, for numerator congruent stimuli, there was not a significant difference between the two shapes: comparing annuli, 0.89 (SD = 0.07), comparing rectangles, 0.89 (SD = 0.08), b = −0.03, SE = 0.11, OR = 0.97, 95% CI[0.78, 1.21], $z = -0.29$, $p = .770$. In contrast, for numerator incongruent stimuli, there was a significant difference, with higher performance with the annuli, 0.69 (SD = 0.14), than with the rectangles, 0.60 (SD = 0.20), b = 0.46, SE = 0.13, OR = 1.59, 95% CI[1.22, 2.07], $z = 3.45$, $p < .001$. In other words, as we found in Experiment 1, when the numerator and proportion values were in conflict, this had a larger negative impact on behavior for stimuli that provide less predictable ways to extract the proportion versus more predictable or simplified processes for extracting the proportion (i.e., rectangles vs. annuli in Experiment 3, blobs vs. rectangles in Experiment 1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Do these effects interact with each other?: Finally, we built a complete model with all main and interaction effects and compared that model to simplified models removing each feature in turn (results of the full model are provided in Supplemental Table S22). When each feature was removed, the simplified model did not include any of the main or interaction effects involving that feature. Overall, we find that the full model is a better fit to the data, however it is less clear cut than in Experiment 1 (see Table 4 for model comparisons). Certainly, the model removing shape as a predictor has a less good fit (using chi-square tests, AIC, and BIC) than the full model. However, the model removing discreteness shows some inconsistency across metrics - worse fit according to BIC but better fit according to AIC and the chi-square test. Here, because most of the metrics suggest the full model and the discreteness predictor significantly interacts with other variables in the full model, we interpret the full model as being a better fit. As in Experiment 1, the overall interaction in the full model between numerator congruency, discreteness, and shape is statistically significant, $b = -0.61$, SE = 0.15, $OR = 0.54$, 95% CI[0.40, 0.73], $z = -4.01$, $p < .001$. Although the exact meaning of this interaction is difficult to interpret, it does suggest that these features are not independent - consistent with Experiment 1 and the overall variation evident in Figure 6.

How does strategy use vary across formats?: In all cases, the strategy with the highest average estimate was the proportion strategy, followed by the numerator strategy, with very little reliance on guessing (Figure 6 and Table 4; see Supplemental Tables S23, S24 for parameter estimates). However, there was also substantial variability across formats, with the non-divided annulus resulting in the highest use of a proportion strategy. Using *t*-tests to compare the estimated weight of the proportion strategy across formats, we find that adding divided lines to the annuli significantly decreased reliance on a proportion strategy, $\Delta M = 0.09$, 95% CI $[0.03, 0.15]$, $t(83.52) = 2.88$, $p = .005$. Additionally, the average weight of the proportion strategy was significantly lower when comparing non-divided rectangles than when comparing non-divided annuli, $\Delta M = 0.22$, 95% CI $[0.13, 0.31]$, $t(61.41) = 4.76$, $p < .001$. However, use of a proportion strategy did not significantly differ between divided annuli and divided rectangles, $\Delta M = 0.04$, 95% CI $[-0.05, 0.13]$, $t(89.37) = 0.88$, $p = .379$ or between divided and non-divided rectangles, $\Delta M = -0.08$, 95% CI $[-0.20, 0.03]$, $t(94.98) = -1.49$, $p = .139$. It is also worth noting that, as in Experiments 1 and 2, strategy use with both the non-divided and the divided rectangles appears to be at least somewhat bimodal, suggesting there may be individual differences.

Discussion

As in Experiments 1 and 2, we find that congruency between the numerator and the overall proportion impacts behavior for all formats tested, but that the strength of this effect varied. In Experiment 3, we had two specific sources of variation to investigate: differences between annuli and rectangles, as another test of predictability in extracting proportional information by using shapes that vary in how much information is required to infer the proportion, and the presence of units, by comparing divided and non-divided shapes. First, we find higher performance on annuli than rectangles. Even with the center angle removed, the high predictability of a pie chart led to more proportional reasoning and fewer errors than the rectangles. Second, the effect of the *presence* of numerical information through unit

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

divisions was inconsistent. The full model fit the data better, on most metrics, than the model without discreteness as a factor and discreteness interacted significantly with other features in the full model. Additionally, strategy analyses reveal that the inclusion of divided lines decreased reliance on a proportion strategy for annuli, but not for rectangles. It could be that because rectangles were already more difficult, even when continuous, adding divided lines did not add to this difficulty, nor substantially change people's strategies. In contrast, for annuli, where participants were very likely to use a proportion strategy, this tendency decreased with the addition of demarcated units.

General Discussion

Across three experiments, we find substantial variation in adults' strategies and ability to compare visually presented proportions in a probability comparison task when those proportions have different visual and spatial features. Here, we summarize major conclusions that emerge across the experiments.

First, adults were impacted by the congruency between the numerator magnitude and the overall proportion magnitude for all formats tested. Given the range of formats we included, this provides robust evidence that people are not encoding holistic proportion estimates that are protected from interference, even for pie charts/annuli that are canonical, familiar, and predictable. Instead, other quantitative features - such as the size of the numerator component - are being encoded and can interfere with proportional reasoning. The generalization of this pattern across stimuli requires us to reevaluate a common interpretation of the cause of numerical interference with discrete proportion stimuli: it may not be about number (or, at least, not entirely), but rather a more general phenomena, whose strength varies due to the spatial properties of the stimulus. Together, these findings highlight the need for additional psychophysical and computational research on the encoding, representation, and comparison processes that support performance on these kinds of tasks.

Second, behavior and strategy use varied systematically as a function of the visual and spatial properties of the stimuli. Our results make clear that the spatial features interact in ways that require fully considering the affordances of each format. For example, although it is typically reported that people make more errors with proportion presented with numerical units (i.e., discretely) than proportions presented without numerical units (i.e., continuously) (Abreu-Mendoza et al., 2023; Boyer et al., 2008; Hurst & Piantadosi, 2024), by expanding the set of contrasts we find notable exceptions to this pattern. For example, we find consistent proportional strategy use with dots grids (a discrete format), but consistent numerator strategy use with separated blobs (a continuous format). What general principles might explain the breadth of variation?

The formats that had the most consistent proportional behavior (i.e., most people were classified as more likely to use a proportion strategy; above chance performance even on numerator incongruent stimuli) had the two sub-components within a single whole that was predictable and organized: dot grids, pie charts, and annuli. In contrast, the formats that had the most consistent numerator-driven behavior (i.e., most people were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

classified as more likely to use a numerator strategy; below chance performance on numerator incongruent stimuli) had spatially separated sub-components, and therefore not a single readily available whole: separated dot clouds, separated blobs. At first glance, the explanation about integrated proportional wholes might also predict that stacked rectangles would be highly proportional, but this isn't what we find. Across the experiments, we can compare four different shapes: blobs, rectangles, pie charts, and annuli. We found that people more consistently used a proportion strategy and were less mislead by inconsistent numerator information with pie charts and annuli, while blobs and rectangles both showed an inconsistent mix of proportional and numerator-driven behavior. Although we did predict that pie charts would be highly proportional and that blobs would not be (due to their high variability and lack of predictable structure), the pattern for stacked rectangles is surprising. One possibility is that the variable widths across stimuli within a trial and across trials led to increased difficulty extracting proportional information. Although the width isn't necessary for encoding proportion from a single stimulus, the comparison context may be facilitated when the dimensions used to encode the proportion (e.g., height for the rectangles) can also be compared across stimuli (e.g., a taller rectangle is a larger denominator than a shorter rectangle). Importantly, because our rectangles varied in width and the blobs varied in both height and width, the only stimulus that could be compared across stimuli within a trial using a single feature were the pie charts/annuli. We also found variation within discrete stimuli that may speak to this issue: dots organized in rows of 10 generally led to consistent proportional responding. Because the dot grids were always in rows of ten, they were in some ways closer to fixed-width rectangles, which may have increased the predictability of the proportional information. Thus, one generalizing principle across the patterns of variation may be that when the proportional whole is clearly available and easy to reference, proportional strategies are used; but when the proportional information is disrupted (e.g., components are presented separately interfering with the whole) then people may focus more on the numerator information. Future work will be needed to test this hypothesis.

Finally, although some formats strongly elicited proportional behavior and others strongly elicited numerator driven behavior, there were several (including stacked rectangles) that were more variable. In particular, the patterns of strategy use in the context of some formats looked bimodal, suggesting individual differences. It is possible that when some features facilitate proportional reasoning (e.g., integrated sub-components), but others hinder the processing of proportional information (e.g., variable widths or having multiple pieces of information that need to be compared), then there are strong individual differences in whether people can overcome the difficulties and attend to proportion, or not. Further work is needed to test these predictions and investigate the cause(s) of this individual variation across various formats (e.g., inhibitory control; Abreu-Mendoza et al. (2020)).

An important next question, based on the patterns we find with discreteness, is whether the explanation suggested here also applies to children's difficulty with discrete visual proportion. There are some differences between our study and typical studies with children that limit the generalizability of these findings to previous developmental work. Research with *discretized* stimuli in children typically include smaller set sizes (e.g., less than 10 units) and long or unlimited display times, both of which are more conducive to *counting* than our much larger set sizes (i.e., up to 53 units) and short presentation times. Thus, it

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

may be that children do apply an over-learned count routine to discrete (and importantly, discretized) stimuli that is specific to *numerical* interference. It is possible that both children and adults are susceptible to the impact of spatial and visual features investigated here, and that children also overapply a count routine when they can. This hypothesis would predict that children are more negatively impacted by the mere presence of numerical information than adults, while also being impacted by visual spatial features that make it difficult to encode the necessary proportional information, such as separating the two subcomponents or making the “whole” less salient or predictable. The relative impact of each of these spatial features across development is an important question for future work.

Taken together, these results have important implications for our understanding of the cognitive processes underlying proportion representation and reasoning. On the one hand, the stark variation across different formats suggests that a single proportion processing system is unlikely to encode and represent proportion holistically and invariably. Instead, proportional reasoning may involve encoding multiple pieces of information and executing computations, which can introduce variation and/or competition from other systems at multiple stages of this process and lead to downstream differences in behavior. Importantly, however, it remains an open question as to how and when the variation arises and what other individual differences may be relevant (e.g., inhibitory control; Abreu-Mendoza et al. (2020)). One possibility is that people’s implementation of different strategies is due to a top-down process, such that some people *opt* to use a numerator heuristic because they determine that a proportion comparison strategy is too challenging or suboptimal. If this is the case, then there may not be an automatically activated proportion processing system, and instead people use distinct computational processes, or strategies, in different contexts or because of individual differences. An alternative possibility is that the variation seen here is driven by more bottom-up processes, such that spatial and visual properties of the stimulus lead to differences in what and how information is encoded from the stimulus, in turn leading to differences in which processes are carried out. Importantly, these two explanations are not mutually exclusive and both may be inducing variation in behavior. Furthermore, the use of a counting strategy explaining some variation in children, but potentially not in adults, suggests that these two factors might vary developmentally. Additional work is needed to carefully compare these alternatives, as well as other factors might impact people’s strategy selection.

Limitations

It is worth noting some limitations of the current study. First, although we designed the “magic ball” probability paradigm to accommodate many possible spatial and visual formats, it cannot easily incorporate area-controlled discrete stimuli. This limitation arises because the “magic ball” scenario is equivalent to picking a colored pixel at random, which makes it inherently rely on overall area. Therefore, in all our stimuli, cumulative surface area was available as an information source that could have been used to compute the proportions. In other words, even for the discrete stimuli, where numerical information is available, cumulative area was also available because all the dots were the same size. If we were to vary dot size and make area not an available cue, our paradigm would need an additional explanation about the irrelevance of dot size on those trials specifically, which

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

would change the instructions across formats – something we were very careful to avoid. In previous work, using a different mathematical approach and less systematic variation within continuous and discrete stimuli, we have found that removing area as an available cue by varying dot size increases adults' tendency to rely on the numerator information, rather than proportion (Hurst & Piantadosi, 2024). One possibility is that this prior finding can also be explained by predictability: variation in dot size within a stimulus increases the irregularity of the stimulus, making it less predictable. Another possibility is that there is a distinction between proportions generated from numerical sets alone and proportion generated from relative area.

In either case, it may be that our choice here to only use area-available stimuli contributed to the smaller difference between discrete and continuous stimuli. To explicitly test whether item size and/or congruency between cumulative area and number also impact behavior and strategy use will require a different paradigm. Furthermore, whether and how the specific magnitude used to infer the proportion (e.g., relative number vs. relative area) impacts the precision or representation of the proportional value is an interesting question that connects to deep issues in numerical cognition about the nature of magnitude representation.

Additionally, the magic ball scenario and wide variety of visual stimuli also limits the direct real-world implications of this work. Although our goal here was to understand the cognitive processes underlying proportional reasoning, an important related goal is to better understand how to present proportional information in educational and decision-making contexts. Substantial prior work has investigated (and debated) what kinds of formats may be best for teaching fractions or communicating about proportional information (Bago & De Neys, 2017; Gunderson et al., 2019; Hurst et al., 2022; Hurst et al., 2020; Lipkus & Hollands, 1999; Rau & Matthews, 2017; Tubau et al., 2019). The current work cannot directly speak to this issue, however, as we did not assess learning and our experimental context and formats are quite different from typical educational settings. Instead, by providing insight into the cognitive processes involved in proportional reasoning, the current work may be used to make predictions about how different formats might be interpreted and used in educational and decision-making contexts, as a function of their general visual and spatial properties.

Second, another well-studied feature of numerical magnitude representations is the impact of ratio or distance between magnitudes on performance: magnitudes that are further apart in terms of ratio (e.g., 1:2 vs. 1:4) are easier to discriminate than values that are closer together (e.g., 1:2 vs. 1:3) (Dehaene, 2011; Halberda & Feigenson, 2008; Moyer & Landauer, 1967). Although some work has begun to address this issue with relational stimuli, including fractions and visually presented ratios (Kalra et al., 2020; Park et al., 2020), we did not focus on this research question here. One possible way to differentiate various strategies people might be using would be to investigate how the ratio between the various quantities involved (i.e., numerator, total, proportion) predict behavior. Future work should include comparisons with a larger range of ratio discrimination levels for all quantities involved to begin testing this question.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Third, we opted to use a between-subject design, where previous work has often used a within-subject design (Abreu-Mendoza et al., 2023; Boyer & Levine, 2015; Hurst & Cordes, 2018). We used a between-subject design because of prior work revealing order effects, where continuous proportional stimuli can shift people's behavior on subsequent discrete proportional judgements (Boyer & Levine, 2015; Hurst & Cordes, 2018). Given the large number of formats tested here, we would have needed substantially more participants to carefully test for order effects in a within-subject design. Thus, although the between-subject analysis introduces some additional noise between formats, it also ensures that we are capturing the default behavior with each format. Future work will be needed to investigate whether including the formats within-subject and varying the order of presentation might shift people's strategies.

Fourth, although our mixture model analysis allowed us to investigate individual and format differences in strategy use, additional work is needed to improve this approach. Parameter estimates for both the numerator comparison strategy and the proportion comparison strategy were atypical in some ways, suggesting that neither strategy fully captured the underlying cognitive process. For the proportion strategy in particular, given that the comparison process was not fully captured by our model, it's likely that both the β and B_1 estimates were incorporating additional variance not well explained by the structure of the model. Additionally, it may be that people use different ways to compute the proportional information across different formats, such as a part:part ratio computation or a part/whole proportional computation. However, our current models cannot capture these differences.

Constraints on Generality

The current samples came from adults in the USA recruited via Prolific. Although we have no a priori reason to believe these findings would not generalize to adults in the USA more broadly, there are a few factors that might impact individual differences and replicability. For example, it may be that educational experience with fractions and proportions contributes to the ease of processing pie charts, which have been a mainstay in American fraction education. However, the current pattern may not replicate in samples where prior fraction education is meaningfully different, or participants have less experience with specific kinds of proportion representations. Additionally, a primary motivation for the current study is the hypothesis that proportional reasoning is highly variable. Here, we investigated the role of spatial features, but it is likely that other features of proportion stimuli and the task procedure may also impact proportional reasoning (e.g., the availability of benchmark heuristics, such as comparing 0.25 vs. 0.75, which cross the half boundary, or 0.25 and 0.33, which do not; duration of stimulus presentation). A direct replication would require using the same stimuli and paradigm, and additional research studies are needed to investigate what other features of the stimuli and task may or may not lead to variation in proportional reasoning.

Conclusion

In summary, we find that adults' ability to compare visual non-symbolic proportions systematically varies as a function of the spatial properties of the stimulus, beyond those typically contrasted in prior work. Our findings suggest that adults' difficulty with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

discrete proportions may not be due entirely, or even primarily, to interference from the numerical information. Instead, a more general explanation that applies to both discrete and continuous proportion might be more appropriate: when the overall proportion, and maybe in particular the “whole”, is difficult to process adults are more likely to use a numerator heuristic strategy instead of a proportional strategy, resulting in worse performance. Furthermore, these findings have implications for our understanding of the processes underlying proportional reasoning, opening new questions about what causes variation in these processes and how they develop. Overall, this work provides a novel explanation of a common behavioral phenomena, contributes important methodological advancements, including the use of stimuli that systematically vary along multiple dimensions as well as the use of mathematical modeling of strategies, and highlights the need for new theoretical approaches that can explain systematic variation in proportional reasoning.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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 K99HD104990 and 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. Data and materials are publicly available: https://osf.io/zqxg7/. We thank Steven T. Piantadosi for helpful feedback on analyses and a previous version of this manuscript. A subset of these data were presented as a talk by MAH at the 2024 Mathematics Cognition and Learning Society Conference in Washington, DC. Specifically, the strategy modeling results from Experiment 1 were presented and discussed with other mathematics cognition, numerical cognition, and education researchers in attendance.

Appendix: List of all comparisons

Appendix Table:

Stimuli values use during the test trials across all Experiments

Larger Proportion			Smaller Proportion			Features of the Comparison			
Numerator	**Other**	**Proportion Magnitude**	**Numerator**	**Other**	**Proportion Magnitude**	**Numerator Congruency**	**Ratio of Proportions**	**Ratio of Numerators**	**Larger Difference**
12	30	0.286	9	23	0.281	congruent	1.018	1.333	numerators
27	20	0.574	12	10	0.545	congruent	1.053	2.250	numerators
18	17	0.514	14	17	0.452	congruent	1.137	1.286	numerators
15	14	0.517	14	17	0.452	congruent	1.144	1.071	proportions
28	19	0.596	17	16	0.515	congruent	1.157	1.647	numerators
40	8	0.833	36	14	0.720	congruent	1.157	1.111	proportions
12	36	0.250	8	31	0.205	congruent	1.220	1.500	numerators
25	18	0.581	10	11	0.476	congruent	1.221	2.500	numerators
38	11	0.776	22	13	0.629	congruent	1.234	1.727	numerators
28	10	0.737	25	18	0.581	congruent	1.269	1.120	proportions
32	15	0.681	9	8	0.529	congruent	1.287	3.556	numerators
33	11	0.750	29	21	0.580	congruent	1.293	1.138	proportions
24	20	0.545	12	17	0.414	congruent	1.316	2.000	numerators

Larger Proportion			Smaller Proportion			Features of the Comparison			
Numerator	Other	Proportion Magnitude	Numerator	Other	Proportion Magnitude	Numerator Congruency	Ratio of Proportions	Ratio of Numerators	Larger Difference
20	28	0.417	11	26	0.297	congruent	1.404	1.818	numerators
33	10	0.767	28	24	0.538	congruent	1.426	1.179	proportions
18	26	0.409	14	35	0.286	congruent	1.430	1.286	proportions
22	9	0.710	18	20	0.474	congruent	1.498	1.222	proportions
30	17	0.638	8	11	0.421	congruent	1.515	3.750	numerators
32	7	0.821	27	23	0.540	congruent	1.520	1.185	proportions
40	7	0.851	24	19	0.558	congruent	1.525	1.667	numerators
20	16	0.556	12	21	0.364	congruent	1.527	1.667	numerators
21	10	0.677	18	25	0.419	congruent	1.616	1.167	proportions
26	12	0.684	20	28	0.417	congruent	1.640	1.300	proportions
15	12	0.556	12	26	0.316	congruent	1.759	1.250	proportions
17	10	0.630	15	30	0.333	congruent	1.892	1.133	proportions
21	25	0.457	7	23	0.233	congruent	1.961	3.000	numerators
18	20	0.474	7	22	0.241	congruent	1.967	2.571	numerators
20	14	0.588	8	20	0.286	congruent	2.056	2.500	numerators
17	32	0.347	7	35	0.167	congruent	2.078	2.429	numerators
22	9	0.710	15	32	0.319	congruent	2.226	1.467	proportions
28	7	0.800	39	10	0.796	incongruent	1.005	1.393	numerators
23	18	0.561	27	23	0.540	incongruent	1.039	1.174	numerators
18	13	0.581	25	20	0.556	incongruent	1.045	1.389	numerators
16	26	0.381	17	30	0.362	incongruent	1.052	1.062	numerators
12	15	0.444	17	24	0.415	incongruent	1.070	1.417	numerators
38	6	0.864	39	11	0.780	incongruent	1.108	1.026	proportions
8	15	0.348	15	33	0.312	incongruent	1.115	1.875	numerators
30	9	0.769	32	16	0.667	incongruent	1.153	1.067	proportions
12	11	0.522	14	17	0.452	incongruent	1.155	1.167	numerators
32	10	0.762	35	18	0.660	incongruent	1.155	1.094	proportions
12	8	0.600	15	14	0.517	incongruent	1.161	1.250	numerators
21	16	0.568	23	24	0.489	incongruent	1.162	1.095	proportions
7	13	0.350	12	28	0.300	incongruent	1.167	1.714	numerators
14	20	0.412	15	30	0.333	incongruent	1.237	1.071	proportions
19	10	0.655	21	20	0.512	incongruent	1.279	1.105	proportions
19	14	0.576	22	27	0.449	incongruent	1.283	1.158	proportions
10	13	0.435	16	32	0.333	incongruent	1.306	1.600	numerators
12	20	0.375	14	36	0.280	incongruent	1.339	1.167	proportions
18	5	0.783	28	22	0.560	incongruent	1.398	1.556	numerators
20	6	0.769	21	18	0.538	incongruent	1.429	1.050	proportions
18	13	0.581	19	28	0.404	incongruent	1.438	1.056	proportions
20	7	0.741	23	22	0.511	incongruent	1.450	1.150	proportions
11	9	0.550	17	30	0.362	incongruent	1.519	1.545	numerators

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Larger Proportion			Smaller Proportion			Features of the Comparison			
Numerator	Other	Proportion Magnitude	Numerator	Other	Proportion Magnitude	Numerator Congruency	Ratio of Proportions	Ratio of Numerators	Larger Difference
7	10	0.412	12	33	0.267	incongruent	1.543	1.714	numerators
8	15	0.348	9	41	0.180	incongruent	1.933	1.125	proportions
12	8	0.600	15	34	0.306	incongruent	1.961	1.250	proportions
6	9	0.400	9	36	0.200	incongruent	2.000	1.500	proportions
8	6	0.571	13	33	0.283	incongruent	2.018	1.625	proportions
7	11	0.389	8	35	0.186	incongruent	2.091	1.143	proportions
15	4	0.789	16	34	0.320	incongruent	2.466	1.067	proportions

References

Abreu-Mendoza RA, Coulanges L, Ali K, Powell AB, & Rosenberg-Lee M (2020). Children's discrete proportional reasoning is related to inhibitory control and enhanced by priming continuous representations. Journal of Experimental Child Psychology, 199. 10.1016/j.jecp.2020.104931

Abreu-Mendoza RA, Powell AB, Renninger KA, Rivera LM, 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. 10.1016/j.cogpsych.2023.101575 [PubMed: 37229979]

Aust F, & Barth M (2024). papaja: Prepare reproducible APA journal articles with R Markdown. 10.32614/CRAN.package.papaja

Bago B, & De Neys W (2017). Fast logic?: Examining the time course assumption of dual process theory. Cognition, 158, 90–109. 10.1016/j.cognition.2016.10.014 [PubMed: 27816844]

Barth M (2023). tinylabels: Lightweight variable labels. Retrieved from https://cran.r-project.org/package=tinylabels

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

Begolli KN, Booth JL, Holmes CA, & Newcombe NS (2020). How many apples make a quarter? The challenge of discrete proportional formats. Journal of Experimental Child Psychology, 192, 104774. 10.1016/j.jecp.2019.104774 [PubMed: 31901724]

Bolker B, & Robinson D (2024). Broom.mixed: Tidying methods for mixed models. Retrieved from https://CRAN.R-project.org/package=broom.mixed

Boyer TW, Bradley L, & Branch Greer N (2024). Children's understanding of relative quantities: Probability judgement and proportion matching. Cognitive Development, 69, 101411. 10.1016/j.cogdev.2023.101411

Boyer TW, & Levine SC (2015). Prompting children to reason proportionally: Processing discrete units as continuous amounts. Developmental Psychology, 51(5), 615–620. 10.1037/a0039010 [PubMed: 25751097]

Boyer TW, Levine SC, & Huttenlocher J (2008). Development of proportional reasoning: Where young children go wrong. Developmental Psychology, 44(5), 1478–1490. 10.1037/a0013110 [PubMed: 18793078]

Cantrell L, & Smith LB (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. Cognition, 128(3), 331–352. 10.1016/j.cognition.2013.04.008 [PubMed: 23748213]

Dehaene S (2011). The number sense: How the mind creates mathematics. Oxford University Press.

Denison S, & Xu F (2012). Probabilistic Inference in Human Infants. Elsevier. 10.1016/B978-0-12-397919-3.00002-2

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

DeWolf M, Bassok M, & Holyoak KJ (2013). Analogical reasoning with rational numbers: Semantic alignment based on discrete versus continuous quantities.

Fabbri S, Caviola S, Tang J, Zorzi M, & Butterworth B (2012). The role of numerosity in processing nonsymbolic proportions. The Quarterly Journal of Experimental Psychology, 65(12), 2435–2446. 10.1080/17470218.2012.694896 [PubMed: 22794101]

Feigenson L, Dehaene S, & Spelke E (2004). Core systems of number. Trends in Cognitive Sciences, 8(7), 307–314. 10.1016/j.tics.2004.05.002 [PubMed: 15242690]

Gebuis T, & Reynvoet B (2012). The Role of Visual Information in Numerosity Estimation. PLOS ONE, 7(5), e37426. 10.1371/journal.pone.0037426 [PubMed: 22616007]

Gohel D, & Skintzos P (2024). Flextable: Functions for tabular reporting. Retrieved from https://CRAN.R-project.org/package=flextable

Grolemund G, & Wickham H (2011). Dates and times made easy with lubridate. Journal of Statistical Software, 40(3), 1–25. Retrieved from https://www.jstatsoft.org/v40/i03/

Gunderson EA, Hamdan N, Hildebrand L, & Bartek V (2019). Number line unidimensionality is a critical feature for promoting fraction magnitude concepts. Journal of Experimental Child Psychology, 187, 104657. 10.1016/j.jecp.2019.06.010 [PubMed: 31374537]

Halberda J, & Feigenson L (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. Developmental Psychology, 44(5), 1457–1465. 10.1037/a0012682 [PubMed: 18793076]

Hoffman MD, & Gelman A (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, 15, 31.

Hollands JG, & Dyre BP (2000). Bias in proportion judgments: The cyclical power model. Psychological Review, 107(3), 500–524. 10.1037//0033-295X.107.3.500 [PubMed: 10941278]

Hurst MA, Boyer TW, & Cordes S (2021). Spontaneous and directed attention to number and proportion. Journal of Experimental Psychology: Learning, Memory, and Cognition. 10.1037/xlm0001084

Hurst MA, Butts JR, & Levine SC (2022). Connecting symbolic fractions to their underlying proportions using iterative partitioning. Developmental Psychology, 58(9), 1702–1715. 10.1037/dev0001384 [PubMed: 35511519]

Hurst MA, & Cordes S (2018). Attending to relations: Proportional reasoning in 3- to 6-year-old children. Developmental Psychology, 54(3), 428–439. 10.1037/dev0000440 [PubMed: 29083218]

Hurst MA, & Levine SC (2022). Children’s understanding of most is dependent on context. Cognition, 225, 105149. 10.1016/j.cognition.2022.105149 [PubMed: 35533418]

Hurst MA (2025). Spatial Features and Proportion Comparison. Data can be retrieved from: https://osf.io/zqxg7. Doi:

Hurst MA, Massaro M, & Cordes S (2020). Fraction magnitude: Mapping between symbolic and spatial representations of proportion. Journal of Numerical Cognition, 6(2), 204–230. 10.5964/jnc.v6i2.285

Hurst MA, & Piantadosi ST (2024). Continuous and discrete proportion elicit different cognitive strategies. Cognition, 252, 105918. 10.1016/j.cognition.2024.105918 [PubMed: 39153444]

Hurst MA, Shaw A, Chernyak N, & Levine SC (2020). Giving a larger amount or a larger proportion: Stimulus format impacts children’s social evaluations. Developmental Psychology, 56(12), 2212–2222. 10.1037/dev0001121 [PubMed: 33090833]

Jeong Y, Levine SC, & Huttenlocher J (2007). The Development of Proportional Reasoning: Effect of Continuous Versus Discrete Quantities. Journal of Cognition and Development, 8(2), 237–256. 10.1080/15248370701202471

Kalra PB, Binzak JV, Matthews PG, & Hubbard EM (2020). Symbolic fractions elicit an analog magnitude representation in school-age children. Journal of Experimental Child Psychology, 195, 104844. 10.1016/j.jecp.2020.104844 [PubMed: 32244000]

Kassambara A (2023). Ggpubr: ‘ggplot2’ based publication ready plots. Retrieved from https://CRAN.R-project.org/package=ggpubr

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. 10.1109/TVCG.2023.3327195 [PubMed: 37883271]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest package: Tests in linear mixed effects models. Journal of Statistical Software, 82(13), 1–26. 10.18637/jss.v082.i13

Leibovich T, Katzin N, Harel M, & Henik A (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. Behavioral and Brain Sciences, 40, e164. 10.1017/S0140525X16000960 [PubMed: 27530053]

Lipkus IM, & Hollands JG (1999). The Visual Communication of Risk. JNCI Monographs, 1999(25), 149–163. 10.1093/oxfordjournals.jncimonographs.a024191

Lourenco SF, & Aulet LS (2022). A theory of perceptual number encoding. 10.31234/osf.io/p7rqn

Lüdecke D (2024). sjPlot: Data visualization for statistics in social science. Retrieved from https://CRAN.R-project.org/package=sjPlot

Matthews PG, Lewis MR, & Hubbard EM (2016). Individual Differences in Nonsymbolic Ratio Processing Predict Symbolic Math Performance. Psychological Science, 27(2), 191–202. 10.1177/0956797615617799 [PubMed: 26710824]

Mock J, Huber S, Bloechle J, Dietrich JF, Bahnmueller J, Rennig J, … Moeller K (2018). Magnitude processing of symbolic and non-symbolic proportions: an fMRI study. Behavioral and Brain Functions, 14(1), 9. 10.1186/s12993-018-0141-z [PubMed: 29747668]

Möhring W, Newcombe NS, Levine SC, & Frick A (2016). Spatial Proportional Reasoning Is Associated With Formal Knowledge About Fractions. Journal of Cognition and Development, 17(1), 67–84. 10.1080/15248372.2014.996289

Moyer RS, & Landauer TK (1967). Time required for judgements of numerical inequality. Nature, 215, 1519–1520. [PubMed: 6052760]

Müller K, & Wickham H (2023). Tibble: Simple data frames. Retrieved from https://CRAN.R-project.org/package=tibble

Odic D, Libertus ME, Feigenson L, & Halberda J (2013). Developmental change in the acuity of approximate number and area representations. Developmental Psychology, 49(6), 1103–1112. 10.1037/a0029472 [PubMed: 22889394]

Odic D, & Starr A (2018). An Introduction to the Approximate Number System. Child Development Perspectives, 12(4), 223–229. 10.1111/cdep.12288 [PubMed: 30534193]

Park Y, Viegut AA, & Matthews PG (2020). More than the sum of its parts: Exploring the development of ratio magnitude versus simple magnitude perception. Developmental Science. 10.1111/desc.13043

Piantadosi ST (2016). Efficient estimation of Weber's W. Behavior Research Methods, 48(1), 42–52. 10.3758/s13428-014-0558-8 [PubMed: 26122978]

R Core Team. (2024). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rau MA, & Matthews PG (2017). How to make 'more' better? Principles for effective use of multiple representations to enhance students' learning about fractions. ZDM, 49(4), 531–544. 10.1007/s11858-017-0846-8

Stan Development Team. (2024). RStan: The R interface to Stan. Retrieved from https://mc-stan.org/

Starr A, DeWind NK, & Brannon EM (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. Cognition, 168, 222–233. 10.1016/j.cognition.2017.07.004 [PubMed: 28715669]

Stevens SS (1957). ON THE PSYCHOPHYSICAL LAW. Psychological Review, 64, 29.

Tubau E, Rodríguez-Ferreiro J, Barberia I, & Colomé À (2019). From reading numbers to seeing ratios: a benefit of icons for risk comprehension. Psychological Research, 83(8), 1808–1816. 10.1007/s00426-018-1041-4 [PubMed: 29931591]

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

Wickham H (2023a). Forcats: Tools for working with categorical variables (factors). Retrieved from https://CRAN.R-project.org/package=forcats

Wickham H (2023b). Stringr: Simple, consistent wrappers for common string operations. Retrieved from https://CRAN.R-project.org/package=stringr

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

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

Wickham H, François R, Henry L, Müller K, & Vaughan D (2023). Dplyr: A grammar of data manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham H, & Henry L (2023). Purrr: Functional programming tools. Retrieved from https://CRAN.R-project.org/package=purrr

Wickham H, Hester J, & Bryan J (2024). Readr: Read rectangular text data. Retrieved from https://CRAN.R-project.org/package=readr

Wickham H, Vaughan D, & Girlich M (2024). Tidyr: Tidy messy data. Retrieved from https://CRAN.R-project.org/package=tidyr

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Public Significance Statement

Both adults and children tend to make errors with proportional information, such as a set of dots where 80% are orange and 20% are blue, and these errors can have negative consequences for math education and everyday decision making. The current study shows that adults' difficulties correctly using proportional information may be due to visual and spatial features of the information, such as the structure of the total quantity (e.g., a circular pie chart vs. an amorphous blob) or the spatial relations between the subparts of the proportion (e.g., a set of intermixed blue and orange dots vs. a set with blue on one side and orange on the other). These task irrelevant features change the strategies people use and the errors they make, providing new insight into how to prevent these errors across contexts.

Author Manuscript

Figure 1:

Example stimuli representative of the various ways prior work has instantiated and compared discrete and continuous stimuli (Panel A) and spatial separateness (Panel B). Note that our descriptions use the term “discretized” for stimuli that are divided part-whole shapes and “discrete” for sets of independent objects, but these terms have been used in other ways in prior work, typically with the term “discrete” being used for both kinds of stimuli.

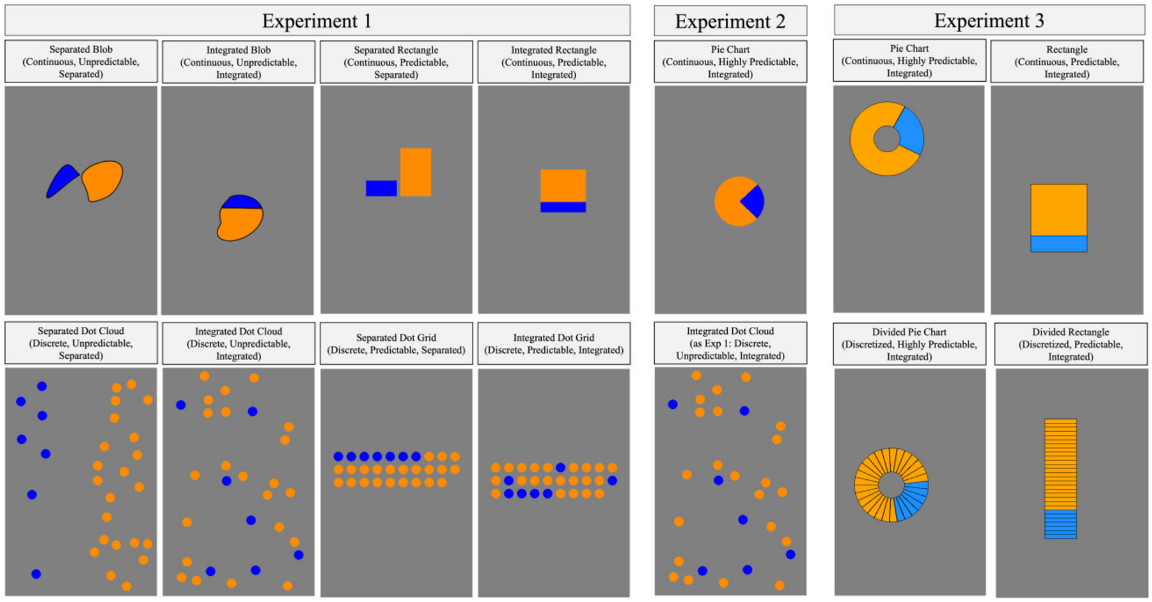

Figure 2.

Example stimuli used across Experiments 1, 2, and 3, as displayed in Panels A, B, and C, respectively. Labels and descriptions provide our operationalization of the three spatial features we use to simplify the stimulus space, with specific contrasts varying across experiments.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

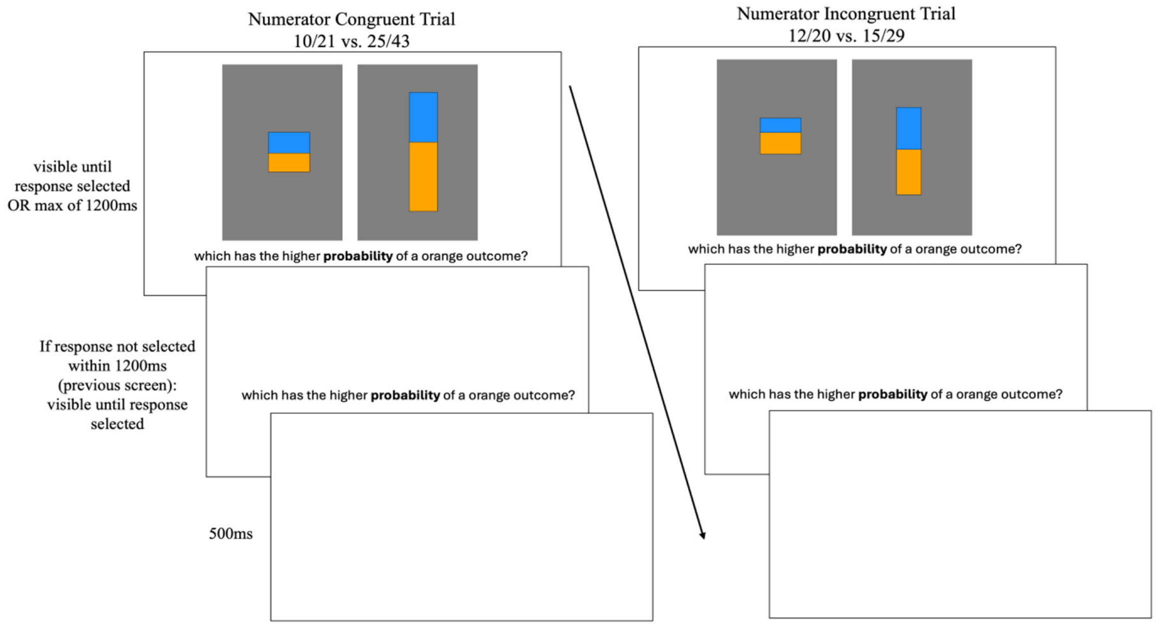

Figure 3.

A schematic of the trial procedure used in all three experiments, using the integrated rectangle format (from Experiment 3) as an example and orange as the assigned numerator color. The left panel displays an example trial where the numerator is congruent with proportion and the right panel displays an example trial where the numerator is incongruent with the proportion. Note that the size of the screen (i.e., the white bounding box around the stimuli) may vary depending on the size of the screen participants used to complete the study. Additionally, the text here has been enlarged for readability in the manuscript.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

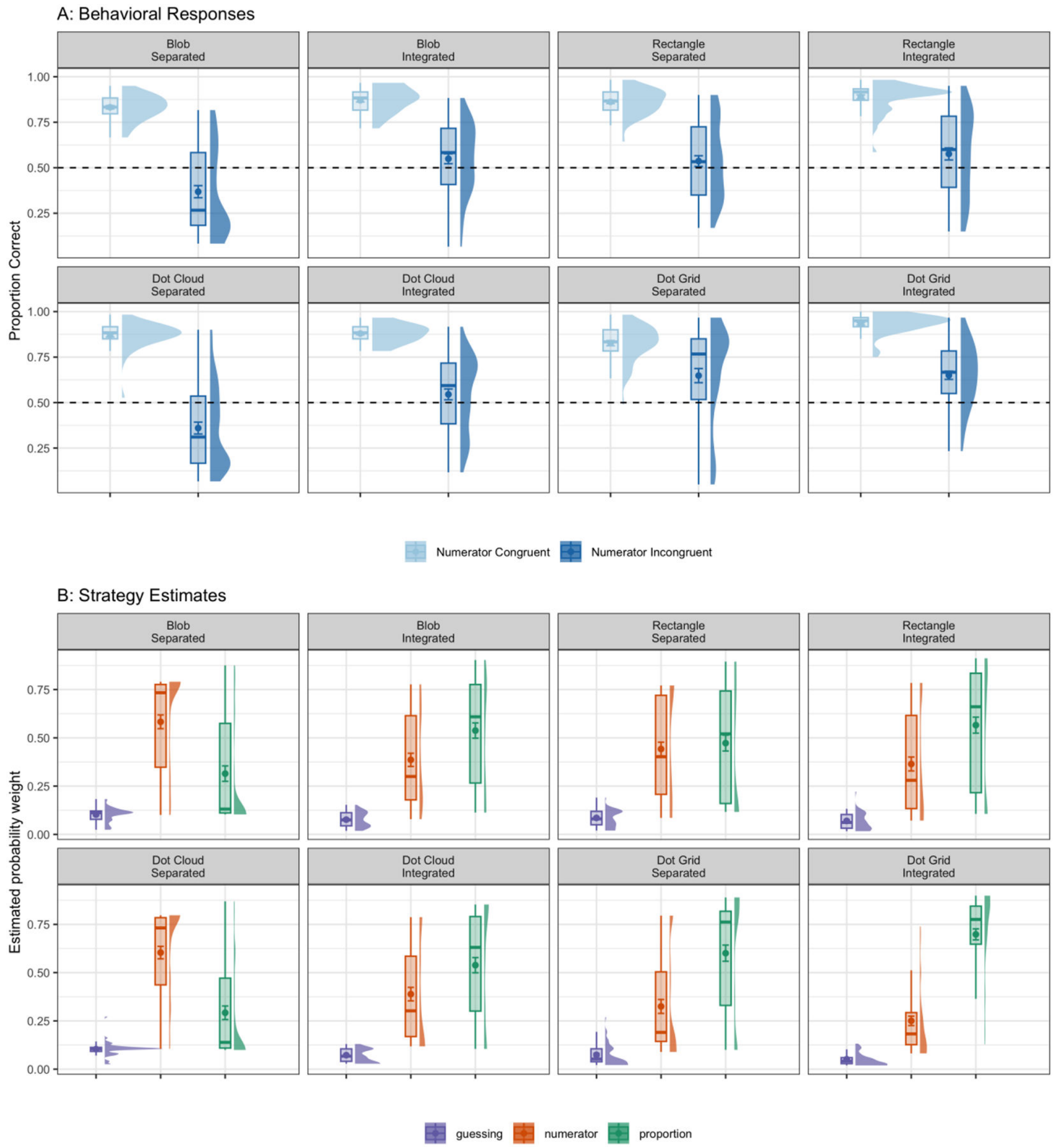

Figure 4.

Data from Experiment 1. Panel A: Proportion of trials correct, aggregated at the participant level for ease of visualization, with chance at 50% (dotted line). There were significant differences between congruent (light blue; left) and incongruent (dark blue; right) trials for each format, but the size of the effect varied across formats. Panel B: Estimated probability weights for each of the three strategies included in the mixture model: guessing (purple), numerator (orange), and proportion (green). Guessing was low for all formats, while the highest probability strategy and the distribution of the individual level data varied across formats. In both panels: half-violin plots provide a smoothed density plot of the distribution of individual participant aggregated data. Box plots provide the median (central line), interquartile range (box), and full range (vertical whiskers). The point and error bars are the mean +/− one standard error. The top row are formats categorized as continuous, and the bottom row as discrete; the left half are categorized as (relatively) unpredictable and the right half as (more) predictable; separated vs. integrated versions of the formats alternate and are labeled.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

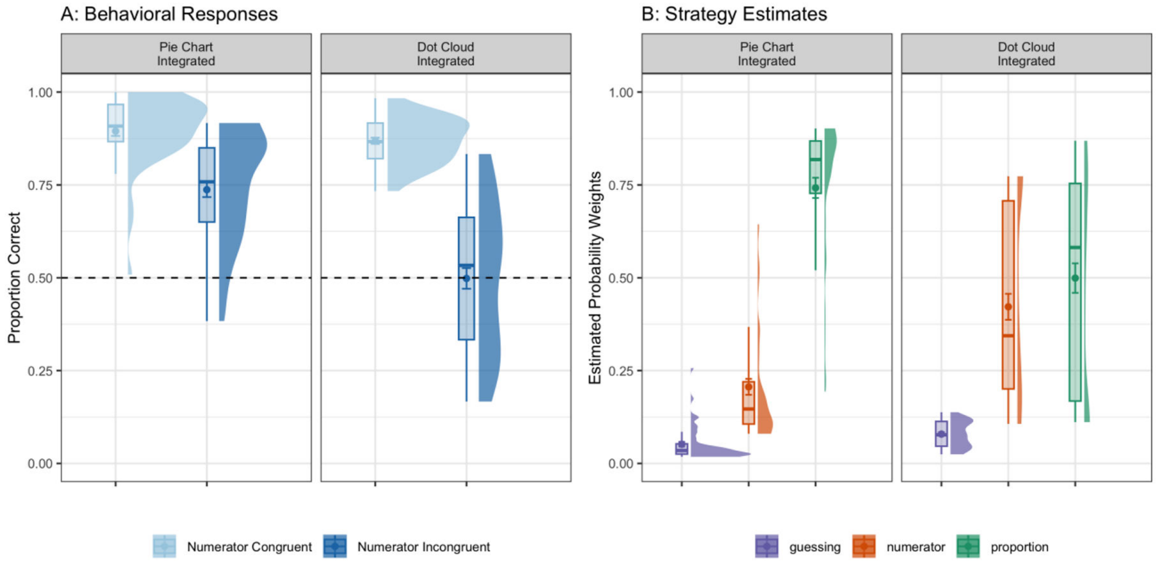

Figure 5.

Data from Experiment 2. Panel A: Proportion of trials correct, aggregated at the participant level for ease of visualization, with chance at 50% (dotted line). There were significant differences between congruent (light blue; left) and incongruent (dark blue; right) trials for each format, but the effect was larger for dot clouds than for pie charts. Panel B: Estimated probability weights from the mixture model across the three mathematically formalized strategies: guessing (purple), numerator (orange), and proportion (green). In both cases, the highest probability strategy is the proportion strategy, but this preference is weaker for dot clouds than for pie charts. In both panels: half-violin plots provide a smoothed density plot of the distribution of individual participant aggregated data. Box plots provide the median (central line), interquartile range (box), and full range (vertical whiskers). The point and error bars are the mean +/− one standard error.

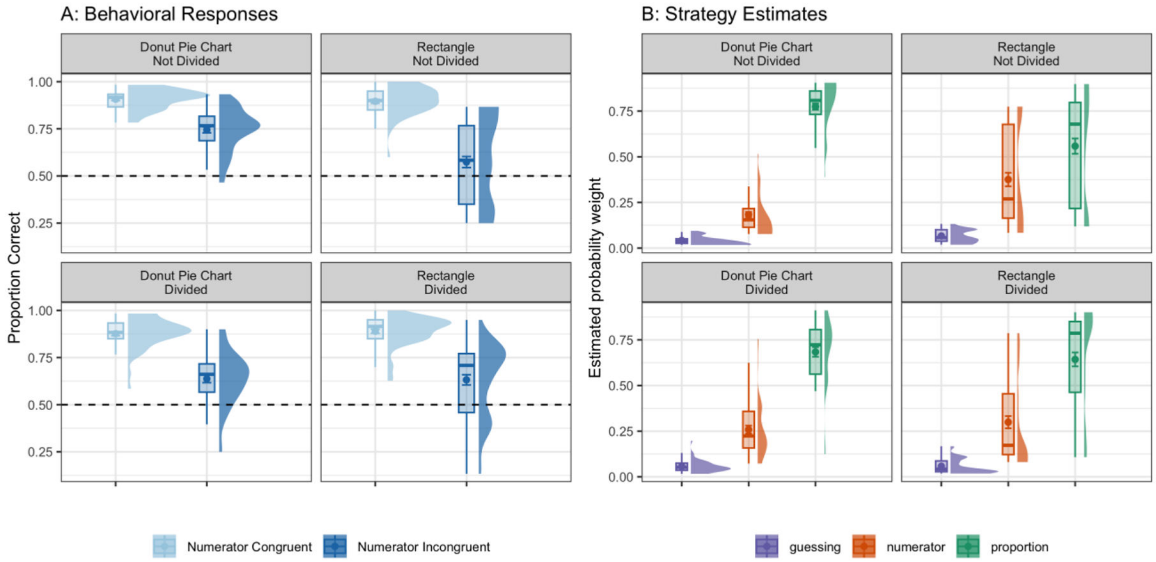

Figure 6.

Data from Experiment 3. Panel A: Proportion of trials correct, aggregated at the participant level for ease of visualization, with chance at 50% (dotted line). There were significant differences between congruent (light blue; left) and incongruent (dark blue; right) trials for each format, but the size of the effect varied across formats. Panel B: Estimated probability weights from the mixture model across the three mathematically formalized strategies. The highest probability strategy, in all cases, was the proportion strategy (green), followed by the numerator strategy (orange), and guessing (purple). In both panels: half-violin plots provide a smoothed density plot of the distribution of individual participant aggregated data. Box plots provide the median (central line), interquartile range (box), and full range (vertical whiskers). The point and error bars are the mean +/− one standard error.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

Numerator congruency effects (including descriptives) analyzed for each format separately.

	Format	Numerator Congruent M(SD)	Numerator Incongruent M(SD)	Odds Ratio for Numerator Congruence Effect (95% CI)	Numerator Incongruent vs. Chance p-value
Experiment 1	separated blobs	0.83 (0.07)	0.37 (0.23)	15.87 (10.87, 23.15)	.001
	integrated blobs	0.87 (0.06)	0.55 (0.19)	10.16 (6.53, 15.79)	.219
	separated rectangles	0.86 (0.07)	0.54 (0.21)	9.49 (6.17, 14.6)	.267
	integrated rectangles	0.89 (0.08)	0.58 (0.23)	9.44 (6.71, 13.29)	.042
	separated dot clouds	0.87 (0.07)	0.36 (0.22)	18.77 (13.94, 25.29)	< .001
	integrated dot clouds	0.88 (0.05)	0.54 (0.21)	10.23 (6.81, 15.37)	.235
	separated grid	0.83 (0.10)	0.65 (0.27)	3.24 (2.49, 4.22)	.001
	integrated grid	0.93 (0.05)	0.65 (0.17)	12.1 (7.81, 18.74)	< .001
Experiment 2	integrated pie charts	0.89 (0.09)	0.74 (0.14)	3.55 (2.42, 5.2)	< .001
	integrated dot clouds	0.87 (0.06)	0.50 (0.2)	11.27 (7.37, 17.23)	.983
Experiment 3	non-divided pie charts	0.91 (0.05)	0.74 (0.11)	4.47 (2.88, 6.94)	< .001
	non-divided rectangles	0.9 (0.07)	0.57 (0.21)	9.19 (6.62, 12.77)	.034
	divided pie charts	0.88 (0.08)	0.64 (0.14)	5.56 (3.74, 8.27)	< .001
	divided rectangles	0.89 (0.08)	0.63 (0.2)	6.31 (4.62, 8.63)	< .001

Note: Numerator Incongruent vs. Chance p-value is calculated from the intercept in an intercept only model including data from only the Numerator Incongruent trials. The Odds of the Numerator Congruency effect is from the binomial mixed effect model, separated by format. 95% CI is calculated using the Wald approximation.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2:

Model comparisons between full model (all three features and interactions) and models with all predictors involving a specific feature removed.

	number of parameters	AIC	BIC	AIC model v. full model	BIC model v. full model	Chisquare full model vs. model
Full Model	18	44,581.45	44,739.29	-	-	-
Remove: Separateness	10	44,815.93	44,903.62	234.48	164.33	250.48 (df = 8) $p < .001$
Remove: Discreteness	10	44,702.92	44,790.61	121.47	51.32	137.47 (df = 8) $p < .001$
Remove: Regularity	10	44,795.85	44,883.54	214.4	144.25	230.40 (df = 8) $p < .001$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3:

Means (N with this strategy as their highest weighted strategy) of the estimated probability weights, summarized across individuals

		Pre-Defined Strategy			
	Format	**guessing**	**numerator**	**proportion**	**Binomial Exact Test**
Experiment 1	Separated Blob	0.10	0.58 (34)	0.31 (15)	.009
	Integrated Blob	0.08	0.39 (18)	0.54 (29)	.144
	Separated Rectangle	0.09	0.44 (25)	0.47 (26)	> .999
	Integrated Rectangle	0.07	0.36 (18)	0.57 (32)	.065
	Separated Dot Cloud	0.10	0.60 (35)	0.29 (13)	.002
	Integrated Dot Cloud	0.07	0.39 (19)	0.54 (30)	.152
	Separated Dot Grid	0.07	0.32 (15)	0.60 (34)	.009
	Integrated Dot Grid	0.05	0.25 (9)	0.70 (45)	< .001
Experiment 2	Integrated Pie Chart	0.05	0.21 (6)	0.74 (44)	< .001
	Integrated Dot Cloud	0.08	0.42 (21)	0.50 (28)	.392
Experiment 3	Non-Divided (Integrated) Pie Chart	0.04	0.18 (1)	0.77 (49)	< .001
	Non-Divided (Integrated) Rectangle	0.07	0.38 (16)	0.56 (31)	.040
	Divided (Integrated) Pie Chart	0.06	0.26 (3)	0.68 (46)	< .001
	Divided (Integrated) Rectangle	0.06	0.30 (13)	0.64 (39)	< .001

Note: Binomial exact test is comparing the number of people whose highest estimated weight was on the numerator strategy vs. proportion strategy

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4:

Model comparisons between full model (both features and interactions) and models with all predictors involving a specific feature removed, for Experiment 3

	number of parameters	**AIC**	**BIC**	**AIC model v. full model**	**BIC model v. full model**	**Chisquare full model vs. model**
Full Model	10	20,768.48	20,849.30	-	-	-
Remove: Discreteness	6	20,790.66	20,839.15	22.18	−10.15	30.18 (df = 4) $p < .001$
Remove: Shape	6	20,840.35	20,888.84	71.87	39.55	79.87 (df = 4) $p < .001$